Computing and Informatics, Vol. 30, 2011, 595–619

VIDEO SHOT BOUNDARY DETECTION USING GENERALIZED EIGENVALUE DECOMPOSITION AND GAUSSIAN TRANSITION DETECTION

Ali Amiri

Computer Engineering Group, Engineering Department
Zanjan University, Zanjan, Iran
e-mail: a_amiri@znu.ac.ir

Mahmood Fathy

Computer Engineering Department
Iran University of Science and Technology (IUST), Tehran, Iran
e-mail: mahfathy@iust.ac.ir

Communicated by François Brémond

Abstract. Shot boundary detection is the first step of the video analysis, summarization and retrieval. In this paper, we propose a novel shot boundary detection algorithm using Generalized Eigenvalue Decomposition (GED) and modeling of gradual transitions by Gaussian functions. Especially, we focus on the challenges of detecting the gradual shots and extracting appropriate spatio-temporal features, which have effects on the ability of algorithm to detect shot boundaries efficiently. We derive a theorem that discuss about some new features of GED which could be used in the video processing algorithms. Our innovative explanation utilizes this theorem in the defining of new distance metric in Eigen space for comparing video frames. The distance function has abrupt changes in hard cut transitions and semi-Gaussian behavior in gradual transitions. The algorithm detects the transitions by analyzing this distance function. Finally we report the experimental results using large-scale test sets provided by the TRECVID 2006 which has evaluations for hard cut and gradual shot boundary detection.

Keywords: Radić's determinant, Generalized Eigenvalue Decomposition (GED), shot boundary detection, Gaussian transition detection

1 INTRODUCTION

The latest developments in multimedia technology, combined with a significant increase in computer performance and the growth of the Internet, have given people access to a large amount of video information. Video applications, which are considerably growing, have initiated the growing demand for innovative technologies and tools for efficient indexing, browsing, and retrieval of video data.

In order to automate the indexing, retrieval, and management of videos, a great deal of research has been done on content-based video retrieval over the last decade [1, 2]. Structural analysis of videos is a basic step in analyzing video content and developing techniques for efficient access, classification, retrieval, and browsing of large video databases. Amongst the various structural levels (i.e., frame, shot, scene, etc.), shot level organization has been regarded suitable for browsing and content-based retrieval [3].

A video shot is defined as a sequence of frames captured by one camera in a single continuous action in time and space [4]. Generally, it is a group of frames that have consistent visual characteristics, such as color, texture, and motion. Depending on whether the transition between shots is abrupt or gradual, shot boundaries can be classified into two types: cut and gradual transitions. The cut transition is the classic, abrupt change where one frame belongs to the disappearing shot and the next one to the appearing shot.

Gradual transitions can be categorized into dissolve, wipe, fade out/in, etc., in accordance with the attributes of different editing effects [5]. For the dissolve transition, the last few frames of the disappearing shot temporally overlap with the first few frames of the appearing shot. During the overlap transition, the intensity of the disappearing shot reduces from normal to zero (fade out), whereas the intensity of the appearing shot increases from zero to normal (fade in). In the fade transition, the disappearing shot fades out into a blank frame, afterwards the blank frame fades into the appearing shot. The wipe transition is in fact a set of shot change methods where the appearing and disappearing shots coexist in different spatial areas of the intermediary video frames, and the area taken by the former expands until it completely replaces the latter. The main focus of many recent works has been on abrupt transitions; gradual transitions are usually harder to identify, because of camera and/or object motions in a shot.

Shot detection has long been on focus in the content-based video analysis community. Over the past decade, shot boundary detection (SBD) has been seriously studied in video retrieval, video summarization, pattern recognition, and multimedia communities [6]. In recent years, more research has been done on automatic shot detection. Since 2001, the TREC Video Retrieval (TRECVID) evaluation testbed has been created to carry out benchmark evaluations of video shot detection tasks [7], and has notably improved the development of SBD techniques. It shows that the detection of abrupt transitions has been tackled quite successfully, whereas the identification of gradual transitions is still a challenging problem [7].

In spite of the latest advances, SBD on large-scale video data is still a very difficult task, with many problematic issues. For example, the problem of how to devise an efficient, unified technique which identifies different types of transitions and is less sensitive to the amount of camera panning and zooming, video object motions, color, and illumination variability in the shot. For this purpose, we suggest a Generalized Eigenvalue Decomposition (GED)-based technique which is not dependent on the content of data and will be utilized to identify various types of transitions, maximize the efficiency of SBD, and decrease the computational cost. We have executed our resolution and assessed it in accordance with the TRECVID benchmark dataset. Our method produced hopeful results, compared with the best results reported in the TRECVID assessments.

The rest of this paper is organized as follows. Section 2 deals with some modern work regarding SBD in past few years. Section 3 addresses the difficulties and challenges regarding shot transition detection and indicates the motivation behind our proposed solution. In Section 4, GED is briefly described. In Section 5, our SBD approach is presented. Section 6 presents experimental assessments of our methods and executions on video SBD tasks in the TRECVID 2006 testbed. Section 7 explains our conclusions and talks about future ideas that will enhance the efficiency of our present solution.

2 RELATED WORKS

SBD, which is also referred to as temporal video segmentation, is the process of detecting the transitions between the adjoining shots [8]. Recently, SBD has been used as a more effective element for all the video retrieval and video summarization systems. In this section, we analyze some of the works which are being done about video shot transition identification and explain some of the common ranking methods for shot identification, particularly for those works that are connected with our suggested method. Starting in the early 1990s, a number of institutions had already begun projects such as QBIC [9], Columbia VideoQ (object-oriented search engine) [10], and the Virage [11], that have connected to digital video libraries in order to handle video content intelligently. During that period, researchers concentrated more on video processing, such as SBD, video retrieval, video object detection, and video summarization.

Until now, numerous methods have been suggested for the identification of shot boundaries, and have produced highly acceptable results. After studying the literature concerning these techniques and methods, we realized that these methods could be classified into two categories: compressed domain methods and uncompressed domain methods. In compressed domain methods, the only data extracted from the videos is the data which is directly accessible from the MPEG streams that are discrete cosine transform (DCT) coefficients, motion vectors, and prediction directions for each block. The absence of the decoding process allows for a much faster computation, but less reliable, especially in the presence of high motion. Pei and Chou [42] suggested a SBD technique in MPEG compressed domain through macroblock type data. In [17], Jiang et al. have suggested an SBD technique on the basis of the DC histogram Chi-square test and macroblock type statistics in one frame of MPEG-2 video sequence. In their algorithm, it is unnecessary to decompress the MPEG-2 video sequence; as a result, the computation is much faster and less complex. The authors of [18] have devised a shot segmentation technique based on grid-mapping dynamic windows in compressed videos. Wang et al. [20] devised a SBD technique for MPEG video based on DC image, which just requires partial decompression from the encode bit-stream and does not require all the histogram comparison between adjoining frames, thus it can make the computation less complex. In addition, the authors of [24] have devised an SBD technique based on macroblock and DC image, which the statistics of macroblock type data is used to coarse detection, and DC images are used to scrutiny detection. In [22], Yang et al. suggested an MPEG-7 descriptor based SBD algorithm for H.264/AVC with a less complex computation.

Uncompressed domain techniques were the main focus in early research on shot detection. These techniques are compared in [40] and [2]. A number of these techniques have been put forward to be used in the identification of abrupt transitions. In some of these methods, an abrupt transition is identified when a specific difference measure between successive frames surpasses a threshold. The difference measure is calculated at either the pixel level or the block level. Considering the limitation of pixel difference algorithms (being highly sensitive to object and camera motions), a large number of researchers recommended using other measures that were based on global information, such as intensity histograms or color histograms [28–31]. At present, the standard color histogram-based algorithm and its variations are greatly used to identify abrupt transitions.

Even these histograms do not precisely exhibit the image difference that large camera motion causes, and therefore, are unable to distinguish between smooth camera motion/parameter changes and gradual scene transitions. Although using more complex characteristics, such as image edges, histograms, or motion vectors [32] ameliorates the situation, it will alleviate but not resolve this issue [14]. In [21], Li et al. put forward a solution to this problem by extracting the color and the edge in different direction from wavelet transition coefficients. They have used a multi-class support vector machine (SVM) classifier that classifies the video shot into three categories: cut transition, gradual transition and normal sequences. Furthermore, the authors of [15] proposed an SBD technique that utilizes global color feature together with the feature of local edge. They eliminated the flash intrusion with the gray value of edge of object, and formulated a gradient detector via edge features based on computing the inter-frame similarity compatible with eigenvalues of the global color. In [23], Wu et al. used both temporal and spatial saliencies that are supposed to produce an information saliency map. They identified the shot transition based on the change of saliency. The authors of [4] have devised a solution to this problem via measuring the changes of data between adjoining images, which are quantized by mutual information (MI) in gray-scale regions of the images. They have also utilized affine image registration to compensate camera panning and zooming. This results in a highly computationally complex approach. An effective method has been suggested in [6], which is on the basis of measures of information theory. The downside of this approach is that it is susceptible to large camera panning and zooming and also flashlights. Also, in [27] Bai et al. proposed a SBD technique that measures the inter-frame dissimilarities via reciprocal information computed separately for each of the HSV color components. They proposed a Petri-net model to describe the combination of boundary frames which demonstrates a shot boundary and utilized for video frames sequence to identify both cut transitions and gradual transitions.

Therefore, the principal problem we have in gradual transition identification is that the comparison that is on the basis of spatial characteristics such as color histogram, edge, and motion vectors without modeling of the temporal link between frames is not appropriate. To solve this problem, many methods explore large windows of frames. However, these approaches are not trifling because the variation between two distinct shots can be mistaken for the object motion variation inside those shots.

Although there have been numerous research attempts, other machine learning and pattern recognition tools have never received enough attention. In [34], Vasconcelos et al. devised a Bayesian formulation for this issue and expanded the standard thresholding model intuitively and adaptively. In [35], Lienhart identifies a number of principal methods that are the foundation of the various SBD techniques and assesses their roles in identifying abrupt cuts, fades, and dissolves. In [36], Ling et al. use specific characteristics, such as intensity pixel-wise difference, color histograms in HSV space, and edge histograms in vertical and horizontal directions as the input vectors to the support vector machine (SVM). The SVM is utilized in order to categorize the frames into four classifications: abrupt, dissolve, fade, and wipe transitions. The authors of [25] suggested a hierarchical learning-based approach for SBD, which is on the basis of kernel-based SVM. In order to make SBD more precise, the authors of [19] have suggested an altered edition of the kernel function of SVM, and have presented several evaluations to compare with other kernel functions. In [16] Hameed proposed a framework for SBD through using different statistical characteristics. After training the framework, some appropriate characteristics are calculated based on these characteristics to detect the shot transitions. An unsupervised method is suggested in [26] for anchor shot identification by using spectral clustering and varied information from heterogeneous multi-modal characteristics. Nonetheless, since temporal features are not taken into account, these techniques are sensitive to flashlights and object motions in real-world programs.

In [37], Xu et al. suggest a shot identification technique for news video based on object segmentation and tracking. They integrate three principal techniques: partitioned histogram comparison, object segmentation, and tracking based on wavelet analysis. The authors of [38] have devised a neural network classifier in order to identify transitions. The classifier is trained with a dissolve synthesizer that produces artificial dissolves. The technique relates to contrast-based and color-based features, and has yielded satisfactory results in comparison with standard techniques that are based on edge change ratios. In [39], the authors suggest co-histograms be utilized for video analysis, which is a statistic graph produced through counting the correspondent pixel pairs of two images. Nevertheless, their technique is not sensitive to camera zooming. A training-based technique is also devised in [4], where a probabilistic-based algorithm is suggested for identifying both abrupt and gradual transitions. After creating priori likelihood functions via training experiments, they consider all related knowledge of SBD, including shot-length distribution and visual discontinuity patterns at shot boundaries.

In recent time, researchers have become conscious of the significance of the temporal modeling of features for video SBD tasks. In [8], Grana et al. develop a linear transition model for SBD; their method concentrates only on gradual transitions with a linear behavior, as well as abrupt transitions. They use an accurate model which provides more discriminative power than with common methods. In [33], Yuan et al. conduct research into the SBD problem: they propose a general, formal framework from the viewpoint of pattern recognition. They research the principal challenges that are demonstrated by the frameworks. In the meantime, they propose a unified SBD technique based on the graph partition model.

We propose a GED and Gaussian transition-based SBD approach, and illustrate its effectiveness via a theoretical and practical analysis. As opposed to the methods mentioned before, our technique is able to detect different types of gradual shots, and is not sensitive to camera zooming and motion, object motions, and illumination changes. Eventually, we assess our method on the TRECVID 2006 dataset.

3 PROBLEMS AND MOTIVATIONS

In this section, we discuss the problem of SBD and attend to several challenging issues. We then discuss the motivations for and philosophy behind our approach to solving these problems.

Recently, in [33], an SBD was defined as a pattern recognition task, which was formed as a classification system with three major modules: representation of a visual content module, construction of a continuity signal module, and classification of continuity values. They studied the visual aspect of a video signal. From this perspective, video can be viewed as a kind of three-dimensional signal in which two dimensions disclose the visual content in the horizontal and vertical frame direction and the third discloses the variations of the visual content over the time axes. This formulation extracts some visual features from each frame and obtains a compact content representation and then calculates the continuity (similarity) values of the adjacent features. In this way, the visual content flow is transformed into a 1-D temporal signal. In the ideal situation, the continuity signal within the same shot maintains large magnitudes, while it drops to low values surrounding the positions of shot transitions. Finally, they classified the boundaries from the non-boundaries or identified the types of the transitions. This formal study of SBD makes it clear that the shot detection task is much more challenging than a traditional SBD task. The following are aspects of the challenge.

1. Representation of visual content and extraction of appropriate features is the significant step in SBD approaches and has a greater effect on the efficiency of other modules. The values of visual features must be constant values throughout a shot and must be irregular during a shot transition. This poses a challenge when searching visual features that satisfy the previous limitations.
2. Only a few spatial features (including the intensity of pixels, color histograms, edge, etc.) have been studied in recent approaches. By using these features, abrupt illumination changes, such as flashlights within shots, often cause significant discontinuities in the inter-frame feature, which means they are often mistaken for shot boundaries. Several illumination-invariant features and similarity metrics have been proposed to deal with the problem. However, these methods are not usually successful because temporal dependencies between the frames have not been considered. Therefore, collecting temporal features from video sequences would be a challenging practice.
3. The spatial features do not clearly model the image difference caused by a large camera or object motion, so one is unable to distinguish between smooth camera motion/parameter changes and gradual scene transitions. The solution to this problem is still an open issue.

To attack these challenges in a unified way, we advise using a GED and Gaussian transition-based approach, which can significantly increase the effectiveness of shot detection tasks while at the same time reducing the computational cost. The main ideas in our solution are summarized below.

1. To handle the representation of visual content, we utilize three-dimensional histograms as spatial features in the RGB color space of each frame. The histogram reflects the overall perspective of each frame and has higher stability but misses local information. To incorporate the spatial information of the color distribution, we partition each frame into 4 blocks of the same size and create a three-dimensional histogram for each of the blocks. We use these histograms as a feature vector of each frame in a video.
2. To solve the loss of temporal features, we use the spatial feature vector of each frame as a column to construct a feature matrix. We apply the GED on this matrix and incorporate the GED components of this matrix as temporal features along the frames.
3. To distinguish between the shot transitions and the image differences caused by a large camera or object motions, we model each shot transition by using a Gaussian function. First, we employ a GED-based distance function to identify the candidates for shot transition. We developed an iterative algorithm which, given a frame of the candidate set, tries to find the best center position for the

transition by minimizing an error function, which computes the fitness of data for the Gaussian model.

4 GENERALIZED EIGENVALUE DECOMPOSITION (GED)

In this section, our objective is to review the generalized eigenvalue decomposition (GED), which various researchers have presented for rectangular matrices [12, 41]. Before doing so, we must introduce some concepts.

We know that studying the eigenvalues and eigenvectors of a square matrix reveals valuable and useful information. It is also well known that the eigenvalue and eigenvector of a matrix in the usual sense can only be defined for square matrices. In [12], the authors extended these concepts for rectangular matrices based on Radic's determinant [13]. The determinant is a straightforward definition for computing the determinant of non-square matrices, which is identical to a conventional determinant used when a matrix is square. It is defined as follows:

Definition 1 (Radić's Determinant [13]). Let $A = [a_{i,j}]$ be an $m \times n$ matrix with $m \leq n$. Determinant of A, is defined as:

$$\det(A) = \sum_{1 \leq j_1 < \ldots < j_m \leq n} (-1)^{r+s} \det \begin{bmatrix} a_{1,j_1} & \ldots & a_{1,j_m} \\ \vdots & \ddots & \vdots \\ a_{m,j_1} & \ldots & a_{m,j_m} \end{bmatrix}, \tag{1}$$

where $j_1, \ldots j_m \in \mathbb{N}$, $r = 1 + \ldots + m$, and $s = j_1 + \ldots + j_m$. If $m > n$, then $\det(A) = 0$.

This determinant is a multilinear operator with respect to the row vectors, and skew-symmetric and has some important properties such as Laplace's expansion with respect to rows [14]. In [12], it is used to define a generalized eigenvalue and eigenvector for rectangular matrices.

Definition 2 (Generalized Eigenvalue and Eigenvector [12]). Let A and B be two rectangular matrices of order $m \times n$ and $m \leq n$. The scalar $\lambda_{A,B}$ and the vector $X \in \mathbb{R}^{n\times 1}$ are called the generalized eigenvalue and generalized eigenvector of matrices A and B, respectively, if and only if satisfying the following conditions:

$$(\lambda_{A,B}A - B)X = 0 \qquad \text{and} \qquad \det(\lambda_{A,B}A - B) = 0.$$

In [12], the authors studied some theorems that discuss the existence of and method of calculating the generalized eigenvalues and eigenvectors. They also presented the Generalized Eigenvalue Decomposition (GED) for rectangular matrices based on these definitions.

Theorem 1 (Generalized Eigenvalue Decomposition(GED)). Let A and B be any two $m \times n$ matrices with $m \leq n$, and suppose that $\lambda_1, \lambda_2, \ldots, \lambda_m$ be generalized eigenvalues and $\{V_1, V_2, \ldots, V_m\}$ be corresponding set of m linearly independent generalized eigenvectors, then the matrix A can be decomposed as

$$A = B.P.\Sigma.P_r^{-1},$$

where $P = [V_1, V_2, \ldots, V_m]$ and $\Sigma = diag(\lambda_1, \ldots, \lambda_m)$ is a diagonal matrix, λ_i is the corresponding eigenvalue of V_i for $i \in \{1, 2, \ldots, m\}$ and P_r^{-1} is the right hand inverse of P.

Proof. See [12]. □

The GED of a rectangular matrix is a factorization of the matrix into a product of four matrices. This theorem is true for any rectangular matrices A and B of the same order. To meet our needs, we assume the matrix B is always equal to a generalized identity matrix, which is defined as:

$$B = I_{m \times n} = \begin{bmatrix} 1 & 0 & \ldots & 0 & | & 0 & \ldots & 0 \\ 0 & 1 & 0 & \vdots & | & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & | & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 1 & | & 0 & \ldots & 0 \end{bmatrix}.$$

In the following section, the intuitive characteristics of GED for SBD in a video sequence are discussed.

5 SHOT BOUNDARY DETECTION USING GED

In this section we will present the proposed SBD algorithm. Initially, to reduce the number of frames to be processed by GED, the input video sequence was sampled with a fixed rate of five frame/second. Our experiments have shown that this sampling rate is sufficient for video programs without many dramatic motions. The algorithm extracts a 500-dimensional feature vector A_i for each frame i in a sampling set. Using A_i as column vector i, we obtain video feature matrix $A = [A_1, \ldots, A_n]$. The feature matrix is fed into the GED-based algorithm to detect shot transitions. Finally, the transition type is detected by monitoring the behavior of the shot transitions. The details of the feature extraction and SBD algorithm are related in the next subsections.

5.1 Feature Extraction

In order to extract spatial features of frames from a wide range of image features, we preferred using color histograms, which are very essential for signifying the overall spatial features of each frame [34]. The combination of color histogram and GED capture the temporary color distribution information for each shot.

Let $F = \{F_1, F_2, \ldots, F_n\}$ be the sampling set of frames of an arbitrary video sequence. For each frame i, we create a 500-dimensional feature vector A_i. Using

A_i as a column vector i, we obtain feature matrix $A = [A_1, \ldots, A_n]$. To compute the feature matrix, in our system implementation, we create three-dimensional histograms in RGB color space with five bins for R, G and B, respectively, resulting in a total of 125 bins. To incorporate spatial information about the color distribution, we partition each frame into 4 blocks of the same size, and create a 3D-histogram for each of the blocks. These four histograms are then concatenated to form a 500-dimensional feature vector for the frame. Finally, using the feature vector of frame i as the i^{th} column, we create the feature matrix A for the video sequence.

5.2 GED-based Distance Metric

Now, we are in a position to prove our main theorem, using the concepts discussed earlier. To do this, we rewrite the GED as follows:

$$A = \underbrace{I_{m\times n}.P}_{\text{Q}}.\Sigma.P_r^{-1} = Q_{m\times m}.\Sigma.P_r^{-1} \tag{2}$$

where $Q = [q_1, \ldots, q_m]$ is an $m \times m$ matrix and $P_r^{-1} = [p'_1, \ldots, p'_m]^T$ is an $m \times n$ matrix. Here, we provide the main theorem that will be utilized in SBD algorithm.

Theorem 2. Let the GED of an $m \times n$ matrix $A = [a_{i,j}]$, where $m \leq n$ is given by Equation (2). If $k \leq r = rank(A)$ and

$$A_k = \sum_{i=1}^{n} \lambda_i q_i p'_i, \tag{3}$$

then

$$\min_{rank(B)=k} = \|A - B\|_2 = \|A - A_k\|_2 = \|\lambda_{k+1}\|_2.$$

Proof. Since $Q^{-1}.A_k.P = diag(\lambda_1, \ldots, \lambda_k, 0, \ldots, 0)$, then $\|A - A_k\|_2 = \lambda_{k+1}$. Now, suppose $rank(B) = k$ for some $B \in R^{m\times n}$. It follows that we can find orthogonal vectors $x_1, \ldots, x_{n-k}$ such that $null(B) = span(x_1, \ldots, x_{n-k})$. A dimension argument shows that:

$$span(x_1, \ldots, x_{n-k}) \cap span(p'_1, \ldots, p'_{k+1}) \neq 0.$$

Let z be a unit 2-norm vector in this intersection. Since $B.z = 0$ and

$$A.z = \sum_{i=1}^{k} \lambda_i (p'_i.z) q_i,$$

we have:

$$\|A - B\|_2^2 \geq \|(A - B).z\|_2^2 = \|A.z\|_2^2 = \sum_{i=1}^{k} \lambda_i^2 (p'_i.z)^2 \geq \|\lambda_{k+1}\|_2^2,$$

This completes the proof of the theorem. $\square$

This theorem could have many applications in solving problems in video and data mining domains. Discarding small generalized eigenvalues is equivalent to discarding linearly semi-dependent or practically non-essential axes of the eigenvector space. In our case, axes with small generalized eigenvalues usually capture either non-essential color variations or noise within the video sequence. The truncated GED, in one sense, captures the most salient underlying structure in the association of histograms and video frames, yet at the same time removes the noise or trivial variations in video frames.

To demonstrate the impact of discarding small generalized eigenvalues on showing the similarity between video frames, we implemented a SBD algorithm in eigenvector space and evaluated the performance of the algorithm using k in the Theorem 2 as a parameter. More precisely, for each frame i, we take the column p'_i of matrix P_r^{-1} as its feature vector and use Equation (4) as the similarity metric to compare it with frame $j = i + 1$:

$$D^k(p'_i, p'_j) = \sqrt{\sum_{h=1}^{k} |\lambda_h|^2 |p'_{h,i} - p'_{h,j}|^2}. \tag{4}$$

In fact, Equation (4) defines a Euclidean distance weighted by the generalized eigenvalues λ_i, k as the parameter that specifies how many eigenvalues are to be used in the metric, and it will be computed by training datasets. We also define the distance between two subsequent frames as follows:

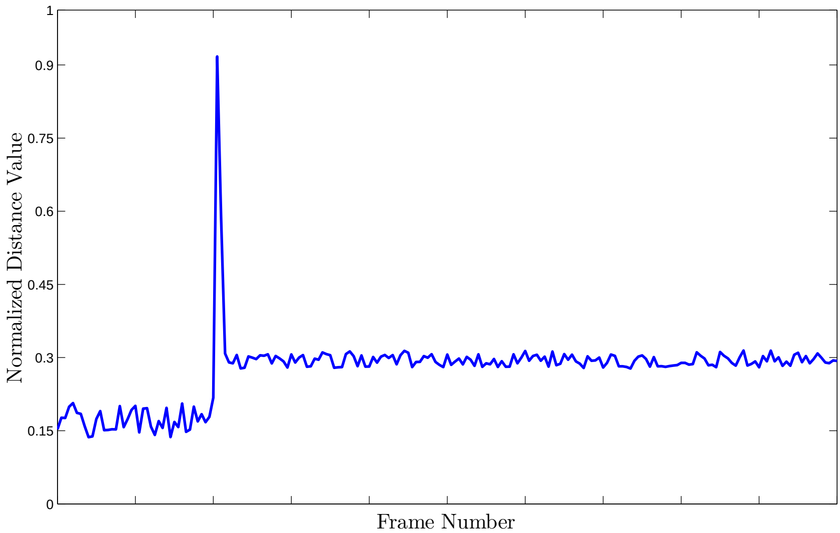
$$\Gamma^k(i) = D^k(p'_i, p'_{i+1}) = \sqrt{\sum_{h=1}^{k} |\lambda_h|^2 |p'_{h,i} - p'_{h,i+1}|^2}. \tag{5}$$

In the next subsection, this distance function will be utilized to detect shot transitions.

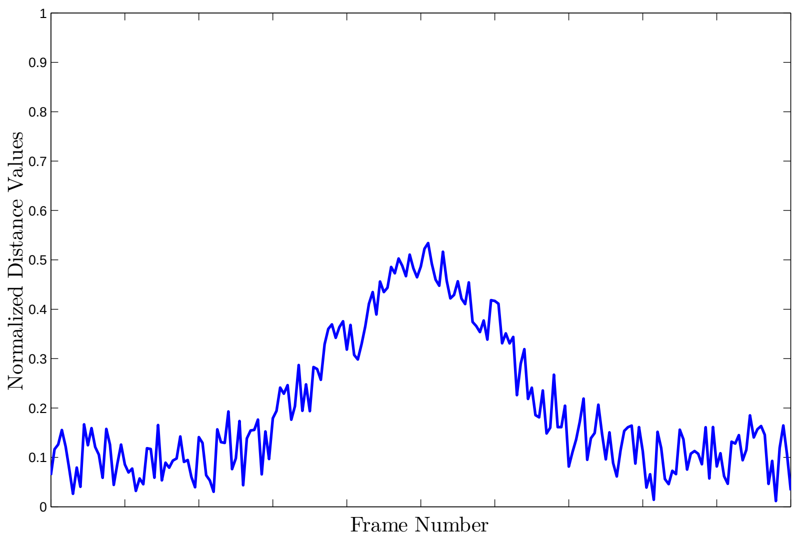
5.3 GED-based Candidate Selection for Gradual Shot Transitions

The distance function in Equation (5) represents the dissimilarity between two frames in generalized eigen space. The intra-shot frames are very similar, so the distance function of Equation (5) will be constant and close to zero. In the inter-shot frames, we also expect the distance function will change gradually. Experimental results show that in the gradual shot boundaries, the distance function has semi-Gaussian behavior (see Figure 1 b)). The proposed algorithm detects the types of shot boundary transitions by analyzing the behavior of this function in detail.

In the hard cut transitions, where one frame belongs to the disappearing shot and the next one to the appearing shot, experimental results show that the distance function changes abruptly and its value is close to one (see Figure 1 a)). Consequently, the hard cut transitions could be detected by using simple thresholding.



a)



b)

Fig. 1. The values of proposed GED-based distance function for a gradual transition and a hard cut transition: a) a hard cut transition, b) a gradual transition

To detect the gradual shot transitions, we employ an algorithm constructed of two steps. The first one finds the candidates for the gradual transition centers, and the second one searches the candidates to find correct transitions. In the first step, the candidates could be computed by applying a threshold on the distance function of Equation (5). The following pseudo code clarifies the mentioned method for hard cut detection and candidate selection of gradual shot centers in detail (see Algorithm (1)).

Algorithm 1 Hard cut detection and candidate selection of gradual shot centers

for each frame f_i in $F = f_1, f_2, \ldots, f_t$ **do**

a) Compute $\Gamma^k(j)$ Using Equation (5)

b)

- **if** $\Gamma^k(j) > HardCutThreshold$ **then**
  Frame f_j is a Hard Cut Transition
- **else**
  if $\Gamma^k(j) > GradualTransitionThreshold$ **then**
  Frame f_j is a candidate for Gradual Transition center
  end if
- **end if**

end for

The $HardCutThreshold$ and $GradualTransitionThreshold$ are computed by using training datasets. To compute the $HardCutThreshold$, we consider all shots of the training set with hard cut transitions. We select, through exhaustive search, the largest real number in [0, 1] that maximizes the sum of precision and recall as $HardCutThreshold$. The $GradualTransitionThreshold$ is a positive real number that is less than $HardCutThreshold$ and is determined similarly by using all gradual shots in the training set.

5.4 Finding the Correct Gradual Shot Transitions

As shown in the previous section (Figure 1), the distance function of Equation (5) has Gaussian behavior during the gradual transitions. Ideally, an arbitrary Gaussian function is defined as:

$$G^{\eta}_{\sigma^2}(t) = G_0 . \exp\left(- \frac{(t-\eta)^2}{\sigma^2} \right), \tag{6}$$

where σ is a real number defined as a variance value of the function and $t = \eta$ is the mean or center of the function.

After finding the candidates to serve as gradual transition centers, it is necessary to analyze the candidates and detect the correct gradual transitions. In an ideal case, a correlation between the distance function of Equation (5) and the Gaussian function of Equation (6) could be a sufficient indicator of correct transition presence, but due to a lack of ideality in most of the shot transitions, we employ an algorithm that explores the candidates at the center of the shot transition and detect the correct transition automatically.

For this reason, we designed an algorithm that uses the distance function of Equation (5) to find the correct transition. It is imperative to observe that in the gradual shots, the transition is not detectable in all shot boundary frames. Consequently, the distance function of Equation (5) does not rise to an absolute maximum at the center of the shot transition. Thanks to the properties of GED in Theorem 2, increasing the size of the window of the shot transition (variance in Equation (6)) makes the values of the function rise to an absolute maximum. Furthermore, in this case, the length of the transition where it is defined as the number of frames, in which the transition is visible, is $L = 2\sigma - 1$. Therefore, we will utilize these properties to detect correct transitions from the obtained candidates of center of shot transitions in the previous section.

Let $\overline{n}$ be an obtained candidate for the center of a gradual shot. Then, the median parameter of Equation (6) is set $\eta = \overline{n}$ and the coefficient parameter of the Gaussian function is set $G_0 = \Gamma^k(\overline{n})$. Now, we should find a value for the variance parameter σ in Equation (6) that makes a maximum adjustment with the distance function of Equation (5). To do this, the following measure is defined:

$$W_\sigma^{\overline{n}} = \sum_{i=0}^{2\sigma} \left\{ \left| G_{\sigma^2}^{\overline{n}}(\overline{n} - i) - \Gamma^k(\overline{n} - i) \right| + \left| G_{\sigma^2}^{\overline{n}}(\overline{n} + i) - \Gamma^k(\overline{n} + i) \right| \right\}. \tag{7}$$

The σ value that minimizes $W_\sigma^{\overline{n}}$ is defined as an optimum value for the variance parameter:

$$\overline{\sigma} = \arg \min_{0 \leq \sigma \Sigma} \left\{ W_\sigma^{\overline{n}} \right\} \tag{8}$$

where Σ is the maximum size that a transition can assume.

Under typical conditions, the algorithm searches the appropriate variance parameter leading to the expected Gaussian shape and finding the correct σ and thus the transition length. If Z is the total number of candidates at the center of gradual shots and since this part of the algorithm is run just for the obtained candidate points, the total number of the computation is definitely less than the product of Σ and Z.

If the median and variance parameters are $\overline{n}$ and $\overline{\sigma}$, respectively, as detected by the algorithm, then the function $G_{\sigma^2}^{\overline{n}}(t)$ is a Gaussian function that has maximum adjustment with the distance function $\Gamma^k(j)$ in Equation (5). We must now verify the significance of the transition and how well the real data fits with the Gaussian function. We introduce the following measure:

$$Height_{\overline{\sigma}}^{\overline{n}} = \Gamma^k(\overline{n}) - \min \left\{ \Gamma^k(\overline{n} - 2\overline{\sigma}), \Gamma^k(\overline{n} + 2\overline{\sigma}) \right\}. \tag{9}$$

This value is the height of the center value with respect to the lower of two values of $\Gamma^k(j)$ that correspond to the extremes of the Gaussian function, and it provides information on the transition significance.

Experimental results show that in real cases the object and camera motion cause some semi-Gaussian behavior in the distance function of Equation (5). In order to address this effect, we must find the hypothesis of having an isosceles Gaussian function and define the fitting error measure as:

$$Error_{\overline{\sigma}}^{\overline{n}} = \frac{1}{4\overline{\sigma}} \sum_{i=1}^{2\overline{\sigma}} \left\{ \left| G_{\overline{\sigma}^2}^{\overline{n}}(\overline{n} - i) - \Gamma^k(\overline{n} - i) \right| + \left| G_{\overline{\sigma}^2}^{\overline{n}}(\overline{n} + i) - \Gamma^k(\overline{n} + i) \right| \right\}. \quad (10)$$

This error sum is divided by the interval's length to achieve a measure that is independent from the transition length. Moreover, a minimum threshold on the $Height_{\overline{\sigma}}^{\overline{n}}$ value, T_H and a maximum threshold on the $Error_{\overline{\sigma}}^{\overline{n}}$ value, T_E are employed to discriminate real gradual shot changes from false ones. The final decision is made based on two parameters. Finally, the analysis of the candidates and detecting the correct gradual transitions could be summarized as in Algorithm 2.

Algorithm 2 Analysis of the candidates and detecting the correct gradual transitions

for each all $\overline{n}$ in the candidates for center of gradual shots **do**

a)
- **for** all $0 \leq \sigma \leq \Sigma$ **do**
  $\overline{\sigma} = arg\min_{0 \leq \sigma \Sigma} \left\{ W_{\sigma}^{\overline{n}} \right\}$ According to Equation (8)
- **end for**

b) Compute $Height_{\overline{\sigma}}^{\overline{n}}$ According to Equation (9)

c) Compute $Error_{\overline{\sigma}}^{\overline{n}}$ According to Equation (10)

d)
- **if** $Height_{\overline{\sigma}}^{\overline{n}} \geq T_H$ and $Error_{\overline{\sigma}}^{\overline{n}} \leq T_E$ **then**
  $Transition(\overline{n}, \overline{\sigma}) = TRUE$
- **end if**

end for

Hence, the correct gradual transitions and the length of the transitions could be extracted by using this algorithm.

6 EXPERIMENTAL RESULTS

The proposed video shot transition detection algorithm was evaluated using a 4-hour video set. All videos were segmented manually by locating hard cuts and gradual transitions together with their length. A total of 1180 shot transitions existed in these video sequences; 1027 were hard cut transitions and 153 were gradual shot boundaries. The video clips were taken from the Internet and TV programs, and included from different movie formats such as AVI, MPEG-7, and SGI. The complete video database will be made available upon request. The details about each video are shown in Table 1.

Video	**Frames**	**Cuts**	**Graduals**
News	95 743	236	27
Cartoon	74 384	142	39
Movie	85 958	369	61
Sport	109 381	187	17
Documentary	57 491	93	9

Table 1. Video set used in experiments

We employed recall and precision as the measures for performance evaluation, which are defined as follows:

- The *Recall* measure, also known as the true positive function or sensitivity, corresponds to the ratio of correct experimental detections over the number of all true detections:

$$Recall = \frac{number\ of\ correctly\ detected\ boundaries}{number\ of\ true\ boundaries}. \tag{11}$$

- The *Precision* measure is defined as the ratio of correct experimental detections over the number of all experimental detections:

$$Precision = \frac{number\ of\ correctly\ detected\ boundaries}{number\ of\ totally\ detected\ boundaries}. \tag{12}$$

Figure 2 shows the evaluation result with the value of k in Equation (5) as a parameter. For each given k value, we empirically determined the SBD threshold so that the best recall and precision were obtained. It can be seen from the figure that when k equals 226, the means of both recall and precision reach their maximum. Hence, based on the above experiments, we set the value of k in Equation (5) to 226 and used this value as well as the proposed algorithm for the SBD task.

An excellent shot transition detector should have high precision and high recall. We compared our algorithm to the techniques proposed in [32, 33], which are free downloadable shot detection software that provides MPEG-7 or XML formatted output. The results of the algorithms applied on the same video sequences are shown in Table 2 (second and third columns). Several false shot cut detections were performed due to camera flashes. It is clear that the proposed video SBD system obtained reasonable performance and that our algorithm is robust.

In the experiments, the video set is sampled with a fixed rate of five frames per second. For feature extraction, each frame is divided into 2×2 blocks. In

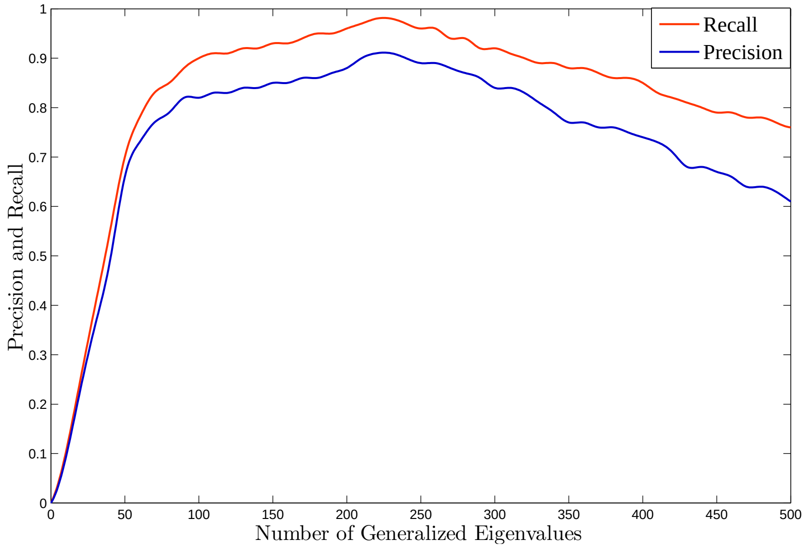


Fig. 2. Performance evaluation of GED based SBD using k as a parameter

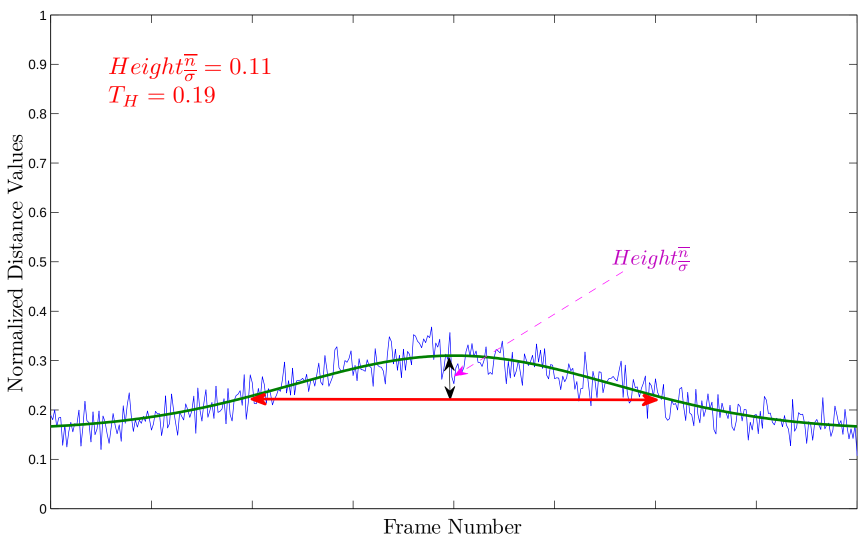
Video	**Our Method**		[44]		[45]	
	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.
News	96	99	91	93	89	94
Cartoon	94	96	74	86	88	91
Movie	92	96	83	92	86	91
Sport	95	98	87	92	79	90
Documentary	93	95	92	94	93	95
overall	94.0	96.8	85.4	91.4	87.0	92.2

Table 2. Comparison of results of the different algorithms in our dataset

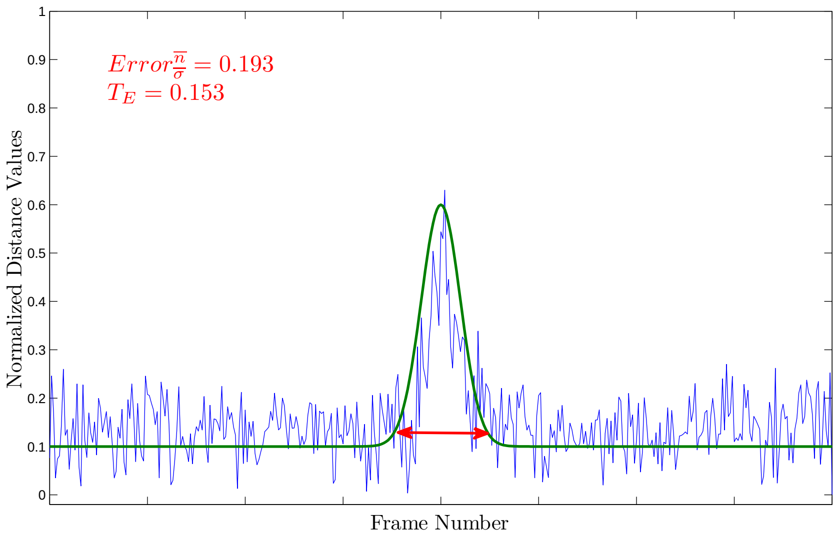
addition, the $HurdCuttThreshold$, $GradualTransitionThreshold$, T_E and T_H were tuned by using the "news_1.avi" as a training set. The algorithm searched for hard cut transitions and candidates of gradual transitions via HurdCuttThreshold and GradualTransitionThreshold, respectively. In order to detect correct gradual transitions, the method presented in Section 5.4 is used. The algorithm fits a Gaussian function with maximum adjustment to the probability function of these candidates, and then computes the $Height\frac{\overline{n}}{\sigma}$ and $Error\frac{\overline{n}}{\sigma}$ for them. The correct gradual transitions are detected based on these parameters. To do this, as discussed in Section 5.4, it uses T_E and T_H thresholds. We chose the thresholds via comprehensive search so as to maximize the sum of precision and recalls. Some of the examples of the detected candidates for gradual shots are shown in Figure 3. Also, in Figure 4, the $Error\frac{\overline{n}}{\sigma}$ diagram for all different types of gradual shot transitions is shown to justify the Semi-Gaussian assumption of the appearance of probability function of gradual transitions. The newscasts from the reference video test set TRECVID 2006 were inserted into the testing set in order to make it possible, in the future, to compare these techniques with other SBD techniques. This set consists of over 6 hours of video sequences that have been digitalized with a frame rate of 29.97 fps and a resolution of 352×264 pixels. To increase the speed of our computations for our experiments, spatially downsampled frames with resolutions of 176×132 pixels were used. The ground truth given by TRECVID was utilized for these video sequences.

The results of experiments are shown in Table 3. It is obvious that the achieved results are better than the results reported in the TRECVID 2006 competition [7]. The best reported hard cut detection results for recall and precision are 90 % and 88 %, respectively, whereas our method produced 98 % recall and 94 % precision. Most false detections were caused by flashing lights and high-speed camera motions. In some cases, false detections appeared in commercials where artistic camera edits were used. The missed shot cut detections were mainly caused by shot changes between two images with very similar spatial color distribution or when the shot change occurred in only a part of the video frame.

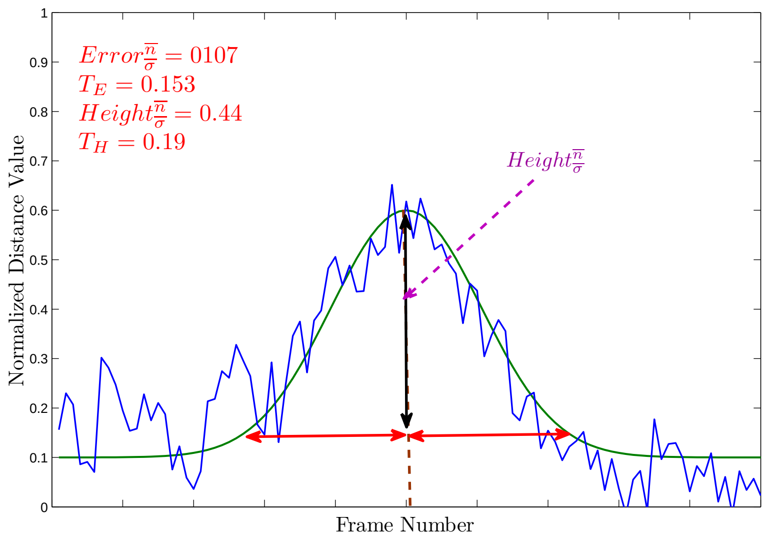
In Figure 5, the overall results of detection from TRECVID-2006 26 participants are shown. The total number of participant groups was 26, but only the best 20 are visible. The proposed algorithm obtained very good results in recall with a range of 92.4 %–97.2 % of correctly detected transitions using the same algorithm for all



a)



b)



c)

Fig. 3. Examples of the detected candidates for gradual transitions, and finding the correct gradual transitions: a) as $Height_{\overline{\sigma}}^{\overline{n}} \geq T_H$, it is an incorrect gradual transition; b) as $Error_{\overline{\sigma}}^{\overline{n}} \leq T_E$, it is an incorrect transition; c) a correct gradual transition

different types of transitions. As the GED-based approach has no prediction for fast camera motions, high light changes, or picture-in-picture changes, the results concerning its accuracy are not as satisfactory.

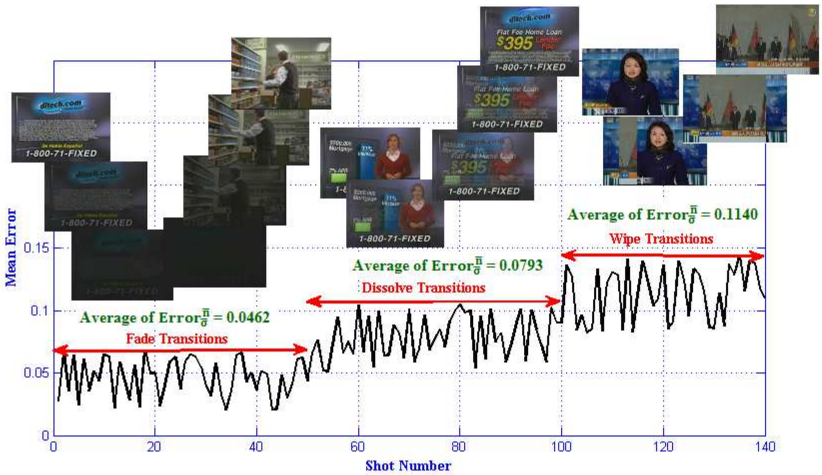


Fig. 4. The diagram of $Error^{\overline{n}}_{\overline{\sigma}} \leq T_E$ for all different types of gradual shot transition

	All		Cuts		Gradual			
							Frame	
Ref	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.	Rec.	Prec.
a	0.6898	0.7425	0.7065	0.7868	0.6446	0.6541	0.7243	0.7850
b	0.8210	0.8986	0.9216	0.8507	0.7416	0.8355	0.8739	0.9261
c	0.5953	0.8317	0.5926	0.8387	0.6030	0.8101	0.8275	0.7984
d	0.6403	0.5723	0.7284	0.5954	0.4031	0.5276	0.5639	0.7834
e	0.8317	0.8217	0.9070	0.8873	0.6420	0.6507	0.8527	0.5637
f	0.5377	0.6044	0.7311	0.6036	0.0159	0.7416	0.2540	0.7056
g	0.3278	0.1595	0.3703	0.1431	0.2126	0.3778	0.4269	0.7766
h	0.7617	0.8687	0.8215	0.8888	0.6013	0.8024	0.7716	0.8486
i	0.7848	0.7344	0.7949	0.8170	0.7565	0.5711	0.7726	0.7000
GED	0.9603	0.9483	0.9729	0.9488	0.9498	0.9247	0.9611	0.9479

Table 3. The result of experiments on TRECVID 2006 data set

7 CONCLUSION

This paper discusses a novel SBD algorithm that was developed based on the Generalized Eigenvalue Decomposition (GED) and modeling of gradual transitions by

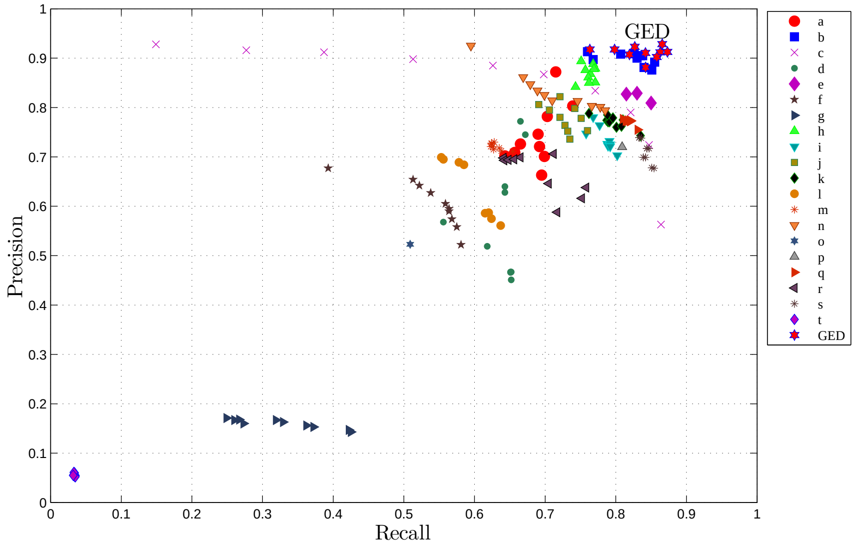


Fig. 5. Results on the overall detection (cuts and transitions) based on the data provided by the organizers in TRECVID 2006. Our approach is labeled GED.

Gaussian functions. We derived a theorem that discussed some new features of GED and we utilized it in defining a new distance function for SBD. The distance function has abrupt changes in hard cut transitions and semi-Gaussian behavior in gradual transitions. The algorithm detects the transitions by analyzing this probability function. The proposed algorithm was implemented and evaluated using the TRECVID benchmark platform. Our results measure up to the best results of other studies. The evaluation demonstrates the effectiveness of the proposed approach on the TRECVID 2006 data collection. The algorithm can be considered robust because it detects a gradual shot efficiently and handles disturbances (high-speed zooming, fast camera and object motions) within a shot.

REFERENCES

[1] Brezeale, D.—Cook, D. J.: Automatic Video Classification: A Survey of the Literature. IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews, Vol. 38, 2008, No. 3, pp. 416–430.

[2] Lienhart, R.: Comparison of Automatic Shot Boundary Detection Algorithms. In: Storage and Retrieval for Image and Video Databases VII, Vol. 3656 of Proceedings of SPIE, San Jose, Ca, USA, January 1999, pp. 290–301.

[3] Cooper, M.—Liu, T.—Rieffel, E.: Video Segmentation via Temporal Pattern Classification. IEEE Transactions on Multimedia, Vol. 9, 2007, No. 3, pp. 610–618.

[4] Hanjalic, A.: Shot Boundary Detection: Unraveled and Resolved? IEEE Transactions on Circuits and Systems for Video Technology, Vol. 12, 2002, No. 2, pp. 90–105.

[5] Cotsaces, C.—Nikolaidis, N.—Pitas, I.: Video Shot Detection and Condensed Representation: A Review. IEEE Signal Processing Magazine, Vol. 23, 2006, No. 2, pp. 28–37.

[6] Cernekova, Z.—Pitas, I.—Nikou, C.: Information Theory-Based Shot Cut/Fade Detection and Video Summarization. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 16, 2006, No. 1, pp. 82–91.

[7] NIST, Homepage of Trecvid Evaluation. Availaible on: `http://www-nlpir.nist.gov/projects/trecvid/`.

[8] Grana, C.—Cucchiara, R.: Linear Transition Detection as a Unified Shot Detection Approach. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, 2007, No. 4, pp. 483–489.

[9] Flickner, M.—Sawhney, H.—Niblack, W.—Ashley, J.—Qian, H.—Dom, B.—Gorkani, M.—Hafner, J.—Lee, D.—Petkovic, D.—Steele, D.—Yanker, P.: Query by Image and Video Content: The QBIC System. IEEE Computer Magazin, Vol. 28, 1995, No. 9, pp. 23–32.

[10] Chang, S. F.—Chen, W.—Meng, H. J.—Sundaram, H.—Zhong, D.: A Fully Automated Content-Based Video Search Engine Supporting Spatio-Temporal Queries. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 8, 1998, No. 5, pp. 602–615.

[11] Hamrapur, A.—Gupta, A.—Horowitz, B.—Shu, C. F.—Fuller, C.—Bach, J.—Gorkani, M.—Jain, R.: Virage Video Engine. In: Storage and Retrieval for Image and Video Databases V, Vol. 3022 of Proceedings of SPIE, San Jose, California, USA, February 1997, pp. 188–197.

[12] Amiri, A.—Fathy, M.: Generalization of Eigenvalue and Eigenvector for Non-Square Matrices Based on Radić's Determinant and its Application to Video Retrieval. To appear in International Journal of Applied and Computational Mathematic (A & CM), 2011.

[13] Radić, M.: Areas of Certain Polygons in Connection with Determinants of Rectangular Matrices. Beitrage zur Algebra and Geometric Contributions to Algebra and Geometry, Vol. 49, 2008, No. 1, pp. 71–96.

[14] Radić, M.: About a Determinant of Rectangular $2 \times n$ Matrix and its Geometric Interpretation. Beitrage zur Algebra and Geometric Contributions to Algebra and Geometry, Vol. 46, 2005, No. 2, pp. 321–349.

[15] Qu, Z.—Liu, Y.—Ren, L.—Chen, Y.—Zheng, R.: A Method of Shot Detection Based on Color and Edge Features. In: Web Society, Proceedings of the First IEEE Symposium, SWS '09, Lanzhou, August 2009, pp. 1–4.

[16] Hameed, A.: A Novel Framework of Shot Boundary Detection for Uncompressed Videos. In: Emerging Technologies, Proceedings of International Conference, ICET '09, Islamabad, October 2009, pp. 274–279.

[17] Jiang, X.—Sun, T.—Li, J.—Chen, X.: A Novel Shot Edge Detection Algorithm Based on Chi-Square Histogram and Macro-Block Statistics. In: Information Scien-

ce and Engineering, Proceedings of International Symposium, ISISE'08, Shanghai, December 2008, pp. 604–607.

[18] Zhang, M. X.—Zheng, J. L.—Li, H.—Chang, J. Y.: A Novel Shot Segmentation Algorithm Based on Grid-Mapping Dynamic Windows in Compressed Videos. In: Wavelet Analysis and Pattern Recognition, Proceedings of International Conference, ICWAPR '09, Baoding, July 2009, pp. 16–21.

[19] Tan, W.—Cao, J.—Li, H.: Algorithm of Shot Detection Based on SVM with Modified Kernel Function. In: Artificial Intelligence and Computational Intelligence, Proceedings of International Conference, AICI '09, Shanghai, November 2009, pp. 11–14.

[20] Wang, X.—Wang, S.—Chen, H.: An Effective Soccer Video Shot Detection Algorithm. In: Computer Science and Software Engineering, Proceedings of International Conference, CSSE '08, Wuhan-Hubei, December 2008, pp. 214–217.

[21] Li, J.—Ding, Y.—Shi, Y.—Zeng, Q.: DWT-Based Shot Boundary Detection Using Support Vector Machine. In: Information Assurance and Security, Proceedings of International Conference, IAS '09, Xi'an, August 2009, pp. 435–438.

[22] Yang, M.—Serrano, J. C.—Grecos, C.: MPEG-7 Descriptors Based Shot Detection and Adaptive Initial Quantization Parameter Estimation for the H.264/AVC. IEEE Transactions on Broadcasting, Vol. 55, 2009, No. 2, pp. 165–177.

[23] Wu, X.—Yuen, P. C.—Liu, C.: Shot Boundary Detection: An Information Saliency Approach. In: Image and Signal Processing, Proceedings of Congress, CISP '08, Sanya-China, May 2008, pp. 808–812.

[24] Tao, L. Z.—Huang, X. L.—Wang, X.: Shot Boundary Detection Based on Macroblock and DC Image. In: Measurement and Service Science, Proceedings of International Conference, MASS '09, Wuhan-Hubei, September 2009, pp. 1–3.

[25] Chavez, G. C.—Precioso, F.—Cord, M.—Foliguet, S. P.—Araujo, A.: Shot Boundary Detection by a Hierarchical Supervised Approach. In: Systems, Signals and Image Processing, Proceedings of 14^{th} International Workshop, MASS '09, Maribor, June 2007, pp. 197–200.

[26] Ma, C.—Lee, C. H.: Unsupervised Anchor Shot Detection Using Multi-Modal Spectral Clustering. In: Acoustics, Speech and Signal Processing, Proceedings of International Conference, ICASSP '08, Las Vegas, April 2008, pp. 813–816.

[27] Bai, L.—Lao, S. Y.—Liu, H. T.—Bu, J.: Video Shot Boundary Detection Using Petri-Net. In: Machine Learning and Cybernetics, Proceedings of International Conference, ICMLC '08, Kunming, July 2008, pp. 3047–3051.

[28] Dailianas, A.—Allen, R. B.—England, P.: Comparison of Automatic Video Segmentation Algorithms. In: Integration Issues in Large Commercial Media Delivery Systems, Vol. 2615 of Proceedings of SPIE, Philadelphia, Pa, USA, October 1996, pp. 2–16.

[29] Ahanger, G.—little, T.: A Survey of Technologies for Parsing and Indexing Digital Video. Visual Communication and Image Representation, Vol. 7, 1996, No. 1, pp. 28–43.

[30] Patel, N. V.—Sethi, I. K.: Video Shot Detection and Characterization for Video Databases. Pattern Recognition, Vol. 30, 1997, No. 4, pp. 583–592.

[31] Tsekeridou, S.—Pitas, I.: Content-Based Video Parsing and Indexing Based on Audio-Visual Interaction. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, 2001, No. 4, pp. 522–535.

[32] Huang, C.-L.—Liao, B.-Y.: A Robust Scene-Change Detection Method for Video Segmentation. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 11, 2001, No. 12, pp. 1281–1288.

[33] Yuan, J.—Wang, H.—Xiao, L.—Zheng, W.—Li, J.—Lin, F.—Zhang, B.: A Formal Study of Shot Boundary Detection. IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, 2007, No. 2, pp. 168–186.

[34] Vasconcelos, N.—Lippman, A.: Statistical Models of Video Structure for Content Analysis and Characterization. IEEE Transactions on Image Processing, Vol. 9, 2000, No. 1, pp. 3–19.

[35] Lienhart, R.: Reliable Ttransition Detection in Videos: A Survey and Practitioner's Guide. International Journal on Image and Graphics, Vol. 1, 2001, No. 3, pp. 469–486.

[36] Ling, X.—Yuanxin, O.—Huan, L.—Zhang, X.: A Method for Fast Shot Boundary Detection Based on SVM. In: Image and Signal Processing, Proceedings of IEEE International Congress, CISP'08, Sanya-China, May 2008, pp. 445–449.

[37] Xu, X. W.—Li, G. H.—Yuan, J.: A Shot Boundary Detection Method for News Video based on Object Segmentation and Tracking. In: Machine Learning and Cybernetics, Proceedings of Seventh IEEE International Conference, ICMLC'08, Kunming, July 2008, pp. 2470–2476.

[38] Lienhart, R.—Zaccarin, A.: A System for Reliable Dissolve Detection in Videos. In: Image Processing, Proceedings of IEEE International Conference, ICIP '01, Thessaloniki-Greece, October 2001, pp. 406–409.

[39] Hao, P.—Chen, Y.: Co-Histogram and Its Application in Video Analysis. In: Multimedia and Expo, Proceedings of IEEE International Conference, ICME'04, Taipei, June 2004, pp. 1543–1546.

[40] Browne, P.—Smeaton, A. F.—Murphy, N.—O'Connor, N.—Marlow, S.—Berrut, C.: Evaluating and Combining Digital Video Shot Boundary Detection Algorithms. In: Machine Vision and Information Processing, Proceedings of 4th Irish Conference, IMVIP'00, Belfast-Ireland, September 2000, pp. 108–114.

[41] Amiri, A.—Fathy, M.: Video Shot Boundary Detection Using Generalized Eigenvalue Decomposition. In: Computational Science and Its Applications, Proceedings of International Conference, ICCSA '09, Suwon-Korea, June 2009, pp. 780–790.

[42] Pei, S.-C.—Chou, Y.-Z.: Efficient MPEG Compressed Video Analysis Using Macroblock Type Information. IEEE Transactions on Multimedia, Vol. 4, 1999, No. 1, pp. 321–333.

[43] Cheng, W.—Xu, D.—Jiang, Y.—Lang, C.: Information Theoric Metrics in Shot Boundary Detection. In: Knowledge-Based Intelligent Information and Engineering Systems, Proceedings of 9th International Conference, KES05, Melbourne-Australia, September 2005, pp. 388–394.

[44] Miene, A.—Dammeyer, A.—Hermes, T.—Herzog, O.: Information Theoric Metrics in Shot Boundary Detection. In: Generalized Documents, Proceedings of ECDL WS, 2001, pp. 39–43.

[45] VCM: Video Content Management by Technologie-Zentrum Informatik. Availaible on: `http://atlas.ced.tuc.gr/`.

Ali Amiri received the M. Sc. Degree in Computer Engineering from Iran University of Science and Technology (IUST) in 2006, and Ph. D. Degree in Artificial Intelligence in February 2011 from IUST, Iran. Since 2011, he has worked in the Engineering Department of Zanjan University as an Assistant Professor. His research interests include video segmentation, video retrieval and summarization, matrix computation, pattern recognition and moving object detection and tracking.

Mahmood Fathy received B. Sc. Degree in Electronics from Iran University of Science and Technology in 1985, M. Sc. in 1987 from Bradford University, UK, in Computer Architecture, and the Ph. D. Degree in Image Processing Computer Architecture in 1991 from UMIST, UK. Since 1991, he has worked in the Computer Engineering Department of IUST as an Associate Professor. His research interests include real-time image processing in particular in traffic engineering and QoS in computer networks including video and image transmission over Internet.