Detection of Gradual Transition in Videos: Approaches and Applications

Hrishikesh Bhaumik
West Bengal University of Technology

Siddhartha Bhattacharyya
VŠB-Technical University of Ostrava

Susanta Chakraborty
Indian Institute of Engineering Science and Technology, Shibpur

Chapter 11
Detection of Gradual Transition in Videos:
Approaches and Applications

Hrishikesh Bhaumik
RCC Institute of Information Technology, India

Siddhartha Bhattacharyya
RCC Institute of Information Technology, India

Manideepa Chakraborty
RCC Institute of Information Technology, India

Susanta Chakraborty
Indian Institute of Engineering Science and Technology, India

ABSTRACT

During video editing, the shots composing the video are coalesced together by different types of transition effects. These editing effects are classified into abrupt and gradual changes, based on the inherent nature of these transitions. In abrupt transitions, there is an instantaneous change in the visual content of two consecutive frames. Gradual transitions are characterized by a slow and continuous change in the visual contents occurring between two shots. In this chapter, the challenges faced in this field along with an overview of the different approaches are presented. Also, a novel method for detection of dissolve transitions using a two-phased approach is enumerated. The first phase deals with detection of candidate dissolves by identifying parabolic patterns in the mean fuzzy entropy of the frames. In the second phase, an ensemble of four parameters is used to design a filter which eliminates candidates based on thresholds set for each of the four stages of filtration. The experimental results show a marked improvement over other existing methods.

1. INTRODUCTION

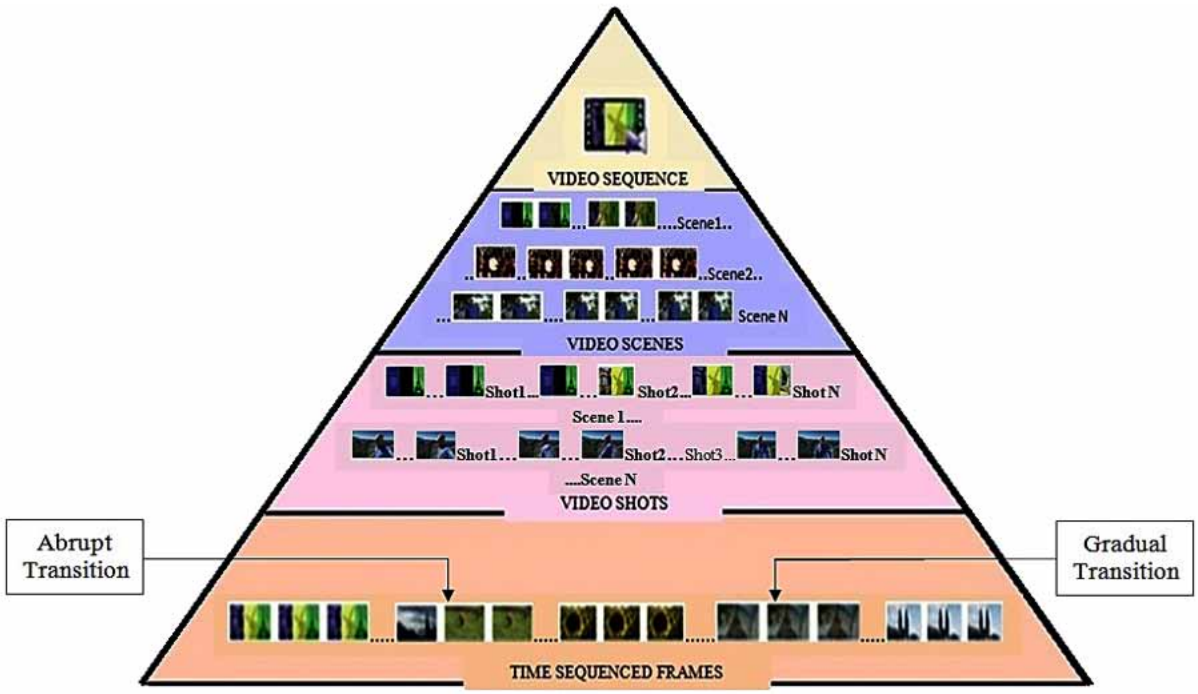
Technological development in the field of multimedia and advances in internet technology along with low cost accessibility to computing resources has led to an increased interest in research on digital videos. This has also been augmented by the fact that memory and capturing devices have witnessed a downward surge in cost. Thus video processing and analysis has become an open area of research for the last few decades. Many new and innovative concepts are being proposed regularly so as to enrich

DOI: 10.4018/978-1-5225-0498-6.ch011

the field of content-based video retrieval (CBVR). The approaches and algorithms developed for CBVR, help to align computer vision in line with human perceptions. In course of defining content-based video retrieval, "content" stands for some features of images such as color, texture, shapes etc. and the term "retrieval" refers to intricate mathematical functions or techniques which can fetch results, relevant for the end user. Thus, content-based video retrieval can be elaborated as the search for videos, where the semantic content matches the query given by a user. In fact, the query given may be using any one of the media types, i.e. text, image, audio and video. In this era, there is an abundant amount of digital information in the form of videos related to music albums, news, documentaries, sports, movies etc. These are available publicly through online digital libraries and repositories. The main problem lies with the amorphous databases or repositories in which these videos are stored. This manifests the immense need for superior search engines which are capable of searching and retrieving videos based on any query media type i.e. text, image, audio, video or a combination of these. This sets the context for the goals with which content-based video retrieval systems are built. Also it enumerates the challenges that may be faced in this research domain.

The basic step towards video content analysis is the automated segmentation of a video into its constituent shots by applying algorithms developed for the purpose. This process is referred to as temporal video segmentation and forms the basis for applications related to video summarization, indexing etc. In other words, video segmentation can be defined as the process of grouping the contents of a video stream into meaningful and manageable fragments. Fragmentation of any video sequence into its constituent units is the prerequisite step for any video processing task. A video stream may be visualized as a conglomeration of a set of different scenes. Scenes in a video are clusters of successive shots having visual similarities. Hence, scenes are composed of shots having semantic similarity. Further scenes can be divided into its composing shots. A shot in a video is a sequence of consecutive frames captured continuously from a single camera having visual continuity. A shot consists of a sequence of temporally related frames. A set of representative frames may be selected which portray the visual content of the shot. These representative frames form the summary of the shot and are called key-frames. The hierarchical elucidation of video segmentation is given below in Figure 1. Shot boundary is the demarcation point which marks the end of one shot and the beginning of another shot. A shot boundary signifies distinct change in the visual contents of consecutive shots and represent discontinuity in high level features including edges, shapes etc. of objects in the spatial and temporal domains. Shot boundaries represent transitions in visual content which are incorporated at the editing stage of the video. The transition effects are categorized into two major categories i.e. abrupt and gradual transitions. In an abrupt transition there is an instantaneous transformation between two consecutive frames elucidated by a sudden change in the visual content of the video. Such abrupt transitions resulting from a change of camera are known as hard cuts and portray points of visual discontinuity. On the other hand, gradual transitions are the editing effects in which two or more shots are combined to enable a smooth changeover from one shot to the next. The change in visual content takes place slowly and continuously over a few frames. Fade-in, fade-out, wipe, whirls and dissolves are some of the types of gradual transitions. In case of wipe, one shot is gradually replaced by another shot using a geometric pattern moving across the screen. There are many types of wipe transitions such as, straight lines, complex shapes, split-screens, horizontal line wipes running from left, right or into the middle of frame. Fade-out transition is one in which one image is gradually replaced by a black screen or by some other image. Fade-in is just the opposite of fade-out in which a solid color or an image gradually gives way to a new image.

Figure 1. Hierarchical representation of video units



Dissolve is a type of gradual transition from one shot to another, in which the current shot fades out and the next shot fades in. In other words, dissolve is the gradual transformation at pixel-intensity level of one image into another. In such cases, the end frames of one shot are combined with the start frames of another shot so that the transition from one shot to the other occurs smoothly. A dissolve effect is characterized by a linear combination of several frames at the end of one shot $S_1 = f_1 \otimes f_2 \otimes f_3 \otimes ...... \otimes f_p$ and the beginning of another shot $S_2 = g_1 \otimes g_2 \otimes g_3 \otimes ...... \otimes g_q$. The resultant of such overlap gives rise to the following sequence:

$D = f_1 \otimes f_2 \otimes .... \otimes d_{p-k+1} \otimes d_{p-k+2} \otimes .... \otimes d_p \otimes g_{p+1} \otimes .... \otimes g_q$, where, d_j is the j^{th} dissolve frame in a sequence of k such frames. A dissolve frame is thus a linear combination of the matrix values representing two frames, one taken from each shot. Thus, the j_{th} dissolve frame may be expressed as $d_j = \alpha * f_{p-k+j} + \beta * g_j$, where α, β lie in the range [0,1]. As α decreases from 1 to 0 and β increases from 0 to 1, the units of the first shot fade away while the components of the second shot gain prominence. Depending on the editing effect, the values of α and β may or may not take a linear form. The rate of shot change is thus dictated by the rate of change of α and β. This also determines the number of frames over which the dissolve transition occurs.

Gradual transitions are hard to detect due to the continuous nature of these effects. Since the advent of Content Based Video Retrieval as a research field, detection of shot boundaries has been a major field of interest for the researchers. Since temporal video segmentation (shot boundary analysis) is the basic step for any Content Based Video Retrieval System, this research domain has attracted much attention ever since exploration in video analysis commenced. It is tough to split gradual transitions in the temporal domain as it is hard to find a clear boundary between two consecutive frames. In such transitions, the low-level and high-level features retrieved from the frames composing the transition, change gradually in the temporal domain. Thus detection of such transitions presents a major challenge to researchers. Among all types of transitions, dissolve is the most complex one to deal with. This is due to the fact that in dissolve, two images are present in superimposed manner and both fade-in and fade-out take place simultaneously. During the course of dissolve detection, there are many parameters which show a distinct pattern for gradual transitions. These parameters include entropy, numbers of edges, standard deviation etc. However, relying on these parameters alone could trigger false detections as well. For instance, when local/global motions are mixed up with other non-dissolve events, it can be easily mistaken and identified as legal dissolves due to their combined characteristics. Detection of abrupt transitions (hard cuts) is now considered to be a solved problem as many approaches with high accuracy have already been designed. However, accurate identification of the duration and points of gradual transitions in a video stream is still an open challenge for researchers.

The motivation behind this chapter is to focus on the approaches and paradigms developed for the detection of gradual transitions. Also a novel two-phased approach for detection of dissolve sequences is presented. During the first phase, candidates for the dissolves are identified. This phase aims at maximizing recall. The second phase involves a filtration stage which eliminates false candidates based on four parameters. Hence, in this stage, precision is maximized without affecting the recall.

The rest of the chapter is organized as follows. In section 2, a background is presented which includes the existing works. The conventional methods developed for transition detection and the soft computing approaches have been discussed in two separate sub-sections of section 2. Section 3 discusses the types of gradual transition. The issues and challenges faced in detection of gradual transition are presented in section 4. The low-level parameters used in dissolve detection are detailed in section 5, while the pitfalls of the parameters and reasons for these are discussed in section 6. A novel method for detection of dissolve sequences is presented in section 7. The experimental results pertaining to the proposed method are enumerated in section 8. The concluding remarks are given in section 9.

2. BACKGROUND

Over the past few decades, there has been an exponential rise in the number of videos available in online repositories. This is due to the decrease in the cost of multimedia devices. Since videos engulf the other three forms of media types i.e. text, image and audio, research in the field of content-based video retrieval (CBVR) has gained momentum. CBVR is a sub-domain of content-based information retrieval. A video can be considered as a set of images having a temporal relationship. Advancements in the field of content-based image retrieval have propelled and complemented extensive research in the domain of CBVR. CBVR stands as a culmination of the various sub-fields of video processing like video indexing, classification, summarization and retrieval. As mentioned in the previous section, video segmentation is the basic step for any video analysis application. It can be classified into two types i.e. spatial and temporal. Spatial video segmentation is mainly concerned with the tracking of objects present in the video. Temporal segmentation of video involves partitioning the video into temporal slices where each slice represents either a shot or a scene. This gives rise to temporal segmentation which entails research domains called shot boundary analysis and scene change detection. Shot boundaries are generated due to coalescing of shots during the editing process. The different edit effects that are incorporated may be abrupt transitions or gradual transitions. Since the advent of research in content-based video retrieval systems, a large number of techniques have been developed for temporal video segmentation. This engulfs algorithms for all types of shot change detection. The different approaches for shot boundary analysis can be broadly categorized into methods which fall under classical approaches and those which employ techniques built on the soft computing paradigm. The next two sub-sections deal with methods under classical and soft computing approaches respectively.

2.1 Classical Approaches to Shot Boundary Analysis

Some of the early works (Boreczky & Wilcox, 1998; Xiong et al., 1997) on shot boundary detection are part of approaches where conventional techniques have been used. In (Boreczhy & Wilcox, 1998) hidden Markov model (HMM) is used for video segmentation. Features for shot boundary detection considered in this work include an image-based distance between consecutive frames of the video, the audio level difference between the pre-frame and post-frame and also motion estimation between frames. These features are integrated on the HMM framework. The method also eliminates the use of threshold as the trained HMM take their place. (Xiong et al., 1997) deal with the problem of video partitioning, key frame computing and key frame pruning. Statistical and spatial information of an image is used for video partitioning. To deal with the problems of key-frame computation and pruning, wavelet decomposition is used to derive parameter vectors. The work also demonstrates a novel image similarity measure which is also used for inter-shot redundancy reduction. In (Abdeljaoued et al., 2000), feature points extracted from images have been used for shot boundary detection. The rate of change of feature points is used as a similarity measure for detecting these transitions. (Truong et al., 2000) developed a mechanism for detection and classification of shot transitions. The work focuses on identifying the points of transition rather than the temporal length of such transitions. (Mich et al., 1999) enumerate the video indexing techniques of these early works in a survey. A detailed analysis of the latest techniques in visual content-based video indexing and retrieval is presented in (Hu et al., 2011). The survey focuses mainly on methods for video structure analysis which include shot boundary detection, key frame extraction and scene segmentation. The approaches for extraction of features used for static summarization, video annotation and video retrieval are also discussed.

The problem of computing a threshold for the various extracted features has been the objective of research for a long time. Many conventional approaches designed for the purpose of shot boundary detection rely on development of a threshold. An approach based on average inter-frame correlation coefficient and block-based motion estimation was proposed in (Porter et al., 2001). This was used to track the image blocks in the video sequence and detect the shot transitions. (Boccignone et al., 2005) used a consistency measure of the eye fixation sequences generated by a human subject looking at the video. The approach detected both abrupt and gradual transitions between shots using a single method. Another approach which focused on a unified model for transition detection was proposed in (Bescós et al., 2005). The approach is based on mapping the inter-frame distances onto a new space which is based on determining a sequence-independent threshold. This mapping also incorporates frame ordering information in the threshold determining process. (Nam and Tewfik, 2005) devised a method based on polynomial data interpolation to detect gradual transitions. B-spline interpolation curve fitting is used to determine the goodness of fit.

(Liu et al., 2004) designed a shot boundary detection method based on constant false-alarm ratio (CFAR). The threshold for determining a cut was based on CFAR. (Liu and Chen, 2002) used temporal statistics based on features extracted from frames of the video. The current frame is matched with the created model. A cut is detected if the features of the current frame do not properly fit the model. Principal Component Analysis (PCA) is primarily used for model creation from the features. (Qi et al., 2003) take a different approach towards the problem of video segmentation. The problem is transformed into a multiple class categorization issue through supervised learning. A manually labeled training dataset is used for classification purposes. The approach achieves better performance than threshold based methods. The work in (Hanjalic, 2002) is based on minimizing the average detection-error probability. The method takes into account apriori knowledge relevant to shot-boundary detection. The factors taken into consideration are shot-length distribution, visual discontinuity patterns at shot boundaries, and characteristic temporal changes of visual features. The approach taken in (Miene et al., 2003) is to combine the results from three features extracted from images. The features taken into consideration are FFT-features, image luminance values, and gray level histogram differences. Adaptive thresholds based on all the three features are taken for determining the shot boundaries. A K-step slipped window and adaptive threshold was used for shot boundary detection in (Qin et al., 2010). (Lefevre et al., 2003) present a review of the video segmentation methods based on time complexity. In the earlier reviews, only recognition rate and ability to categorize the shot transitions were taken into consideration. The review is helpful for researchers since it deals with image features like pixel, histogram, block characteristics, motion, etc. A survey on temporal video segmentation is detailed in (Koprinska & Carrato, 2001). It describes the steady development of techniques for the uncompressed domain methods which were modified and applied into the compressed domain. The review also deals with the topic of camera operation recognition.

Gabor filtering technique was used for video cut detection in (Barbu, 2009). In this approach, tri-dimensional images are computed from 2D Gabor filtering. The use of a threshold was eliminated by using an unsupervised automatic distance classifier method. A real-time adaptive threshold technique was proposed in (Yasira & Natarajan, 2009). The proposed algorithm has a low computational complexity (which is imperative for application in real-time systems) and can deal with changing light effects. Another method proposed in (Liu et al., 2008) was reported to be effective for flashlight and fast background changes. These effects often cause false alarms. To counter this, a post refinement strategy using local feature analysis was used in the work. A linear transition detector was designed in (Grana & Cucchiara, 2007) which provided a unified approach for shot boundary analysis. The method was tested on publicly available videos of sports and also the TRECVID 2005 dataset.

(Cerneková et al., 2006) focused on detection of abrupt transitions, fade-in and fade-outs by taking into consideration the mutual information and joint entropy on the set of video frames. The method was tested on the TRECVID 2003 benchmark having videos containing high object and camera motion. The approach is reported to work well for detecting abrupt transitions and fades. Another approach which used the mutual information computed on the components of the HSV color model was proposed in (Bai et al., 2008). In this work a Petri-Net model was used to denote the boundary frames of the shots. The method can be applied for the detection of both hard cuts and gradual transitions.

The significant features of the 3D structure of a scene are modeled and tracked over the frames of a video in (Donate & Liu, 2010) for detecting the shot boundaries. The SLAM technique is used for the purpose. Shot transitions are detected by observing the feature tracking ability of the system. Color coherence change was used in (Tsamoura et al., 2008a) for detecting gradual transitions. The method shows lower sensitivity to local or global motion. Also this approach eliminates the need for threshold selection. Multiple features are used in (Tsamoura et al., 2008b) by the authors resulting in a meta-segmentation scheme. Each of the features used in the work could separately be employed for the detection of gradual shot transitions. Conditional Random Fields (CRFs) are useful for parameter estimation and inference. CRF was employed for the detection of gradual transitions in (Yuan et al., 2007). The approach developed in (Šarić et al., 2008), utilizes twin comparison method and absolute difference between frames for computing the ratio of dominant colored pixels to total number of pixels. The detection of gradual transitions was carried out using fractal analysis. The classifier used for the work is based on Clonal Selection Algorithm. (Chavez et al., 2006) used a kernel-based SVM classifier to detect shot boundaries. The method gets rid of the need for a threshold or any pre-processing step to compensate motion or post-processing filtering in order to eliminate falsely detected transitions. A high speed approach was devised in (Kawai et al., 2007) in which the detection was based on using multiple features. In this work, only those parts of the video were analyzed which are the likely candidate regions for a dissolve. The algorithm was tested on videos taken from the TRECVID 2007 dataset. A formal study of color-structure descriptors used for shot boundary detection is presented in (Abdelali et al., 2009). A comprehensive study of the shot boundary detection techniques has been enumerated in (Yuan et al., 2007). The study elaborates three techniques i.e. visual content representation, construction of signal based on similarity and classification of similarity values. The work also presents a general framework for detection of transitions. It conglomerates various shot boundary detection techniques using a graph partition model.

An intensity-based method was proposed in (Ionescu et al., 2011) for dissolve detection which was able to cope with constraints of the animated movie domain. In this method the concept of twin threshold was used in place of a global threshold. The approach allows to reduce false detections caused by steep intensity fluctuations as well as to retrieve dissolve caught up in other visual effects. Temporal segmentation of video using frame and histogram space was introduced in (Joyce & Liu, 2006). In this method, certain properties of a dissolves' trajectory in image-space are used to implement a simple threshold-based detector. A motion tolerant method for dissolve detection was proposed in (Su et al., 2005). In this approach, detection of dissolve occurs through manifestation of disturbance caused by the motions of the objects. Thereafter, a filter is used to eliminate confusion in the detected dissolved frames. Moreover, to model the behavior of a dissolve transition, classification of the pixels are done in three different categories i.e. proponents, fence sitters and opponents. Proponent pixels are those in which the intensity change is either monotonously increasing or decreasing. Pixels fall in the fence sitter category if the intensity of these pixels remains unchanged in an observation window. Opponents are those pixels which do not fall into the above two categories.

Morphological operators were used in (Naranjo et al., 2007) for the detection of gradual transitions. The algorithm for determining dissolve sequences is based on the computation of a simple metric between frames. In combination with the variance between the frames, morphological filtering is used to detect the dissolve effects. (Volkmer et al., 2004) used average frame similarity and adaptive threshold for detection of gradual transitions. In this approach the frames were grouped into two different sets, pre-frames and post-frames. For each of the two sets, the distance between each frame in that set and the current frame are determined. The average of these intra-set distances gives a final value, which is the average distance between that set and the current frame. The computation results in two values, one each for the pre- and post-frame sets. The ratio of these values are calculated, which is referred to as the Pre-Post ratio. The value of this ratio is used to detect gradual transitions. A chromatic video edit model for gradual transitions detection was portrayed in (Song et al., 1997). This was based on the assumption that discontinuity values belonging to a transition consist of two piece-wise linear functions of time. One function is monotonically decreasing and the other one is increasing. Such linearity is not present outside the transition area. A search for segments which are close-to-linear in the series of discontinuity values is carried out by investigating the first and second derivatives of the slope with respect to time. A close-to-linear segment is found if the second derivative is less than a pre-specified percentage of the first derivative.

An algorithm for detection of gradual transitions was proposed in (Yoo et al., 2006) which was based on the fact that most of gradual curves can be characterized by variance distribution of edge information in the frame sequences. Average edge frame sequence was obtained by performing Sobel edge detection. Features were extracted by comparing variance with those of local blocks in the average edge frames. These features were further processed by the opening operation to obtain smooth variance curves. The lowest variance in the local frame sequence was chosen as a gradual transition point. (Apostolidis and Mezaris, 2014) proposed a method for fast shot segmentation by combining global and local visual descriptors. In this approach, both abrupt and gradual transitions could be detected. The detection was based on the visual similarities of the neighboring frames of a video. For assessing frame similarity, SURF is used for local descriptor and HSV histograms are used for global descriptor. GPU-based processing is used for accelerating the analysis. Firstly, abrupt transitions are detected between the consecutive frames where there is a huge change in the visual content, expressed by a very low similarity value. Thereafter, gradual transitions are detected by computing the value of similarities in the identification of frame-sequences where a progressive change of the visual content is present. Finally, a post-processing step is performed which aims at identifying outliers due to object/camera movement and flash-lights. A concept based on accumulating histogram difference (AHD) and support points was proposed in (Ji et al., 2010). The algorithm is able to detect fades and dissolves. Another advantage of this method is that it can eliminate false detection caused by flashlight.

(Drew et al., 2000) introduced a method for detecting dissolve and wipe transitions. The approach was based on spatio-temporal images of chromatic histogram differences. According to this method all the pixels are used to create spatio-temporal images and each column of the frame is used to form a chromatic histogram. This histogram is then intersected with the histogram from previous frames. At the time of wipes, the edges appear very strong. For dissolve detection, another approach was proposed which was based on color-distance based on the 2D Cb-Cr histograms. In another work, (Rong et al., 2005) Expectation Maximization (EM) curve fitting was used on the frame to frame difference curve for detecting the gradual transitions. The peak contours are approximated by a combination of Gaussian and uniform distributions. The weight of uniform component, the average height and the relative height of the peaks are given as input features to the decision tree classifier in order to discriminate between gradual transitions and cuts. Also the method is able to differentiate between different types of gradual transitions i.e. wipes, fades and dissolves based on the flexibility of the EM curve.

Hilbert transform and feature vectors from Gray Level Co-occurrence Matrix (GLCM) are used for shot boundary detection in (Priya and Domnic, 2012). Contourlet Transform is employed by (Rao and Patnaik, 2014) for shot transition detection where features from each sub-band are extracted for determining the shot boundary. In (Lu and Shi, 2013), an approach was devised for fast detection of shot boundaries. Adaptive thresholds were used in the work for predicting the shot boundaries and width of gradual transitions. The candidates for gradual transitions are selected and singular value decomposition (SVD) is used to speed up SBD. In several approaches the problem of detecting gradual video transitions is viewed as a classification problem. In one such work (Koumousis et al., 2012), Iterative Self Organized Data Analysis (ISODATA) classification algorithm is used. The value of the kappa coefficient is computed from the confusion matrix and is used to identify the transitions. (Manjunath et al., 2011) proposed a nonparametric shot boundary detection technique which could be applied to real time systems. Eigen gap analysis was performed to detect the shot boundaries. (Sowmya and Shettar, 2013) analyzed block based histogram method and block based Euclidean distance methods by varying the block sizes. Two important conclusions were derived from the work. Firstly, the performance of the histogram method improves with increase in block size. Secondly, Euclidean distance approach performs better than the histogram method. Feature selection is an important step in Content Based Video Retrieval (CBVR). The efficacy of the retrieval algorithm is directly related to the features chosen for the purpose. (Patel and Meshram, 2012) elaborate some of the interesting features that can be extracted from video data. Several similarity measurement techniques are also described which are used in video indexing and retrieval. Also some of the predominant research issues in the area of CBVR are discussed.

Pixel value differences of consecutive frames are computed for determining an automatic threshold in (Kundu and Mondal, 2012). The method described in the work integrates an outlier removal algorithm and a false alarm elimination scheme for accurate detection of shot boundaries. Pixel-wise difference was also used in (Patel et al., 2013). In addition, color histogram method is included in the approach to improve performance. Some of the methods for video segmentation take into consideration the motion information of objects. One important work in this direction is (Poleg et al., 2014). The researchers work on egocentric videos (Poleg et al., 2014) which involve videos where the camera is under constant motion. Motion vectors are used to compute Cumulative Displacement Curves which help in temporal segmentation of such videos. A major point of concern for researchers in the field of shot boundary detection is the number of false detections due to rapid object motion. This issue was addressed in the work (Yu et al., 2011). Twin-comparison was used for approximate detection of dissolve sequences. Curve fitting is done in accordance with two parameters i.e. percentage of dissolve-supporters in a frame and the variance of coefficients in the dissolve model. The curve is used to remove non-dissolve transitions. Daubachy 4 Wavelet was integrated along with HSV Color Space for the detection of shot boundaries in (Tariq et al., 2014). Wavelet transform was also used in the work (Thakare, 2012) for temporal video segmentation. The detection of transition sequences in videos of the compressed domain present its own challenges. A fast method for detection of shot boundaries in MPEG videos was proposed in (Ma et al., 2012). The I frames in the MPEG videos are first decoded to generate the DC images components. The histogram differences are computed for detecting the approximate shot boundaries. Using the movement information of B frames, the abrupt transitions are captured precisely. The gradual transitions are located using difference values of N successive I frames. (Mahesh and Kuppusamy, 2012) perform hybridization of the frame difference as well as consecutive frame intersection methods for achieving better results. The efficacy of the method is analyzed by computing statistical measures and kappa coefficient. Tracking of figure-ground segmentation method was proposed in (Li et al., 2013) for unsupervised video segmentation. (Zajić et al., 2011) propose a method for shot boundary detection method using Multifractal Analysis (MA). In the proposed method, low-level features such as color and texture are extracted from the frames of the video. Thereafter, a feature matrix is constructed by concatenating the features. Each row of the feature matrix corresponds to feature vector of a frame. Multifractal analysis is then applied to detect the shot boundaries.

2.2 Soft Computing Approaches to Shot Boundary Analysis

Soft computing has been used extensively by researchers for detection of shot boundaries and several methods have been developed over time. A news video parsing system (Gao and Tang, 2002) was developed for temporal video segmentation as well as detection of news reader. In order to detect the shots in the video, fuzzy c-means algorithm is used. The shots are classified into two categories i.e. those containing the news reader and the other having shots of news footage. This is implemented by means of a graph-theoretical approach. The concept of fuzzy logic has been used for detection of shot boundaries in (Küçüktunç et al., 2010; Bhaumik et al., 2014). Fuzzy color histogram was used in (Küçüktunç et al., 2010) for video segmentation. The work was targeted at copy video detection as a CBVR application. The color histogram is generated with a fuzzy function on the L*a*b* color space. (Bhaumik et al., 2014) use spatio-temporal fuzzy hostility index on the pixels of the frames composing a video to find dissimilarity patterns for detection of abrupt transitions. A fuzzy logic based approach to integrate several features for shot boundary analysis is demonstrated in (Fang et al., 2006). The approach consists of two modes, one for detection of hard cuts and another for detection of gradual transitions. A mode-selector is used to switch between the two modes for efficient detection of shot boundaries. (Jadon et al., 2001) proposed a fuzzy logic based framework for temporal video segmentation. The frame-to-frame property difference values are fuzzified using Rayleigh distribution. The difference values were characterized by fuzzy terms which were used to design fuzzy rules for detecting abrupt and gradual changes. The discrimination power of rough sets was harnessed in (Shirahama et al., 2012) for event retrieval in video archives. In this approach, multiple classification rules are extracted using rough set theory to retrieve parts of event shots. Partial supervised learning is used to train the classifiers for extracting rules and achieving a higher accuracy of retrieval.

(Chan and Wong, 2011) demonstrate an approach for shot boundary detection based on Genetic Algorithm by optimizing a traditional scoring based metrics. The approach eliminates the use of thresholds which have always proved to be a bottleneck in maximizing both precision and recall. The method is based on the edge-change ratio metric. A different approach using genetic algorithm is adopted in (Chiu et al., 2000). A series of string representation is manipulated by the algorithm. The segmentations are evaluated by defining a similarity adjacency function.

Support Vector Machine (SVM) is the most popular and widely used soft computing paradigm for temporal video segmentation. (Ngo, 2003) demonstrated a robust dissolve detector based on SVM. The method extracts multi-resolution temporal slices from the frames of the video. At low resolution, the problem of dissolve detection is reduced to detection of abrupt transitions. Gabor wavelet features are extracted in the high resolution space surrounding the position of hard cuts in low resolution space. These output features are fed to a support vector machine for pattern classification. (Li et al., 2009) present a method based on a multi-class SVM to categorize the shots into cut transition, gradual transition and normal sequences. The algorithm extracts color and edge features in different directions from wavelet transition coefficients. The classification is based on the feature vectors taken from all frames within a temporal window. SVM has also been used in (Chasanis et al., 2009) for detection of both abrupt cuts and dissolve sequences in videos. A learner based strategy is adopted using a set of features which have discrimination power to differentiate between abrupt and gradual transitions. Features based on colour histogram, variation between consecutive frames and the contextual information at a time were taken into consideration. The SVM is capable of locating and characterizing transitions. (Sun et al., 2011) organize the features into a multi-dimension vector by using the method of sliding window. The approach for shot boundary detection is based on SVM and is optimized by particle swarm and Tabu searches respectively. The work demonstrates that optimization by Tabu search is more efficient than particle swarm optimization. (Ling et al., 2008) devised another method for rapid detection of shot boundaries based on SVM. The smooth intervals in the original video sequence are eliminated by detecting changes in gray level variance. Thereafter, the new frame sequence is called reordered frame sequence (RFS). Feature vectors are formed from parameters like intensity pixel-wise difference, color histogram differences in HSV space and edge histogram differences in X and Y direction. These vectors are given as input to the SVM to detect the cuts. Temporal multi-resolution is applied to the frames in RFS in order to detect gradual changes.

Methods for hardware implementation of the various algorithms have been developed for shot boundary detection. To this effect, (Hsu et al., 2009) implemented a FPGA based fully parallel digital Support Vector Machine (SVM) classifier which was used to detect shot boundaries in a continuous video stream. Particle Swarm Optimization based classifier for detection of abrupt and gradual transitions was devised in (Meng et al., 2009). Difference curves of U-component histograms from YUV model are taken as features. A sliding window mean filter is used to filter the difference curves. Further a KNN Classifier based on PSO is used to detect and classify the shot transitions. An unsupervised shot transition detection method based on Auto-associative Neural Network (AANN) was proposed in (Geetha & Palanivel, 2012). The AANN was able to classify the type of transition i.e. abrupt or gradual. The approach is tested on different genres of videos to determine its efficiency.

Although shot boundary detectors based on soft computing paradigms have proven efficacy, researchers have embarked on hybridization of the various soft computing paradigms in order to achieve better results. A fuzzy clustering neural network for detection of abrupt transition in videos was developed in (Shen & Cao, 2011). An amalgamation of genetic algorithm and SVM was presented in (Sun et al., 2011). In this work, a multi-dimensional vector was prepared from features of the pixel and compressed domains. GA was used to optimize the parameters of the SVM kernel. The main feature of the model is that it overcomes the difficulty of parameter selection for the SVM. Feature selection is a problem which has triggered enormous research over the last two decades. (Gao et al., 2005) proposed a method for selection of appropriate features from the entire gamut of the video feature space. For this purpose, rough sets and fuzzy c-means clustering was used for feature reduction and rule generation. A drawback of the fuzzy c-means algorithm is that it assumes a consistent contribution from each feature of the samples. To overcome this, a feature weighting technique was developed in (Bao et al., 2006) using Variable Precision Rough-Fuzzy Sets and incorporated into the fuzzy c-means algorithm. Fuzzy set theory and Adaboost for the detection of shot transition in videos was used in (Zhao & Cai, 2006). Videos were classified into six categories based on camera motion and color changes. Several features from the compressed domain were used to classify the shots into three categories i.e. abrupt cut, gradual transition and shots with no changes. Several techniques were proposed to increase the performance of Adaboost. The approach taken by (Lienhart, 2001) for detecting dissolve transition is mainly based on multi resolution detection approach and machine learning algorithms such as neural networks, support-vector machines etc. The detection is based on three principal ideas. A dissolve synthesizer was created for emulating dissolves of any duration from videos. The method incorporates two new features for extracting the characteristics of dissolves. Finally, the concepts of machine learning were utilized for reliable object detection.

3. TYPES OF GRADUAL TRANSITIONS

A video is created by the amalgamation of different types of scenes. The transformation from one scene to other is a result of incorporating various types of transition effects between the different scenes of the video. The two main classes of transition effects are abrupt and gradual transition. A classification of different types of transition effects is given below in Figure 2.

According to the above mentioned categorization, a brief insight into the different types of transition effects are given in further sub-sections.

3.1 Abrupt Transition

Abrupt transition refers to a sudden change in frame contents that takes place between successive frames of a video. Abrupt transitions make up almost 98% of the edit effects. Abrupt transitions are also termed as "hard cuts". During a hard cut, there is a substantial change in the visual contents occurring abruptly. The last frame before the hard cut is called the pre-cut frame and the first frame of the next shot after the hard cut is termed as post-cut frame. Pictorial depiction of a hard cut is given below in Figure 3.

3.2 Gradual Transition

Gradual transition is one in which the conversion from one shot to another takes place progressively over a set of frames. Gradual transition is often a contextual change occurring over a period of time. During the editing stage, these transitions are introduced and have different lengths. The various types of gradual transitions have been dealt with in further sub-sections.

Figure 2. Categorization of different types of transition effects

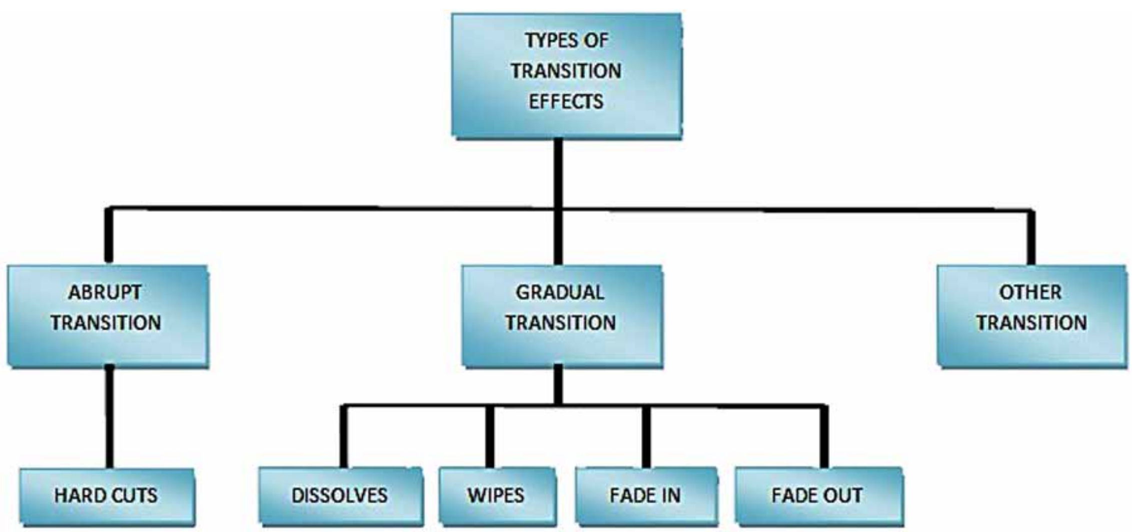
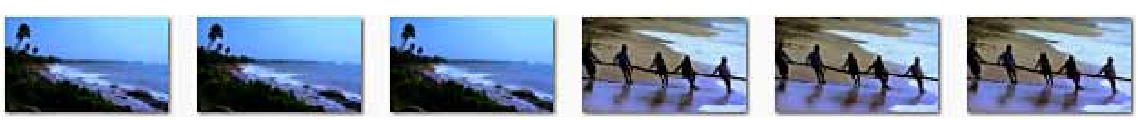


Figure 3. An abrupt transition



3.2.1 Dissolve

Dissolve is a type of gradual transition in which the last few frames of the vanishing shot temporally overlap with the fast few frames of the appearing shot. During the process of overlapping, the intensity of the vanishing shot decreases gradually and simultaneously the intensity of the appearing shot increases linearly. At the end of the transition the appearing shot gains prominence and the vanishing shot disappears completely. A dissolve transition is illustrated in Figure 4.

3.2.2 Wipes

Wipe is another type of gradual transition in which both the appearing frames of one shot and the disappearing frame of the other shots coexist in different spatial regions within two or more intermediate frames between a pair of different shots. This transition gives a concept of entering image and exiting image within the frames of the transition (the intermediate frames). At the end, the entering image is present and the exiting image gets completely removed as depicted in Figure 5.

Figure 4. Dissolve transition

Figure 5. Gradual transition (wipe)

3.2.3 Fade-in

Fade-in is a gradual transition which initiates with a few black or single colored frames. Gradually, the black or single colored frame disappears and the frames of the appearing shot gets prominent. In this transition the black frames and the starting frames of the next shots are superimposed. The intensity of the black frame decreases steadily and the frames of the next shot become prominent. The fade-in transition is depicted in Figure 6.

3.2.4 Fade-out

Fade-out is just the reverse of fade-in transition. Here the frames of a shot loose intensity and gradually turn into black or any other single colored frame. The visual content of the frames during a fade-out transition thus disappears. Figure 7 illustrates a fade-out transition.

Figure 6. Gradual transition (fade-in)

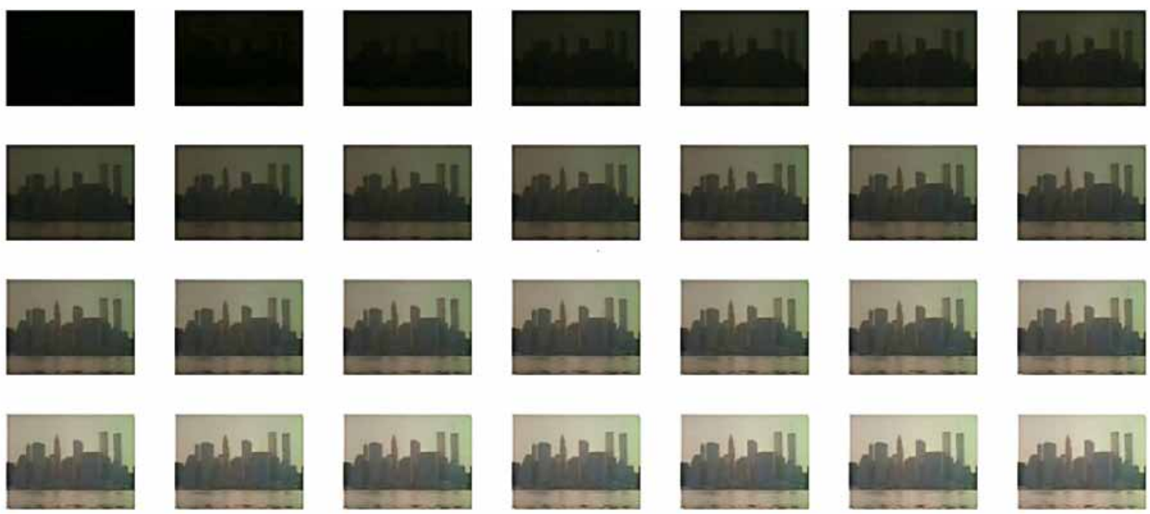


Figure 7. A fade-out gradual transition

3.2.5 Other Types of Gradual Transition

There are some special and rare transition effects that are used in some videos. These fall into the class of other transition effects. Swirl is an example of such transition where the pixels spin like a twister while a transition from one scene to the next occurs.

4. ISSUES AND CHALLENGES OF GRADUAL TRANSITIONS

The gradual transition takes place over a certain number of consecutive time sequenced frames. The contents of the video frames change progressively from one scene to another since the transition region is composed of the last few frames of one shot and an equal number of frames taken from the beginning of the next shot. In some videos, more complex edits are inserted such that a transition region is the conglomeration of frames taken from more than two shots. Determining the frames which compose the transition region become tough since the contents of frames in such regions is the combination of composing shots. The properties of the frames in the gradual transition region are close to the frames of the composing shots. The features extracted from the video frames can be broadly divided into three categories i.e. low-level, mid-level and high-level features. Low-level features include RGB values, HSV components, color histograms, statistical features like mean, variance, skewness of the pixel values, entropy, texture, wavelet features etc. Mid-level features include feature point detectors and descriptors such as SIFT (Lowe, 1999), SURF (Bay et al., 2008), DAISY (Tola et al., 2010), GIST (Oliva, 2001), BRIEF (Calonder, 2010) etc. Given an image, these feature point detectors can be used to extract salient positions in the image which could be used to capture the medium level semantics of the image. These feature points can also be used to track and recognize objects in an image. The mid-level features capture points on the objects rather than the whole semantic meaning. On the other hand, high-level features include shapes of objects, edges in the frames, optical flow (Barron et al., 1994), motion vectors (Wang et al., 2007), event modeling (Li & Sezan, 2001), etc. The high-level features are closely connected to the semantic content of a video such as scenes, objects etc. and are more natural to humans than the low-level features. As such, the low-level features are not capable of clearly distinguishing the frames which lie in the gradual transition region. Using low-level features such as color, histograms, intensity or statistical measures may lead to several false detections. The other challenges associated with accurate detection of dissolves pertain to object and camera movement. Several features have been extracted which compensate for object and camera movements. These ensure that false alarms caused due to these effects are minimized. The problem of flash light compensation has also been discussed in the literature and several methods relevant to it have been developed. Many approaches have been devised over the years to detect dissolve transitions. However, the efficiency of these approaches has not been satisfactory. In case of gradual transition detection, the appraisal of the algorithms depends on its ability to determine the location and duration of such transition. Another important factor to be considered is its capacity to judge the type of gradual transition such as dissolve, fade-in, fade-out, wipe etc. Other important issues include the robustness of the algorithm for application to various encoded file types in the compressed and uncompressed domains. The ability of the approaches developed for application to real-time systems is also an important criterion. A repository of such algorithms in web-executable format will be helpful for easy analysis and comparison of the developed methods as suggested in (Gargi et al., 2000). The reviews presented in (Sao and Mishra, 2014; Thakre, 2014; Saini and Gupta, 2015; Mittalkod et al., 2011) provide an insight into the challenges faced in this domain of research.

5. PARAMETERS FOR DETECTING GRADUAL TRANSITIONS

There are certain parameters that show distinct characteristics and behaviors when applied to the videos having gradual transitions. Based on these parameters, the gradual transitions can be detected. Some of these parameters have been elaborated in further sub-sections.

5.1. Fuzzy Hostility Index (FHI)

The measure of the amount of homogeneity or heterogeneity of a pixel with its neighborhood is termed as the Fuzzy Hostility Index (Bhattacharyya et al., 2009) of that pixel. In a fuzzy system, all elements have a certain degree of membership to a fuzzy set. A pixel having a gray level value in the range [0, 255] can be mapped to the range [0, 1]. The value in the range [0, 1] denotes the membership of that pixel to the fuzzy set WHITE denoted by μ_p. The degree of homogeneity or heterogeneity in the nth-order neighborhood of a pixel can be enumerated by the fuzzy hostility index (FHI).

The FHI (ζ) of a pixel in its $n-$order neighborhood is given by the following equation:

$$\zeta = \frac{3}{2^{n+1}} \sum_{i=1}^{2^{n+1}} \frac{\left|\mu_p - \mu_{qi}\right|}{\left|\mu_p + 1\right| + \left|\mu_{qi} + 1\right|} \tag{1}$$

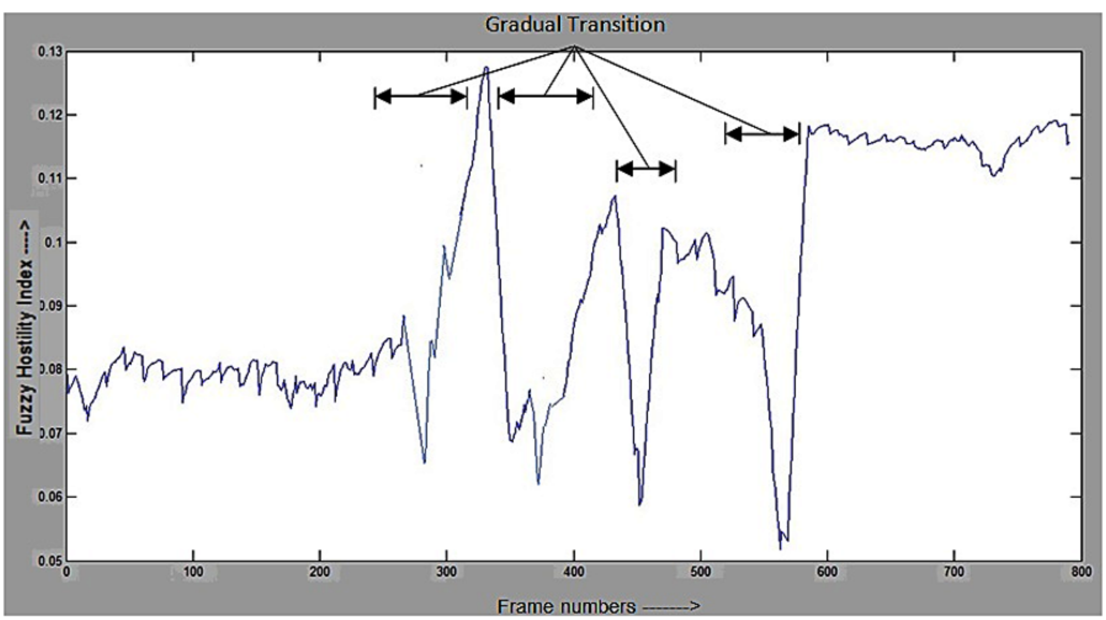
where μ_p is the fuzzy membership value of the candidate pixel and μ_{qi}; i =1, 2, 3, . . ., 2^{n+1} are the fuzzy membership values of its neighbors in an $n-$order neighborhood. The value of the fuzzy hostility index ζ lies in [0, 1]. Higher value of ζ indicates lower neighborhood homogeneity and lower value of ζ indicates the higher neighborhood homogeneity. If $\zeta = 1$, then heterogeneity is maximum. If $\zeta = 0$ then there is total homogeneity in the neighborhood. The fuzzy hostility map (FHM) is the visual representation of the FHI values of pixels in an image. In a FHM, the edges of objects become prominent because the edges represent heterogeneous regions on an image. The same is illustrated in Figure 8 and Figure 9.

The mean FHI of the images where the gradual transitions are present i.e. the regions of gradual transition are lower than the mean FHI corresponding to pure frames. This is due to the fact that the pixel values are more homogeneous in frames of the dissolve region. The frames have lower contrast than a pure frame. This occurs as more than one image contributes to the gradual transition. If the mean FHI of all the pixels in an image is computed and plotted, then a sharp dip would occur in the regions of gradual transition. The same can be observed in the graph enumerated in Figure 10.

Figure 8. Original image

Figure 9. Fuzzy hostility map

Figure 10. Mean FHI curve showing the dips during gradual transition



5.2. Edge Change Ratio (ECR)

Edge Change Ratio (Zabih et al., 1999) is a technique which was applied to detect shot boundaries. It takes into account the number of incoming and outgoing edges between each pair of consecutive frames in a video sequence. If the amount of change is greater than a specified threshold, a shot boundary is detected. For determining the ECR between a pair of frames, a number of intermediate steps are followed. The first step involves edge detection of the pair of frames from the gray scale converted image of the original one. The number of edge pixel is counted in the edge detected consecutive frames and are stored as p_{n-1} and p_n. In the next step, dilation is performed on the pair of frames followed by inversion of background. Thereafter AND operation is done between the dilated image of frame f_n and image of frame f_{n-1} obtained after edge detection. The resulting image denotes the outgoing edge pixels. The number of edge pixels is represented by EC_{n-1}^{out}. The same operation is performed between the dilated image of frame f_{n-1} and the image of frame f_n obtained after edge detection. The number of edge pixels in this case is represented by EC_n^{in} which denotes the number of incoming edge pixels. Edge Change Ratio of the frames under consideration is given by the maximum of $\frac{EC_{n-1}^{out}}{p_{n-1}}$ and $\frac{EC_n^{in}}{p_n}$.

ECR is computed from the following equation:

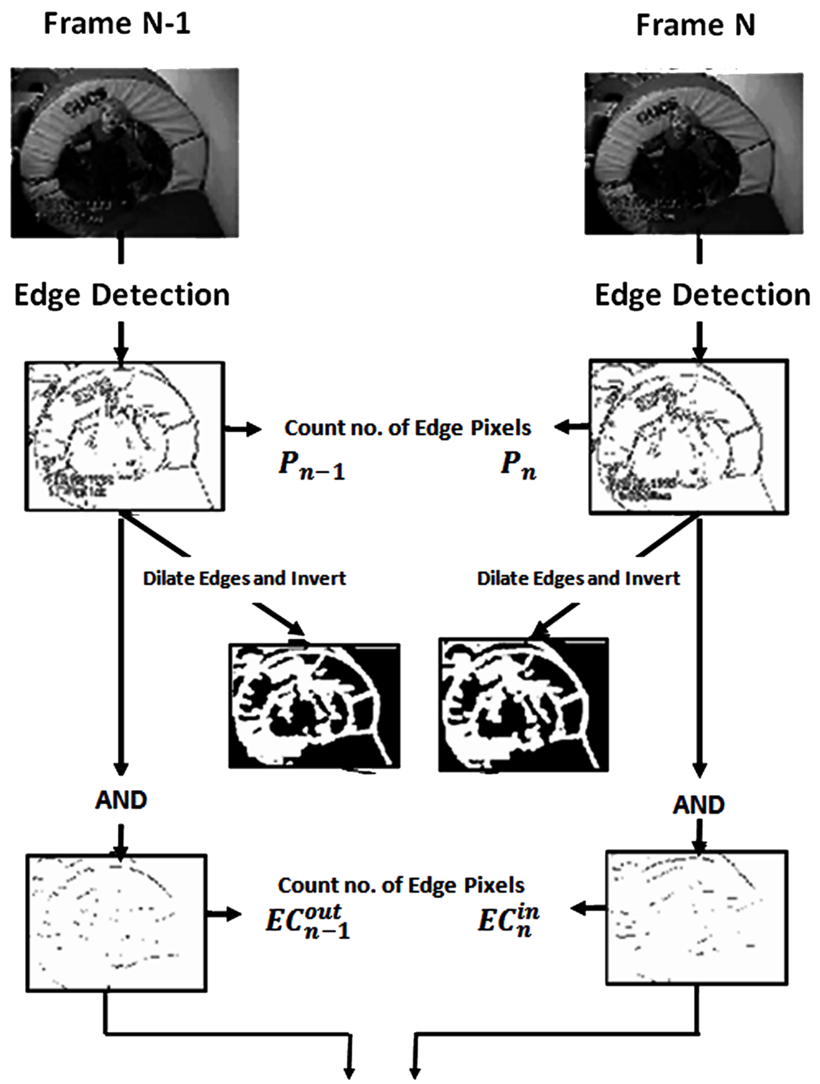
$$ECR_n = \max\left(ECR_n^{in}, ECR_{n-1}^{out}\right) = \max\left(\frac{\Sigma EC_n^{in}}{\Sigma p_n}, \frac{\Sigma EC_{n-1}^{out}}{\Sigma p_{n-1}}\right) \tag{2}$$

The pictorial representation of the method is given below in Figure 11. The value of ECR is larger at places where a gradual transition effect is present. As the number of edge changes is more at these places due to a combination of two or more images, the ECR is observed to be high.

5.3. Fuzzy Entropy

A measure of the amount of disorder in a system can be termed as entropy (Baber et al., 2011). In case of an 8-bit image, entropy can be defined as the spread of state corresponding to the gray level values which individual pixels can adopt. If an 8-bit pixel is taken into consideration, there are 256 spread of states. If all such states are equally occupied, as in the case of images which have been histogram equalized perfectly, the spread of states is a maximum. The entropy of images in such cases is maximum. On the other hand, if the image has pixel values concentrated on two states, then the entropy is low. If all the pixels have identical values, the entropy of the image is zero. If the entropy of an image decreases the information content of the image also decreases. The information content and the quality of the information content in an image can be computed as the logarithm of the probability of the amount of information conveyed by an image. The average entropy level of dissimilar shots resides at different levels. At the time of a gradual transition in a video, the entropy level shifts from the earlier level (of the disappearing shot) following a parabolic path and stabilizes at the new level (occupied by the appearing shot). This is illustrated in Figure 12. This is due to the fact that there is an increase in disorder in pixel

Figure 11. Steps of edge change ratio

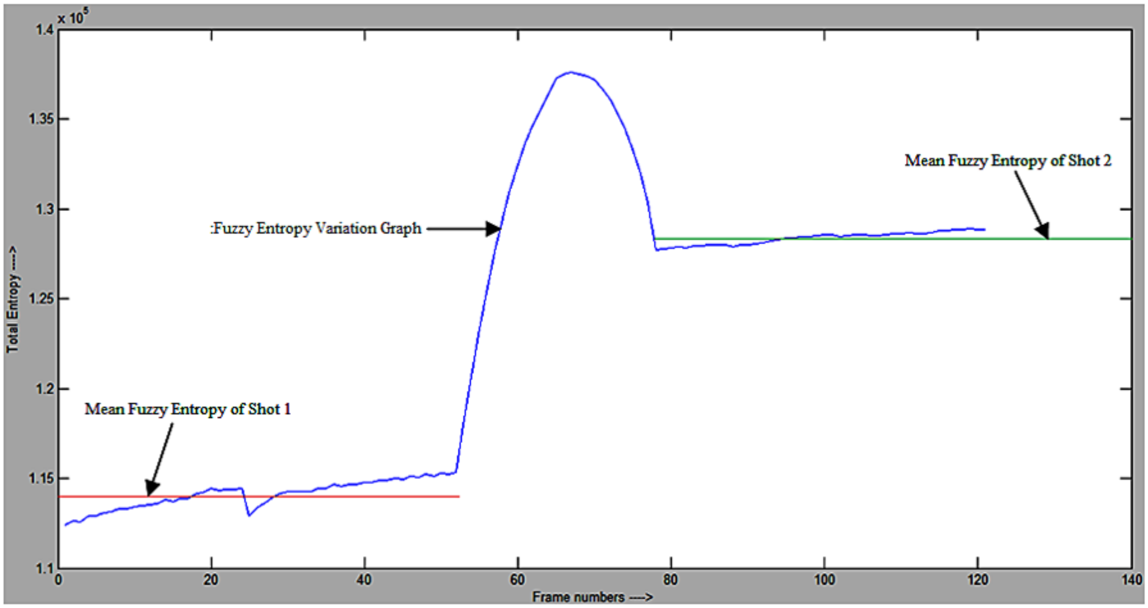


$$ECR_n = \max\left(ECR_n^{in}, ECR_{n-1}^{out}\right) = \max\left(\frac{\Sigma EC_n^{in}}{\Sigma p_n}, \frac{\Sigma EC_{n-1}^{out}}{\Sigma p_{n-1}}\right)$$

values during a dissolve transition which can be attributed to the intermingling of the image fading in and the one fading out. For a set of images participating in a dissolve, the entropy reaches maximum when the images are combined in equal proportions i.e. at the point where the intensity of the frame is exactly midway between the participating frames of the shots. At this point, the visual content portrayed by the frame is minimum.

If the pixel values of a gray scale image are mapped to the range [0, 1] by dividing each of the elements by 255, the values represent the membership of the pixels to two complementary fuzzy sets,

Figure 12. Variation of fuzzy entropy during dissolve sequence



$WHITE$ and $BLACK$. The membership function to $WHITE$ is given by μ_W, where $\mu_W = pixelvalue / 255$. The membership value of $BLACK$ is given by $\mu_B = 1 - \mu_W$. The fuzzy entropy (η) of an image having resolution $n \times m$ may be measured by the equation:

$$\eta = \sum_{i=1}^{n}\sum_{j=1}^{m} I(i,j) * \log\frac{1}{I(i,j)} + (1 - I(i,j)) * \log\frac{1}{1 - I(i,j)} \tag{3}$$

where, $I(i,j)$ is the fuzzy value of the pixel at position (i,j) in the fuzzy matrix.

During a dissolve transition, one shot gains prominence over the other. For a time, sequenced set of images obtained after disintegrating a video, the fuzzy entropy plot takes a form which is illustrated in Figure 13. The curve takes a parabolic nature at places where a dissolve transition occurs. The general equation of a 2nd degree curve may be represented as $y = ax^2 + bx + c$. which is useful for detecting candidate dissolve transitions.

5.4. Standard Deviation of Pixel Intensities

A gray scale image can be represented as a 2D matrix consisting of pixel values of an image. In case of a dissolve sequence or any other gradual transition, the matrix is a combination of two or more image matrices. Due to blending of these pixel values, the resultant images have a lower contrast than the original images combined to form such a transition. Thus, the variance of pixel values in the dissolve image is lower than the original images. If the values of standard deviation (Lienhart, 1999) are plotted, it will result in a sharp dip at places which imply the presence of gradual transitions in the video sequence. The same is evident from Figure 14.

Figure 13. Parabolic nature of fuzzy entropy during gradual sequence

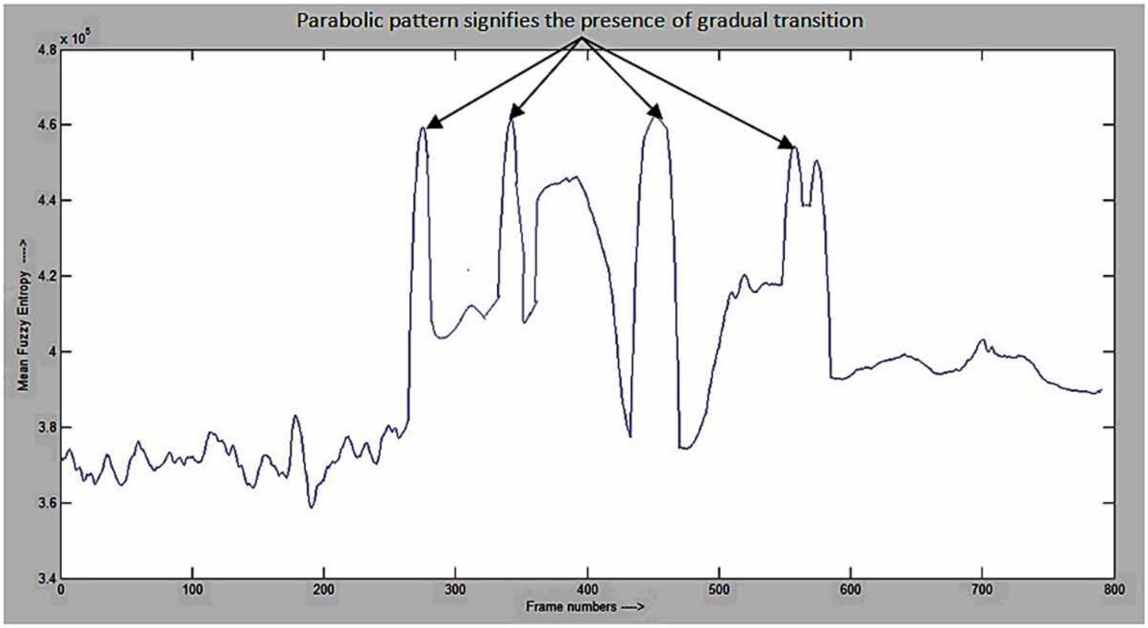
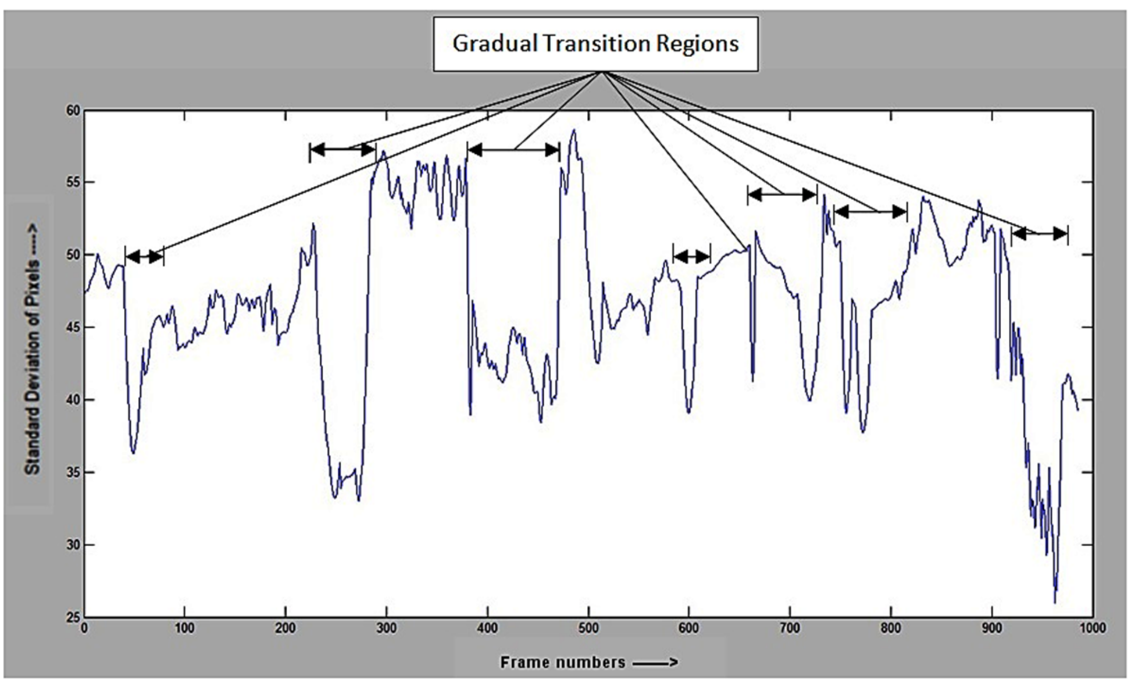


Figure 14. Plot of standard deviation of pixel intensities



5.5. Count of the Edge Pixels

Edges are basically the line segments or curved line segments formed by set of points. These occur at places of high contrast within an image. These high contrast or high frequency regions usually occur at the boundaries of objects. Edges can also be viewed as places of discontinuity where the pixel neighborhood represents heterogeneity. Detecting such edges by some set of mathematical methods which aim to categorize the set of points in digital images, at which the image intensity changes sharply or has discontinuities, is known as edge detection. Edge detection is a deep-seated tool in image processing. One such edge detector is Canny Edge Detector. Canny Edge Detector (Canny, 1986) is a multi-stage algorithm used to detect a wide range of edges in an image. Canny edge detection algorithm is mainly biased on five steps. Primarily a Gaussian filter is introduced to eradicate the noise in an image and make the image smoother. Thereafter the intensity gradient of the image is determined and non-maximal suppression is applied to get rid of spurious response to edge detection. To determine the potential of edges, double threshold is used, and then the edges are tracked by hysteresis. Lastly, all the edges that are weak and not connected to strong edges are suppressed. Hence the remaining edges are finalized as the result of the Canny edge detection algorithm.

The number of edges will be more in the frames containing dissolve transitions than the pure frames because the dissolve transition frames are formed by the overlapping of two or more frames. Thus, if there is an abrupt change in the number of edge pixels in a sequence of frames, then it becomes a candidate for dissolve transition.

6. PITFALLS OF PARAMETERS IN DETECTING GRADUAL TRANSITIONS

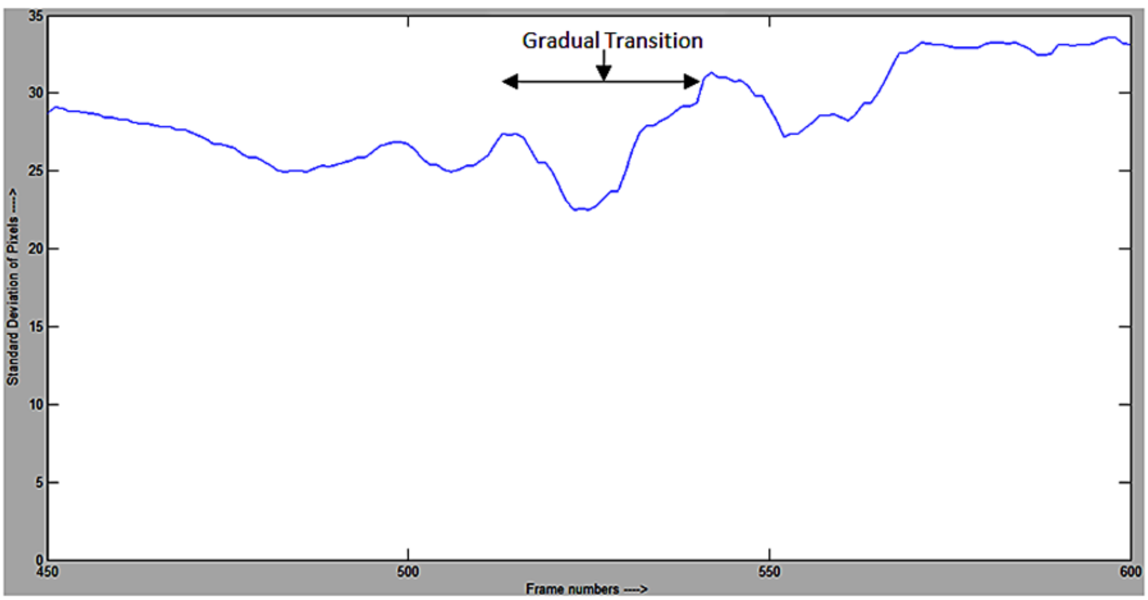
The parameters enumerated in section 5 show distinct characteristics for enabling the detection of dissolve transition sequences. These parameters are useful in detecting dissolve sequences to a limited extent. The problems in relying on a single parameter are multi-fold. Although a parameter can provide an indicative pattern for a candidate dissolve, the precision level may fall due to false alarms in different situations. Also the problem of providing a decent threshold automatically is a major problem. The further subsections discuss some cases where false detection is possible due to different situations.

6.1 False Detection of Gradual Transition Based On Standard Deviation of Pixels

As discussed in section 5.4, the standard deviation of pixel values corresponding to the frames in a video falls due to presence of dissolve transitions as contrast of the frames decreases. However, there are cases where this property may lead to false detection. Figure 15 and Figure 16 depict an example of a false positive from the video “What’s love got to do with it” by Tina Turner where the standard deviation value dips due to a close-up on the hair portion of the singer. Since most pixels have nearly the same value, so a fall in the standard deviation value of the pixel intensities triggers a dip in the plot.

Figure 15. Sequence from Tina Turner video

Figure 16. Standard deviation of pixel intensities dip during sequence in Figure 15

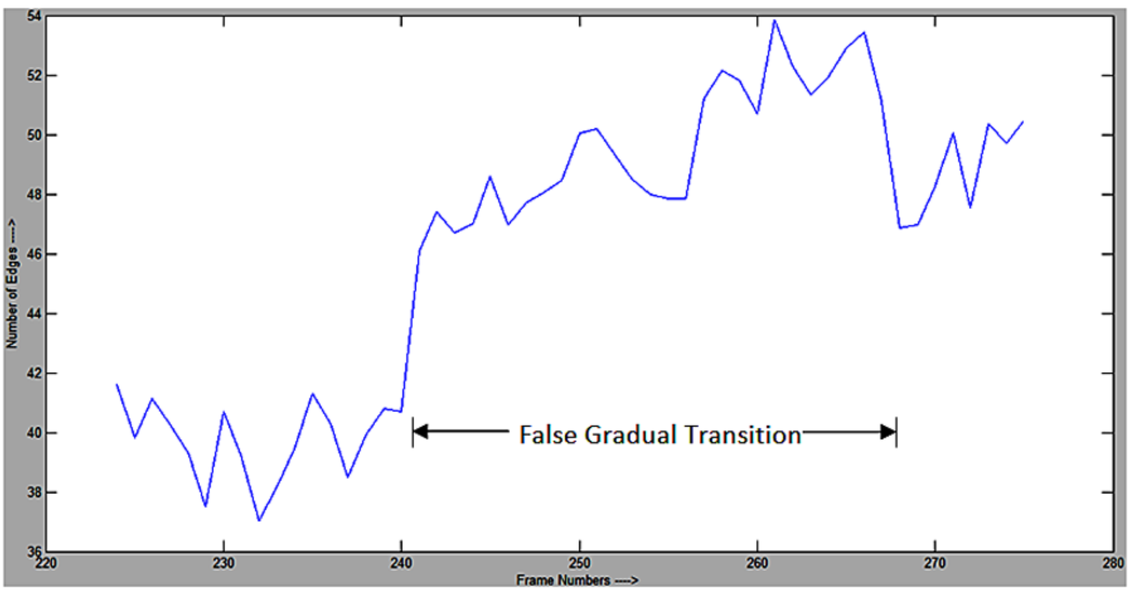


6.2 False Detection of Gradual Transition Based On Count of Edge Pixels

The number of edges increases during the occurrence of a gradual transition due to the combination of multiple images during a video sequence. Although this concept is found to be true for most video sequences, there are incidences of it being false. The false detection can occur due to various reasons. Such a false positive is demonstrated from the video "What's love got to do with it" by Tina Turner. The sequence shown in Figure 17 portrays such a situation where the number of edges rises rapidly due to appearance of buildings in the background. The rise in the number of edges and edge pixels thereof is attributed to the emergence of the various structural composites of the buildings in the scene. The graph plotted for the number of edges against the frame numbers is shown in Figure 18.

Figure 17. Sequence from Tina Turner video

Figure 18. Rise in number of edges during sequence of Figure 17



7. PROPOSED METHOD

In this section, a method for detection of dissolve sequences is elaborated. The concept used here is to employ a sequence of filters based on various low-level parameters extracted from the video sequence. In order to achieve high accuracy, a two-phased approach is used. The main target in the first phase is to increase the recall to 1 without being concerned about the precision of detection. Achieving a good recall ensures that all dissolve sequences are identified along with few other candidates which may not be legal dissolve sequences. The second phase involves an ensemble of filters which serve to eliminate the spurious dissolve sequences which were introduced in the first phase. Based on a particular low-level parameter, each filter is designed to remove fake dissolve sequences introduced in Phase 1. Thus, the second phase serves to increase precision without sacrificing recall. The design of the filters in the second phase is such that each filter employs a very conservative threshold. This ensures that after each filtration stage some of the spurious dissolve sequences that were introduced in the first phase are eliminated progressively after each filtration stage. Hence, the precision of the system improves after each stage of filtration without affecting the high recall achieved in the first phase. The two-phased dissolve detection approach is elaborated in further subsections.

7.1 First Phase

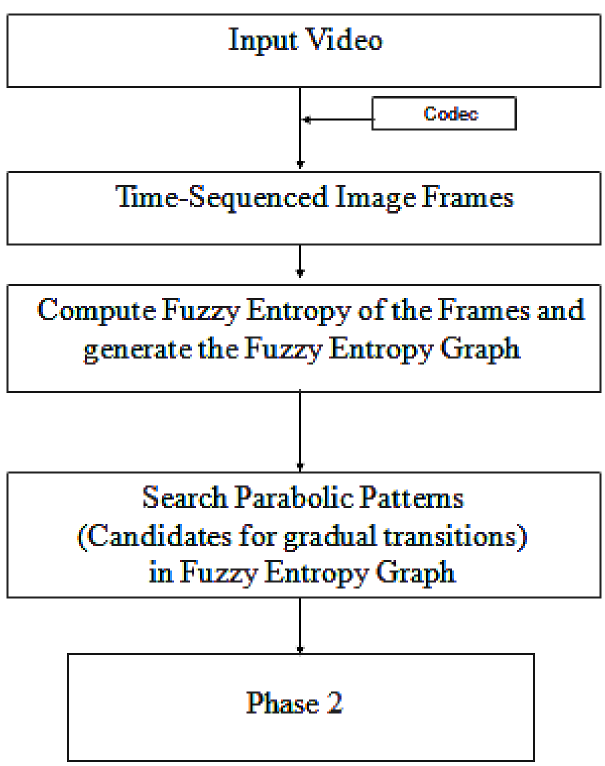
Initially, videos are fragmented into frames by means of a codec. As explained in section 5.3, the fuzzy entropy of the time sequenced frames is computed individually. The fuzzy entropy sequence generated from the frames is then analyzed for detecting parabolic sequences. A parabola is a second degree polynomial characterized by two properties:

The gradient on the curve (dy / dx) decreases monotonically as the curve is traversed in a direction of increasing x.

The change of gradient (d^2y / dx^2) is negative all along the curve in the direction of increasing x.

The above two properties are tested for the fuzzy entropy sequence. This helps in detecting all the dissolve sequences present in the video. However, there may be many false positives which may be introduced in this stage. The main target here is to include all the dissolve transitions even though all identified sequences may not be correct. These are termed as "candidate dissolves". In Figure 19, the flow of Phase 1 is enumerated. The output of Phase 1 is given as input to Phase 2, i.e the "candidate dissolves" are filtered in Phase 2.

Figure 19. Flow diagram of the first phase



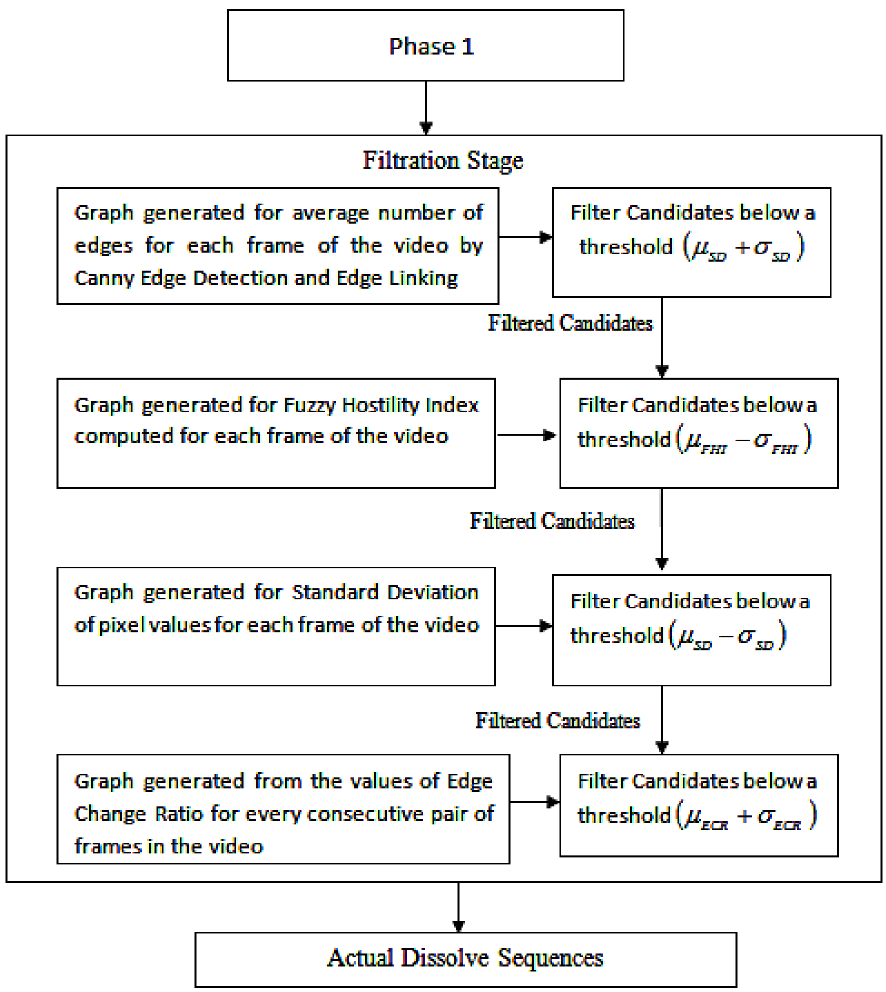
7.2 Second Phase

The second phase is the filtration stage for elimination of sequences which were wrongly identified as dissolves. This phase consists of four stages and each stage is based on a particular parameter. The four parameters which are considered include number of edge pixels, fuzzy hostility index, standard deviation of pixel intensities and edge change ratio. The purpose of this phase is to improve the precision at each stage of filtration. It is pertinent to mention here that if the true positives are filtered out unintentionally, it would lead to a fall in the recall value. The target of this phase is to curtail the false positive caused due to the first phase, as far as possible. The flow diagram of phase two is shown in Figure 20.

7.2.1 Filtration and Elimination Using Edge Detector

Edge detection is performed on the time sequenced image frames of the video by a Canny Edge Detector as explained in section 5.5. The number of edges increases abruptly during a dissolve transition. For

Figure 20. Flow diagram of Phase 2 of the proposed method



computing the threshold, the mean (μ_{ED}) and standard deviation (σ_{ED}) of the number of edges over all frames is computed. In this work, a low threshold has been selected at every stage of the filtration process to ensure that the true dissolve candidates are not eliminated. At the same time, it is ensured that the precision value increases as the selected candidates pass each stage of the filtration process. As each dissolve candidate is a collection of frame numbers, the values corresponding to each frame are tested against the threshold $\left(\mu_{ED} + \sigma_{ED}\right)$. A dissolve candidate is eliminated if none of the values corresponding to the frames in it cross the threshold.

7.2.2 Filtration and Elimination Using Fuzzy Hostility Index

The candidate dissolve sequences which pass the previous stage are fed into this stage. The average fuzzy hostility index is computed for all frames and stored in a row matrix (M_{FHI}). The mean (μ_{FHI}) and standard deviation (σ_{FHI}) of the values in M_{FHI} is calculated for determining the threshold. As explained in section 5.1, the average FHI value of the frames in the gradual transition region is lower than pure frames. Hence, in this stage, the threshold taken is $\left(\mu_{FHI} - \sigma_{FHI}\right)$. It may be seen that such a threshold is very conservative and value of average FHI for at least one frame of a candidate dissolve region must be equal or lower than this threshold in order to be categorized as a gradual transition. The candidates which do not meet the criteria are eliminated.

7.2.3 Filtration and Elimination Using Standard Deviation of Pixel Values

Due to multiple frames participating in a dissolve, the standard deviation of the pixel intensity values decrease. The same has been discussed in section 5.4. The standard deviation of the pixel intensity values for each frame in the video is stored in a row matrix $\left(M_{STD}\right)$. The mean $\left(\mu_{STD}\right)$ and standard deviation $\left(\sigma_{STD}\right)$ of the values in $\left(M_{STD}\right)$ is computed for determining the threshold. The threshold taken in this stage of filtration is $\left(\mu_{STD} - \sigma_{STD}\right)$. The candidates of dissolve transition filtered from the earlier stage are fed into this stage for further filtration. A candidate for dissolve transition is eliminated if no frame of the candidate has standard deviation of the pixel intensity values lower than or equal to $\left(\mu_{STD} - \sigma_{STD}\right)$.

7.2.4 Filtration and Elimination Using Edge Change Ratio

The Edge Change Ratio (ECR) is an important indicator for dissolve sequences as explained in section 5.2. During a gradual transition sequence, the ECR plot shows an abrupt change. The mean $\left(\mu_{ECR}\right)$ and standard deviation $\left(\sigma_{ECR}\right)$ of the ECR for successive frames of the video are computed. The dissolve candidates filtered from the earlier stage are tested against a threshold $(\mu_{ECR} + \sigma_{ECR})$. If the ECR value for any two successive frames of a candidate dissolve is more than the threshold, it is retained in the candidate list, else eliminated. At the end of this stage, the final set of dissolve sequences is obtained. The final set of dissolve sequences are such that all members in it have passed the threshold levels set for the different parameters.

8. EXPERIMENTAL RESULTS AND ANALYSIS

The proposed method for gradual transition detection was tested on a video data set consisting of eight videos. The source code for the proposed method was written using MATLAB. The characteristics of the data set are given in Table I. All the videos in the dataset are in MP4 file format (ISO/IEC 14496-14:2003), commonly named as MPEG-4 file format version 2. For evaluating the performance of the proposed method, recall (*R*) and precision (*P*) are computed as follows:

$$\text{Recall} = (D_d - D_f) / D_t$$

$$\text{Precision} = (D_d - D_f) / D_d$$

where,

D_d : Dissolve sequences detected by algorithm.
D_f : False dissolve sequences detected.
D_t : Actual dissolve sequences present in the video.

The above two parameters are also taken to evaluate the performance of the proposed method to the existing methods.

8.1 The Video Dataset

Out of the eight videos in the dataset, V1 is a highlights video of a Wimbledon tennis match. The videos labeled V2-V6 are Hindi movie songs. The dataset also contains two documentary videos V7 and V8. Originally these videos contained mostly abrupt transitions (hard cuts). To test the robustness of the proposed method, dissolve sequences of various lengths were introduced into the videos at places where hard cuts originally existed. The details of the dataset are given in Table 1.

8.2 Experimental Results

The proposed method for dissolve detection is a two-phased method in which the first phase is used for parabolic pattern detection. The candidates detected in the first phase are filtered out based on threshold fixed for different parameters. The number of candidates detected in the first stage and filtered in the next stage is enumerated in Table 2. Further, Table 3 shows the details of precision and recall values obtained by running the algorithm on the videos of the test set. The span of the dissolve detected by the algorithm is within an error range of one frame on either side of the dissolve span.

8.3 Comparison with Other Existing Methods

The proposed method for dissolve detection was compared with three other existing methods like Edge Change Ratio (ECR) (Qi et al., 2010), Entropy and Local Descriptor (ELD) (Hanjalic,2002) and Edge-

Table 1. Test video dataset

Video	Length (mm:ss)	Resolution	FPS	No. of Frames	No. of Dissolves Sequences
V1	02:58	640 × 360	25	4468	43
V2	02:42	640 × 360	25	4057	70
V3	04:10	640 × 360	25	6265	172
V4	03:27	640 × 360	25	4965	77
V5	03:31	640 × 360	25	5053	138
V6	05:58	1280 × 544	24	8602	83
V7	51:20	640 × 360	25	74020	941
V8	44:14	480 × 360	25	66339	626

Table 2. Results on test video dataset

Video	Actual Dissolves Sequences	Parabolic Patterns Detected (1st Phase)	Dissolves Detected after Filtration Phase	Correct Detections	False Alarms
V1	43	666	46	43	3
V2	70	1449	76	68	8
V3	172	2751	177	170	7
V4	77	2154	81	75	6
V5	138	1922	135	133	2
V6	83	2573	89	82	7
V7	941	1876	952	939	13
V8	626	11146	633	625	8

Table 3. Recall and precision

Video	Actual Dissolves Sequences	Correct Detections	False Alarms	Recall	Precision
V1	43	43	3	1.0	0.934
V2	70	68	8	0.971	0.894
V3	172	170	7	0.988	0.960
V4	77	75	6	0.974	0.926
V5	138	133	2	0.963	0.985
V6	83	82	7	0.988	0.921
V7	941	939	13	0.997	0.986
V8	626	625	8	0.998	0.987

Table 4. Comparison with other methods

	ECR		ELD		EC		Proposed Method	
Video	**R**	**P**	**R**	**P**	**R**	**P**	**R**	**P**
V1	0.71	0.55	0.81	0.76	0.58	0.43	**1.0**	**0.93**
V2	0.54	0.63	0.76	0.64	0.62	0.56	**0.97**	**0.89**
V3	0.75	0.81	0.89	0.80	0.79	0.68	**0.98**	**0.96**
V4	0.44	0.41	0.92	0.87	0.46	0.32	**0.97**	**0.92**
V5	0.69	0.87	0.85	0.79	0.53	0.70	**0.96**	**0.98**
V6	0.59	0.51	0.94	0.91	0.62	0.46	**0.98**	**0.92**
V7	0.62	0.73	0.72	0.62	0.68	0.61	**0.99**	**0.98**
V8	0.56	0.76	0.81	0.77	0.70	0.55	**0.99**	**0.98**

based Contrast (EC) (Miene *et al*., 2003). For each of the existing methods, the recall and precision values are computed for videos in the dataset. The proposed method is seen to outperform all the existing methods compared. A summary of the comparison is shown in Table 4.

9. CONCLUSION

Detection of dissolve transitions involve a lot of complexity because of the inherent structure of these edit effects. Although the method discussed here was seen to work well for detection of dissolves involving two shots, it is yet to be tested on transitions involving more than two shots. The method was seen to outperform traditional methods used for dissolve detection. As this method uses an ensemble of low-level features, the time taken for execution is relatively high. The method in its present form is not suitable for application in real-time systems. Detection of gradual transitions still remains an area of research as the methods developed do not show consistent performance for all types of video. As detection of edit effects has a great impact on the overall advances in the domain of video analysis, therefore research in this field will continue for some time to come.

REFERENCES

Abdelali, A. B., Nidhal Krifa, M., Touil, L., Mtibaa, A., & Bourennane, E. (2009). A study of the color-structure descriptor for shot boundary detection. *International Journal of Sciences and Techniques of Automatic control and computer engineering*, 956-971.

Abdeljaoued, Y., Ebrahimi, T., Christopoulos, C., & Ivars, I. M. (2000, September). A new algorithm for shot boundary detection.*Proceedings of the 10th European Signal Processing Conference* (pp. 151-154). SPIE.

Apostolidis, E., & Mezaris, V. (2014, May). Fast shot segmentation combining global and local visual descriptors. *Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6583-6587). IEEE. doi:10.1109/ICASSP.2014.6854873

Baber, J., Afzulpurkar, N., Dailey, M. N., & Bakhtyar, M. (2011, July). Shot boundary detection from videos using entropy and local descriptor. *Proceedings of the 2011 17th International Conference on Digital Signal Processing (DSP)* (pp. 1-6). IEEE. doi:10.1109/ICDSP.2011.6004918

Bai, L., Lao, S. Y., Liu, H. T., & Bu, J. (2008, July). Video shot boundary detection using petri-net. *Proceedings of the 2008 International Conference on Machine Learning and Cybernetics* (Vol. 5, pp. 3047-3051). IEEE.

Bao, Z., Han, B., & Wu, S. (2006). A novel clustering algorithm based on variable precision rough-fuzzy sets. In *Computational Intelligence* (pp. 284–289). Springer Berlin Heidelberg. doi:10.1007/978-3-540-37275-2_36

Barbu, T. (2009). Novel automatic video cut detection technique using Gabor filtering. *Computers & Electrical Engineering*, *35*(5), 712–721. doi:10.1016/j.compeleceng.2009.02.003

Barron, J. L., Fleet, D. J., & Beauchemin, S. S. (1994). Performance of optical flow techniques. *International Journal of Computer Vision*, *12*(1), 43–77. doi:10.1007/BF01420984

Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, *110*(3), 346–359. doi:10.1016/j.cviu.2007.09.014

Bescós, J., Cisneros, G., Martínez, J. M., Menéndez, J. M., & Cabrera, J. (2005). A unified model for techniques on video-shot transition detection. *IEEE Transactions on* Multimedia, *7*(2), 293–307.

Bhattacharyya, S., Maulik, U., & Dutta, P. (2009). High-speed target tracking by fuzzy hostility-induced segmentation of optical flow field. *Applied Soft Computing*, *9*(1), 126–134. doi:10.1016/j.asoc.2008.03.012

Bhaumik, H., Bhattacharyya, S., & Chakraborty, S. (2014, April). Video Shot Segmentation Using Spatio-Temporal Fuzzy Hostility Index and Automatic Threshold. *Proceedings of the 2014 Fourth International Conference on Communication Systems and Network Technologies (CSNT)* (pp. 501-506). IEEE. doi:10.1109/CSNT.2014.106

Boccignone, G., Chianese, A., Moscato, V., & Picariello, A. (2005). Foveated shot detection for video segmentation. *IEEE Transactions on* Circuits and Systems for Video Technology, *15*(3), 365–377.

Boreczky, J. S., & Wilcox, L. D. (1998, May). A hidden Markov model framework for video segmentation using audio and image features. *Proceedings of the 1998 IEEE International Conference onAcoustics, Speech and Signal Processing* (*Vol. 6*, pp. 3741-3744). IEEE. doi:10.1109/ICASSP.1998.679697

Calonder, M., Lepetit, V., Strecha, C., & Fua, P. (2010). Brief: Binary robust independent elementary features. *Computer Vision–ECCV*, *2010*, 778–792.

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6), 679-698.

Černeková, Z., Pitas, I., & Nikou, C. (2006). Information theory-based shot cut/fade detection and video summarization. *IEEE Transactions on* Circuits and Systems for Video Technology, *16*(1), 82–91.

Chan, C., & Alexander, W. (2011). Shot boundary detection using genetic algorithm optimization. *Proceedings of the 2011 IEEE International Symposium on Multimedia (ISM)* (pp. 327-332). IEEE. doi:10.1109/ISM.2011.58

Chasanis, V., Likas, A., & Galatsanos, N. (2009). Simultaneous detection of abrupt cuts and dissolves in videos using support vector machines. *Pattern Recognition Letters*, *30*(1), 55–65. doi:10.1016/j.patrec.2008.08.015

Chavez, G. C., Precioso, F., Cord, M., Philipp-Foliguet, S., & Araujo, A. D. A. (2006). Shot boundary detection at TRECVID 2006.*Proc. TREC Video Retrieval Evaluation*.

Chiu, P., Girgensohn, A., Polak, W., Rieffel, E., & Wilcox, L. (2000). A genetic algorithm for video segmentation and summarization. *Proceedings of the 2000 IEEE International Conference on Multimedia and Expo ICME '00* (Vol. 3, pp. 1329-1332). IEEE. doi:10.1109/ICME.2000.871011

Donate, A., & Liu, X. (2010, June). Shot boundary detection in videos using robust three-dimensional tracking. *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 64-69). IEEE. doi:10.1109/CVPRW.2010.5543811

Drew, M. S., Li, Z. N., & Zhong, X. (2000). Video dissolve and wipe detection via spatio-temporal images of chromatic histogram differences. *Proceedings of the 2000 International Conference on Image Processing* (Vol. 3, pp. 929-932). IEEE. doi:10.1109/ICIP.2000.899609

Fang, H., Jiang, J., & Feng, Y. (2006). A fuzzy logic approach for detection of video shot boundaries. *Pattern Recognition*, *39*(11), 2092–2100. doi:10.1016/j.patcog.2006.04.044

Gao, X., & Tang, X. (2002). Unsupervised video-shot segmentation and model-free anchorperson detection for news video story parsing. *IEEE Transactions on* Circuits and Systems for Video Technology, *12*(9), 765–776.

Gao, X. B., Han, B., & Ji, H. B. (2005). A shot boundary detection method for news video based on rough sets and fuzzy clustering. In *Image Analysis and Recognition* (pp. 231–238). Springer Berlin Heidelberg.

Gargi, U., Kasturi, R., & Strayer, S. H. (2000). Performance characterization of video-shot-change detection methods. *IEEE Transactions on* Circuits and Systems for Video Technology, *10*(1), 1–13.

Geetha, M. K., & Palanivel, S. (2012). Unsupervised Approach for Retrieving Shots from Video. *International Journal of Computers and Applications*, *60*(6).

Grana, C., & Cucchiara, R. (2007). Linear transition detection as a unified shot detection approach. *IEEE Transactions on Circuits and Systems for Video Technology*, *17*(4), 483–489. doi:10.1109/TCSVT.2006.888818

Hanjalic, A. (2002). Shot-boundary detection: Unraveled and resolved? *IEEE Transactions on* Circuits and Systems for Video Technology, *12*(2), 90–105.

Hsu, C. F., Ku, M. K., & Liu, L. Y. (2009, September). Support vector machine FPGA implementation for video shot boundary detection application. *Proceedings of theIEEE InternationalSOC Conference SOCC '09* (pp. 239-242). IEEE. doi:10.1109/SOCCON.2009.5398049

Hu, W., Xie, N., Li, L., Zeng, X., & Maybank, S. (2011). A survey on visual content-based video indexing and retrieval. *IEEE Transactions on* Systems, Man, and Cybernetics, Part C: Applications and Reviews, *41*(6), 797–819.

Ionescu, B., Vertan, C., & Lambert, P. (2011, May). Dissolve detection in abstract video contents. *Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 917-920). IEEE. doi:10.1109/ICASSP.2011.5946554

Jadon, R. S., Chaudhury, S., & Biswas, K. K. (2001). A fuzzy theoretic approach for video segmentation using syntactic features. *Pattern Recognition Letters*, *22*(13), 1359–1369. doi:10.1016/S0167-8655(01)00041-1

Ji, Q. G., Feng, J. W., Zhao, J., & Lu, Z. M. (2010, September). Effective dissolve detection based on accumulating histogram difference and the support point. *Proceedings of the 2010 First International Conference on Pervasive Computing Signal Processing and Applications (PCSPA)* (pp. 273-276). IEEE. doi:10.1109/PCSPA.2010.73

Joyce, R. A., & Liu, B. (2006). Temporal segmentation of video using frame and histogram space. *IEEE Transactions on* Multimedia, *8*(1), 130–140.

Kawai, Y., Sumiyoshi, H., & Yagi, N. (2007, November). Shot Boundary Detection at TRECVID 2007. Proceedings of TRECVID ‘07.

Koprinska, I., & Carrato, S. (2001). Temporal video segmentation: A survey. *Signal Processing Image Communication*, *16*(5), 477–500. doi:10.1016/S0923-5965(00)00011-4

Koumousis, K. I., Fotopoulos, V., & Skodras, A. N. (2012, October). A new approach to gradual video transition detection. *Proceedings of the 2012 16th Panhellenic Conference on Informatics* (pp. 245-249). IEEE. doi:10.1109/PCi.2012.85

Küçüktunç, O., Güdükbay, U., & Ulusoy, Ö. (2010). Fuzzy color histogram-based video segmentation. *Computer Vision and Image Understanding*, *114*(1), 125–134. doi:10.1016/j.cviu.2009.09.008

Kundu, M. K., & Mondal, J. (2012, December). A novel technique for automatic abrupt shot transition detection. *Proceedings of the 2012 International Conference on Communications, Devices and Intelligent Systems (CODIS)* (pp. 628-631). IEEE. doi:10.1109/CODIS.2012.6422281

Lefèvre, S., Holler, J., & Vincent, N. (2003). A review of real-time segmentation of uncompressed video sequences for content-based search and retrieval. *Real-Time Imaging*, *9*(1), 73–98. doi:10.1016/S1077-2014(02)00115-8

Li, B., & Sezan, M. I. (2001). Event detection and summarization in sports video. *Proceedings of the IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL ‘01)* (pp. 132-138). IEEE. doi:10.1109/IVL.2001.990867

Li, F., Kim, T., Humayun, A., Tsai, D., & Rehg, J. (2013). Video segmentation by tracking many figure-ground segments. *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2192-2199). doi:10.1109/ICCV.2013.273

Li, J., Ding, Y., Shi, Y., & Zeng, Q. (2009, August). DWT-based shot boundary detection using support vector machine. *Proceedings of the Fifth International Conference on Information Assurance and Security IAS'09* (Vol. 1, pp. 435-438). IEEE. doi:10.1109/IAS.2009.16

Lienhart, R. W. (1999). Comparison of automatic shot boundary detection algorithms. *Proc. of SPIE Conf. Image and Video Processing VII*, (pp. 290-301).

Lienhart, R. W. (2001, January). Reliable dissolve detection. In Photonics West 2001-Electronic Imaging (pp. 219-230). International Society for Optics and Photonics.

Ling, X., Yuanxin, O., Huan, L., & Zhang, X. (2008, May). A method for fast shot boundary detection based on SVM. Proceedings of the Congress on Image and Signal Processing CISP'08 (Vol. 2, pp. 445-449). IEEE. doi:10.1109/CISP.2008.605

Liu, S., Zhu, M., & Zheng, Q. (2008, October). Video shot boundary detection with local feature post refinement. *Proceedings of the 9th International Conference on Signal Processing ICSP '08* (pp. 1548-1551). IEEE.

Liu, T. Y., Lo, K. T., Zhang, X. D., & Feng, J. (2004). A new cut detection algorithm with constant false-alarm ratio for video segmentation. *Journal of Visual Communication and Image Representation*, *15*(2), 132–144. doi:10.1016/j.jvcir.2003.10.001

Liu, X., & Chen, T. (2002, May). Shot boundary detection using temporal statistics modeling. *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)* (Vol. 4, pp. IV-3389). IEEE. doi:10.1109/ICASSP.2002.5745381

Lowe, D. G. (1999). Object recognition from local scale-invariant features. *Proceedings of the seventh IEEE international conference on Computer vision* (Vol. 2, pp. 1150-1157). IEEE. doi:10.1109/ICCV.1999.790410

Lu, Z. M., & Shi, Y. (2013). Fast video shot boundary detection based on SVD and pattern matching. *IEEE Transactions on* Image Processing, *22*(12), 5136–5145.

Ma, C., Yu, J., & Huang, B. (2012). A rapid and robust method for shot boundary detection and classification in uncompressed MPEG video sequences. *Int. J. Comput. Sci. Issues, 5*, 368-374.

Mahesh, K., & Kuppusamy, K. (2012). A New Hybrid Algorithm for Video Segmentation. In *Advances in Computer Science, Engineering & Applications* (pp. 587–595). Springer Berlin Heidelberg. doi:10.1007/978-3-642-30157-5_59

Manjunath, S., Guru, D. S., Suraj, M. G., & Harish, B. S. (2011, March). A non parametric shot boundary detection: an eigen gap based approach.*Proceedings of the Fourth Annual ACM Bangalore Conference* (p. 14). ACM. doi:10.1145/1980422.1980436

Meng, Y., Wang, L. G., & Mao, L. Z. (2009, July). A shot boundary detection algorithm based on particle swarm optimization classifier. *Proceedings of the 2009 International Conference on Machine Learning and Cybernetics,* (Vol. 3, pp. 1671-1676). IEEE. doi:10.1109/ICMLC.2009.5212297

Mich, O., Brunelli, R., & Modena, C. M. (1999). A survey on the automatic indexing of video data. *Journal of Visual Communication and Image Representation*, *10*(2), 78–112. doi:10.1006/jvci.1997.0404

Miene, A., Hermes, T., Ioannidis, G. T., & Herzog, O. (2003, November). Automatic shot boundary detection using adaptive thresholds.*Proc. TRECVID Workshop* (pp. 1-7).

Mittalkod, S. P., & Srinivasan, G. N. (2011). Shot Boundary Detection Algorithms and Techniques: A Review. *International Journal of Computer System Engineering*.

Nam, J., & Tewfik, A. H. (2005). Detection of gradual transitions in video sequences using b-spline interpolation. *Multimedia*. *IEEE Transactions on*, *7*(4), 667–679.

Naranjo, V., Angulo, J., Albiol, A., Mossi, J. M., Albiol, A., & Gomez, S. (2007). Gradual transition detection for video partitioning using morphological operators. *Image Analysis & Stereology*, *26*(2), 51–61. doi:10.5566/ias.v26.p51-61

Ngo, C.-W. (2003). A robust dissolve detector by support vector machine.*Proceedings of the eleventh ACM international conference on Multimedia*. ACM. doi:10.1145/957013.957072

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*(3), 145–175. doi:10.1023/A:1011139631724

Patel, B. V., & Meshram, B. B. (2012). *Content based video retrieval systems.* arXiv preprint arXiv:1205.1641

Patel, U., Shah, P., & Panchal, P. (2013). Shot Detection Using Pixel wise Difference with Adaptive Threshold and Color Histogram Method in Compressed and Uncompressed Video. *International Journal of Computers and Applications*, *64*(4).

Poleg, Y., Arora, C., & Peleg, S. (2014). Temporal segmentation of egocentric videos.*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2537-2544).

Porter, S. V., Mirmehdi, M., & Thomas, B. T. (2001, September). *Detection and Classification of Shot Transitions*. BMVC.

Priya, G., & Domnic, S. (2012). Transition detection using Hilbert transform and texture features. *American J. of Signal Proc.*, *10*, 35–40. doi:10.5923/j.ajsp.20120202.06

Qi, Y., Hauptmann, A., & Liu, T. (2003, July). Supervised classification for video shot segmentation. *Proceedings of the 2003 International Conference on Multimedia and Expo ICME'03* (Vol. 2, pp. II-689). IEEE.

Qin, T., Gu, J., Chen, H., & Tang, Z. (2010, September). A fast shot-boundary detection based on k-step slipped window. *Proceedings of the 2010 2nd IEEE International Conference on Network Infrastructure and Digital Content* (pp. 190-195). IEEE. doi:10.1109/ICNIDC.2010.5657841

Rao, P. C., & Patnaik, M. R. (2014). Contourlet Transform Based Shot Boundary Detection. *International Journal of Signal Processing*. *Image Processing and Pattern Recognition*, *7*(4), 381–388. doi:10.14257/ijsip.2014.7.4.36

Rong, J., Ma, Y. F., & Wu, L. (2005, January). Gradual transition detection using em curve fitting. *Proceedings of the 11th InternationalMultimedia Modelling Conference MMM '05*(pp. 364-369). IEEE.

Saini, S., & Gupta, P. (2015). Video Shot Boundary Detection Using Various Techniques. *International Journal of Emerging Technologies and Innovative Research*, *2*(4), 1109–1115.

Sao, N., & Mishra, R. (2014). A survey based on video shot boundary detection techniques. *International Journal of Advanced Research in Computer and Communication Engineering*, *3*(4).

Šarić, M., Dujmić, H., & Baričević, D. (2008). Shot boundary detection in soccer video using twin-comparison algorithm and dominant color region. *Journal of Information and Organizational Sciences*, *32*(1), 67–73.

Shen, S., & Cao, J. (2011, March). Abrupt shot boundary detection algorithm based on fuzzy clustering neural network. *Proceedings of the 2011 3rd International Conference on Computer Research and Development*.

Shirahama, K., Matsuoka, Y., & Uehara, K. (2012). Event retrieval in video archives using rough set theory and partially supervised learning. *Multimedia Tools and Applications*, *57*(1), 145–173. doi:10.1007/s11042-011-0727-z

Song, S. M., Kwon, T. H., Kim, W. M., Kim, H., & Rhee, B. D. (1997, December). Detection of gradual scene changes for parsing of video data. In Photonics West'98 Electronic Imaging (pp. 404-413). International Society for Optics and Photonics.

Sowmya, R., & Shettar, R. (2013). Analysis and verification of video summarization using shot boundary detection. *Am Int J Res Sci Technol Eng Math*, *3*(1), 82–86.

Su, C. W., Liao, H. Y. M., Tyan, H. R., Fan, K. C., & Chen, L. H. (2005). A motion-tolerant dissolve detection algorithm. *IEEE Transactions on* Multimedia, *7*(6), 1106–1113.

Sun, X., Zhang, Y., Hao, X., & Min, W. (2014). Shot Boundary Detection Based on SVM Optimization Model. *Open Automation and Control Systems Journal*, *6*(1), 393–397. doi:10.2174/1874444301406010393

Sun, X., Zhao, L., & Zhang, M. (2011, August). A Novel Shot Boundary Detection Method Based on Genetic Algorithm-Support Vector Machine. *Proceedings of the 2011 International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)* (Vol. 1, pp. 144-147). IEEE. doi:10.1109/IHMSC.2011.41

Tariq, A., Flail, N., & Ghazi, A. (2014). Using Daub achy Wavelet for Shot Boundary Detection. *IOSR Journals*, *1*(16), 66–70.

Thakare, S. (2012). Intelligent processing and analysis of image for shot boundary detection. *International Journal of Emerging Technology and Advanced Engineering*, *2*(2), 208–212.

Thakre, K. S. (2014). Analysis and Review of Formal Approaches to Automatic Video Shot Boundary Detection. *Analysis*, *3*(1).

Tola, E., Lepetit, V., & Fua, P. (2010). Daisy: An efficient dense descriptor applied to wide-baseline stereo. *IEEE Transactions on* Pattern Analysis and Machine Intelligence, *32*(5), 815–830.

Truong, B. T., Dorai, C., & Venkatesh, S. (2000, October). New enhancements to cut, fade, and dissolve detection processes in video segmentation.*Proceedings of the eighth ACM international conference on Multimedia* (pp. 219-227). ACM. doi:10.1145/354384.354481

Tsamoura, E., Mezaris, V., & Kompatsiaris, I. (2008a, October). Gradual transition detection using color coherence and other criteria in a video shot meta-segmentation framework. *Proceedings of the 15th IEEE International Conference on Image Processing ICIP '08* (pp. 45-48). IEEE. doi:10.1109/ICIP.2008.4711687

Tsamoura, E., Mezaris, V., & Kompatsiaris, I. (2008b, June). Video shot meta-segmentation based on multiple criteria for gradual transition detection. *Proceedings of the International Workshop on Content-Based Multimedia Indexing CBMI '08* (pp. 51-57). IEEE. doi:10.1109/CBMI.2008.4564927

Volkmer, T., Tahaghoghi, S. M., & Williams, H. E. (2004, June). Gradual transition detection using average frame similarity. *Proceedings of theConference onComputer Vision and Pattern Recognition Workshop CVPRW'04* (pp. 139-139). IEEE. doi:10.1109/CVPR.2004.357

Wang, T., Wu, Y., & Chen, L. (2007, April). An approach to video key-frame extraction based on rough set. *Proceedings of the International Conference on Multimedia and Ubiquitous Engineering MUE'07* (pp. 590-596). IEEE. doi:10.1109/MUE.2007.65

Xiong, W., Lee, C. M., & Ma, R. H. (1997). Automatic video data structuring through shot partitioning and key-frame computing. *Machine Vision and Applications*, *10*(2), 51–65. doi:10.1007/s001380050059

Yasira Beevi, C. P., & Natarajan, S. (2009). *An efficient video segmentation algorithm with real time adaptive threshold technique*. Citeseer.

Yoo, H. W., Ryoo, H. J., & Jang, D. S. (2006). Gradual shot boundary detection using localized edge blocks. *Multimedia Tools and Applications*, *28*(3), 283–300. doi:10.1007/s11042-006-7715-8

Yu, F., Lu, Z., & Li, Y. (2011). Dissolve detection based on twin-comparison with curve fitting. *International Journal of Innovative Computing, Information, & Control*, *7*, 2417–2426.

Yuan, J., Li, J., & Zhang, B. (2007, September). Gradual transition detection with conditional random fields.*Proceedings of the 15th international conference on Multimedia* (pp. 277-280). ACM. doi:10.1145/1291233.1291291

Yuan, J., Wang, H., Xiao, L., Zheng, W., Li, J., Lin, F., & Zhang, B. (2007). A formal study of shot boundary detection. *IEEE Transactions on* Circuits and Systems for Video Technology, *17*(2), 168–186.

Zabih, R., Miller, J., & Mai, K. (1999). A feature-based algorithm for detecting and classifying production effects. *Multimedia Systems*, *7*(2), 119–128. doi:10.1007/s005300050115

Zajić, G. J., Reljin, I. S., & Reljin, B. D. (2011). Video shot boundary detection based on multifractal analysis. *Telfor Journal*, *3*(2), 105–110.