A Novel Video Retrieval System Using GED-based Similarity Measure

Ali Amiri, Mahmood Fathy, and Atusa Naseri
Computer Engineering Department, Iran University of Science and Technology
1684613114 Narmak, Tehran, Iran
{a_amiri, Mahfathy, A_Naseri}@iust.ac.ir

Abstract

In this paper, we propose a novel video retrieval system using Generalized Eigenvalue Decomposition (GED). The system contains two major subsystems: database creation and database searching. In both subsystems, we propose new methods for shot-feature extraction, feature dimension transformation and feature similarity measuring base on GED. Experimental results confirm the effectiveness of our proposed system.

Keywords: *We Video retrieval, Spatio-Temporal Features, dimension transformation, Generalized Eigenvalue Decomposition.*

1. Introduction

Today people have access to a tremendous amount of video information in the internet. With a large video data collection, it is infeasible for a human to classify or cluster the video scenes or to find either the appropriate video scene, or the desired portions of the video. Video retrieval is an essential technology to design video search engines with serve as an information filter and sifts out an initial set of relevant videos from database.

A large number of approaches have been attempted for forming automatic content-based retrieval of video. After reviewing the literature of methods, we found that these approaches could be divided into two categories: video segmentation-based and motion-based features. Video segmentation is a fundamental step in analyzing video sequence content and in devising methods for efficient access, retrieval and browsing of large video databases. Generally speaking, there are two typical video segmentation methods, i.e. shot-based and object-based. To date many works focus on breaking video into shots, and then searching the video for appropriate shots. Object-based methods segments video into objects and use some of suitable object features for hierarchical indexing of video contents. Motion based approaches use the trajectories of objects to indexing and browsing video contents.

In this paper, we will focus on shot-based video segmentation and video retrieval methods. In shot based video retrieval, finding the interested video scene requires an efficient video shot boundary detection algorithm. Shot is a sequence of frames captured by one camera in a single continuous action in time and space [1]. Shot boundary detection is an important task in managing video database for indexing, browsing and other content-based operations. Video shot boundaries need to be determined possibly automatically to allow content based video retrieval manipulation. Early many researches were mainly focused on extending efficient shot boundary detection algorithms [1]-[7].

Video has both spatial and temporal dimensions and hence a good video retrieval should capture the spatio-temporal contents of the scene. Therefore, when shot boundaries are detected, it is need to extract some spatio-temporal features from shots, which could be utilized to compare visual content of video shots for finding desired video scenes. A number of algorithms for extraction of spatio-temporal features from shots have been reported in the literature [8]-[12]. Successful as they are, existing video retrieval methods are not without problems. For instance, some of these approaches have not appropriate method for extraction of spatio-temporal features from shots. Also, finding desired shots are computationally intensive and extremely sensitive to similarity measure and visual features. These problems have seriously hindered their practical utilities. In view of the aforementioned problems, in this paper, we set out to develop a visual content based video retrieval system based on GED with the following two objectives in mind.

(1). The feature extraction is efficient and use generalized eigenvalue of feature-shot matrices to extract proper spatio-temporal features from shots.

(2). the similarity measure and feature transformation approach is efficient.

The rest of this paper is organized as follows. In Section 2, a brief description of the GED is presented. Section 3 presents our video retrieval system. Section 4 shows empirical evaluations of our solutions and implementations on video retrieval task. Section 5 sets out our conclusions and discusses some future directions to improve the performance of our current solution.

2. Generalized Eigenvalue Decomposition (GED)

In this section, our objective is to review the generalized eigenvalue decomposition (GED) which presented by authors in [13] for rectangular matrices. To commence, we must introduce some concepts.

We know that the studying of eigenvalues and eigenvectors of a square matrix represent valuable and useful information of it. Also, it is well-known that the eigenvalue and eigenvector of a matrix in the usual sense can only be defined for square matrices. In [13], the authors extended these concepts for rectangular matrices based on Radic's determinant [14]. The determinant is a straightforward definition for computing the determinant of non-square matrices which is identical to conventional determinant when matrix is square. It is defined as follows:

Definition 1 (Radic's Determinant [14]). Let $A = [a_{i,j}]$ be an $m \times n$ matrix with $m \leq n$. Determinant of A, is defined as:

$$\det(A) = \sum_{1\leq j_1<\cdots<j_m\leq n}(-1)^{r+s}\det\left[\right. \qquad (1)$$

Where , and . If , then .

This determinant is multilinear operator with respect to the rows vectors, and skew-symmetric and has some important properties such as Laplace's expansion with respect to rows [15]. In [13], it is used to define generalized eigenvalue and eigenvector for rectangular matrices.

Definition 2 (generalized eigenvalue and eigenvector [13]). Let A and B are two rectangular matrices of order $m \times n$ and $m \leq n$. The scaler $\lambda_{A,B}$ and the vector $X \in R^{n\times 1}$ are called generalized eigenvalue and generalized eigenvector of A and B respectively, if and only if satisfying the following conditions:

and .

In [13], the authors have studied some theorems which discuss about the existence and method of calculation of the generalized eigenvalues and eigenvectors. Also, they have presented the Generalized Eigenvalue Decomposition (GED) for rectangular matrices based on these definitions.

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

$$A = B.P.\Sigma.P_r^{-1} \qquad (2)$$

Where $P = [V_1, V_2, \dots, V_m]$ is an $n \times m$ column-orthogonal matrix, $\Sigma = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ is a diagonal matrix, λ_i is the corresponding eigenvalue of V_i for $i \in \{2, 3, \dots, m\}$ and P_r^{-1} of order $m \times n$ orthogonal matrix is the right hand inverse of P. Also, if $\mathrm{rank}(A) = r$, then Σ satisfies:

$$|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_r| > |\lambda,$$

Where $|\lambda_i|$ is the magnitude of λ_i, for $i = 1, \dots, m$.

Proof. See [13].

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 same order. In our needs, we assume the matrix B as always equal to generalized identity matrix which defined as follow:

$$B = I_{m \times n} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \qquad (3)$$

In the following section, the intuitive characteristics of GED for shot boundary detection in a video sequence are studied.

3. Video Retrieval Using GED

In this section, we set out to develop a shot-based video retrieval system using generalized eigenvalue decomposition. The block diagram of the proposed system is shown in Fig.1. It contains the following main components: 1) video database creation; 2) video retrieval.

The video database creation subsystem performs video analysis to efficiently index the video using proper spatio-temporal features. At first, raw video data is segmented into a sequence of video shots using an efficient shot boundary detection algorithm. To date many efficient algorithms have been proposed for shot boundary detection. Here, we used our recent shot boundary detection algorithm to segment video sets into shots [16]. Feature-shot matrices of shots are extracted to feed into the dimension transformation algorithm. The transformed matrices and their correspond shot numbers are inserted in the video database.

The second subsystem performs search and retrieval of desired video scenes from video database. Spatio-temporal features of the shot of user interest are extracted to feed into dimension transformation algorithm. The transformed matrix is used in search algorithm to find desired scenes. The details of the other components of the proposed system are studied in the next sections.

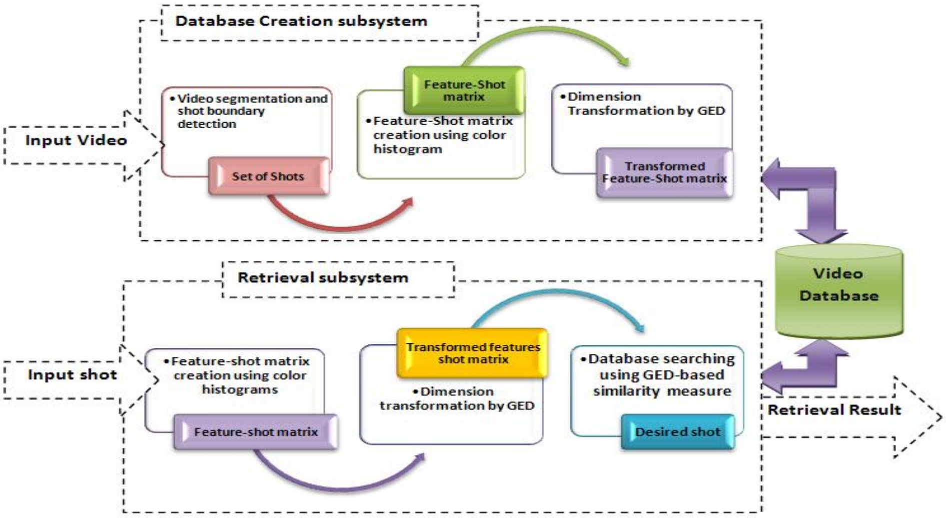
Database Creation subsystem
Input Video
•Video segmentation and shot boundary detection
Set of Shots
Feature-Shot matrix
•Feature-Shot matrix creation using color histogram
•Dimension Transformation by GED
Transformed Feature-Shot matrix
Video Database
Retrieval subsystem
Input shot
•Feature-shot matrix creation using color histograms
Feature-shot matrix
Transformed features shot matrix
•Dimension transformation by GED
•Database searching using GED-based similarity measure
Desired shot
Retrieval Result

Figure 1. The block diagram of the proposed video retrieval algorithm.

3.1. Feature-Shot Matrix Extraction

In order to extract spatio-temporal features of shots, from a wide range of image features, we preferred color histograms to represent each shot frame. Color histograms are very good features for signifying overall spatial features of each frame [17]. The combination of color histogram and GED capture the information of temporary color distribution for each shot. Let $F = \{F_1, F_2, \dots, F_n\}$are sampling set of frames of an arbitrary video shot. To reduce the number of frames to be processed by the GED, we roughly sample the input video sequence with a fixed rate of five frames per second. Our experiments have shown that this sampling rate is sufficient for video programs without many dynamic motions, such as news, documentaries, talk shows, etc. For each frame i in the sampling set, we create an m-dimensional feature vector A_i. Using A_i as column vector i, we obtain the feature shot matrix $A^{(k)} = [A_1, A_2, \dots, A_n]$ for shot k. Now, for computing of feature shot matrix $A^{(k)}$ for shot k, in our system implementation, we create three-dimensional histograms in the RGB color space with five bins for R, G and B, respectively, resulting in a total of 15 bins. To incorporate spatial information of the color distribution, we divide each frame into 2×2 blocks, and create 3D-histogram for each of the blocks. These four histograms are then concatenated together to form a 60 dimensional feature vector for the frame. Using the feature vector of frame i as the i^{th} column, we create the feature-shot matrix $A^{(k)}$ for the video sequence.

3.2. Dimension Transformation Using GED

Obviously, the feature shot matrices of the various video shots will be have different number of columns. It is impossible to compare matrices of different dimension through the use of conventional distance metrics. Therefore, in this section we propose a straightforward technique to transform the feature shot matrices to the matrices of equal dimension by using GED decomposition.

Given an $m \times n$ feature shot matrix A, where $m \leq n$, From theorem 1 the GED of A could be represented as:

$$A = \underbrace{I_{m\times n} P_{n\times m} \Sigma_{n}}_{V} \qquad (4)$$

Hence, by using GED, we will be able to decompose the matrix A into the multiplication of matrix V of order $m \times m$ and matrix T of order $m \times n$. This valuable property of GED could be used to transform matrices of different dimensions to the matrices of same dimensions.

Suppose that $A^{(1)}, A^{(2)}, \dots, A^{(k)}$ are feature shot matrices of k video shots of order $m \times n_1, m \times n_2$, …, $m \times n_k$, respectively. Then, according to Eq.(4), these matrices could be decomposed as:

$$\begin{aligned} A^{(1)}_{m\times n_1} &= V^{(1)}_{m\times m} T^{(1)}_{m\times n_1} \\ &\vdots \\ A^{(k)}_{m\times n_k} &= V^{(k)}_{m\times m} T^{(k)}_{m\times n_k} \end{aligned} \qquad (5)$$

Where, all matrices: $V^{(1)}, V^{(2)}, \dots, V^{(k)}$ are of order $m \times m$. So, these decompositions could be assumed as a mapping of matrices $M^{(i)}$ of order $m \times n_i$ to matrices $V^{(i)}$ of order $m \times m$, with $i = 1, \dots, k$ and their inverses are the matrices $T^{(1)}, T^{(2)}, \dots, T^{(k)}$ respectively. Consequently, by using the Eq.(4), we could transform the feature shot matrices of different dimensions to the matrices of same dimensions. This approach will be used in creating of video databases, indexing of video shots and comparing the similarity of video shots.

3.3. Video Retrieval Algorithm

As demonstrated in Fig. (1), the video retrieval system is composed of database creation and retrieval subsystems.

I. Database creation, which builds a video database from scratch with the following steps: (a) Collect the video sequences to be indexed by the database. (b) Segment each sequence into individual shots and record the boundary of each shot in the database. (c) Create the feature shot matrix for each shot as explained in section (3.1). (d) Compute the GED decomposition of the feature shot matrices and compute the matrices V and T for each feature shot matrix according to the Eq.(4). Store the matrices V and T of each shot into database and use the matrix V as the feature matrix of the shot.

II. Database searching, which find the user interested video shots from the video database by using an appropriate similarity metric. At first, compute the matrices V and T of each shot of input video sequence similar to the database creation subsystem. The similarity metric is defined using the generalized eigenvalues of feature shot matrices. Let $V^{(i)}$ be the feature matrix of shot i which computed according to the Eq.(4), We have:

$$V^{(i)} = \underbrace{I_{m\times n} P_{n\times m}}_{Q_{m\times m}} \Sigma_{m\times m} = Q_{m\times m}\Sigma_{m\times m} = \left[q_1^{(i)}, \dots, q_m^{(i)}\right] \mathrm{diag}(\lambda_1^{(}$$

So, we obtain,

(6)

Now, we define the following matrix norm for $V^{(i)}$:

$$\| \qquad (7)$$

Where the notation $\|.\|_F$, shows the Frobenius vector norm.

Now, suppose that V (i) and V (j) be the feature matrices of input shot and an arbitrary original video shots (in database), respectively. From Eq. (6) we have:

$$V^{(i)} = \left[\lambda_1^{(i)} q_1^{(i)}, \ldots, \lambda_m^{(i)} q_m^{(i)}\right],$$
$$V^{(j)} = [\lambda_1^{(j)} q_1^{(j)}, \ldots, \lambda_m^{(j)} q_m^{(j)}]$$

By using Eq. (7), we compute the norms of these matrices. Now, we define the following distance metric:

$$D^{(T)}\left(V^{(i)}, V^{(j)}\right) = \frac{\min\{\|V^{(i)}\|^{(T)}, \|V^{(j)}\|^{(T)}\}}{\max\{\|V^{(i)}\|^{(T)}, \|V^{(j)}\|^{(T)}\}} \; ; \qquad (8)$$

The result of comparison by using this distance is a value between zero and one. For two very similar shots, the distance value will be closed to one and for two shots with different visual contents the distance will be closed to zero.

4. Experimental Results

In order to validate the effectiveness of the proposed shot boundary detection algorithm, some of experiments are presented in the following.

The video types and the number of frames for each video are summarized in Table 1. Among the videos, the News is from MPEG-7 test set [34], the Cartoon is from an episode of "Tom and Jerry," the Movie is from the feature film "You've Got Mail," and the rest are from broadcast TV programs including sports and documentaries.

Table 1. The video Types that have been used in the experiments.

Video Type	Number of Frames
News	95743
Cartoon	74384
Movie	85958
Sport	109381
Documentary	57491

The video retrieval algorithm support user interactive video search. In this search the user must present a sample shot of the desired video scene. The algorithm matches the sample shot with all the shots stored in database by using Eq.(8) and retrieves 8 shots whose similar to input shots.

The video retrieval system is evaluated using two common measures, recall and precision, which are defined as follows:

- The Recall measure, also known as the true positive function or sensitivity, which corresponds to the ratio of correct experimental detections over the number of all true detections. It measures the ability of a system to present all relevant items:

$$Recall = \frac{number\ of\ r}{number\ of\ rel} \qquad (9)$$

- The Precision measure defined as the ratio of correct experimental detections over the number of all experimental detections. It measures the ability of a system to present only relevant items.

$$Precision = \frac{number\ of\ r}{total\ num}, \qquad (10)$$

The values of recall and precision is in the range of [0 1]. A high recall indicates the capability of retrieving correct shots, while a high precision indicates capability of avoiding false matches.

In order to verify the best value for T in Eq. (8), we computed the precision and recall for different values of T. Fig. (2) shows the evaluation result with the value of T as a parameter. It can be seen from the figure that when T equal 36, both recall and precision reach their maximum.

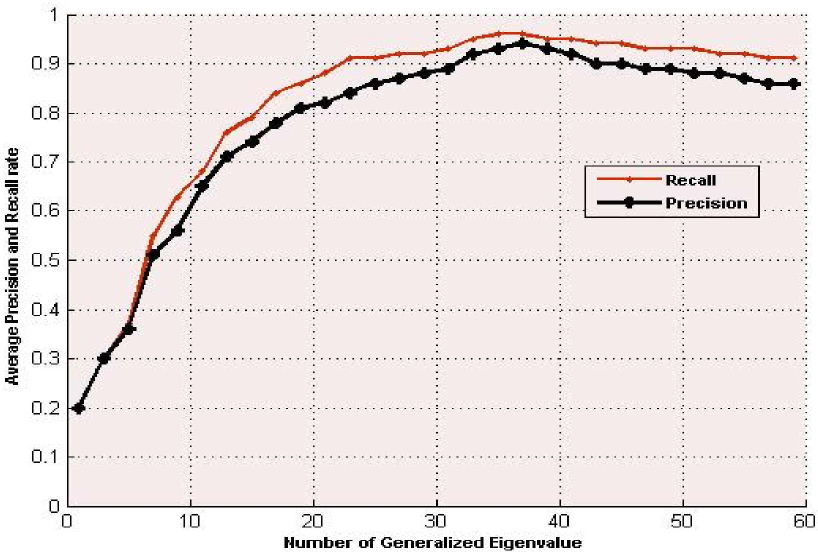

Figure 2. Average Precision and Recall with the value of T as a parameter

Therefore, base on above experiments, we set the value of T to 36. Table 2 shows experimental results using the test video set. By exploring the results, we achieved that many of false detection is due to a lot of .ash effects in a shots and high speed object motions. But, it does not influence on the efficiency of our system, because there is not more flash effects and high speed motions in the video sets that captured from the real world.

Table 2. Experimental results for video retrieval using proposed system

Video Type	*Precision*	*Recall*
News	92.5	95.6
Cartoon	91.4	95.2
Movie	89.0	94.5

Sport	90.5	93.3
Documentary	93.3	96.6
Average	91.3	95.0

Also, for comparison, we implemented the same video retrieval system using raw feature-shot matrices without dimension and feature space transformation phase. We employed Frobenius norm of feature-shot matrices in Eq. (8) to compute similarity of two different shots of video. Table 3 shows experimental results using the test video set for this case.

Table 3. Experimental results using raw feature-shot matrices without using feature transformation and proposed distance metric

Video Type	*Precision*	*Recall*
News	75.9	80.5
Cartoon	64.5	73.6
Movie	62.8	71.0
Sport	71.5	78.4
Documentary	65.3	70.5
Average	68.0	74.8

It is clear that the video retrieval system has obtained reasonable performance.

5. Conclusion

In this paper, a novel video retrieval algorithm is developed based on the Generalized Eigenvalue Decomposition (GED). We propose a system that is able to retrieve video shots according to their amounts of visual similarities, color distribution uniformities, and visual changes. The experimental results confirm its high efficiency. Also, more work for video indexing, summarization, classification and data dimension reduction based on generalized eigenvalue decomposition will be reported along this line in the near future.

References

[1] J. Yuan, H. Wang, L. Xiao, W. Zheng, J. Li, F. Lin, and B. Zhang, A Formal Study of Shot Boundary Detection, IEEE Transaction on Circuits and Systems For Video Technology, 17(2)(2007), pp. 168-186.

[2] M. Cooper, T. Liu, and E. Rieffel, Video Segmentation via Temporal Pattern Classification, IEEE Transaction on Multimedia, 9(3)(2007), pp. 610-618.

[3] C. Grana, and R. Cucchiara, Linear Transition Detection as a Unified Shot Detection Approach, IEEE Transaction on Cicuits and Systems For Video Technology, 17(4)(2007), pp. 168-186.

[4] X. Qian, G. Liu, and R. Su, Effective Fades and Flashlight Detection Based on Accumulating Histogram Difference, IEEE Transaction on Circuits and Systems For Video Technology, 16(10)(2006), pp. 1245-1258.

[5] J. Nam, and A. H. Tewfik, Detection of Gradual Transitions in Video Sequences Using B-Spline Interpolation, IEEE Transaction on Multimedia, 7(4)(2005), pp. 667-679.

[6] A. Hanjalic, Shot-Boundary Detection: Unraveled and Resolved, IEEE Transaction on Circuits and Systems For Video Technology, 12(2)(2002), pp. 90-105.

[7] W. J. Heng, and K. N. Ngan, Shot Boundary Refinement for Long Transition in Digital Video Sequence, IEEE Transaction on Multimedia, 4(4)(2002), pp. 434-445.

[8] X. F. Yang, Q. Tian, and P. Xue, E.cient Short Video Repeat Identification With Application to News Video Structure Analysis, IEEE Transaction on Multimedia, 9(3)(2007), pp. 600-609.

[9] N. M. Loccoz, E. Bruno, and S. M. Maillet, Interactive Retrieval of Video Sequences from Local Feature Dynamics, Lecture Notes in Computer Science , 3877(2006), pp. 128-140.

[10] H. Lu, B. C. Ooi, H. T. Shen, and X. Xue, Hierarchical Indexing Structure for E.cient Similarity Search in Video Retrieval, IEEE Transaction on Knowledge and Data Engineering, 18(11)(2006), pp. 1544-1559.

[11] J. Shao, Z. Huang, H. T. Shen, X. Zhou, E. P. Lim,and Y. Li, Batch Nearest Neighbor Search for Video Retrieval, IEEE Transaction on Multimedia, 10(3)(2008), pp. 409-420.

[12] C. G. M. Snoek, B. Huurnink, L. Hollink, M. D. Rijke, G. Schreiber, and M.Worring, Adding Semantics to Detectors for Video Retrieval, IEEE Transaction on Multimedia, 9(5)(2007), pp. 975-986.

[13] A.Amiri, M.Fathy, "Generalized Eigenvalue Decomposition Based on Eigenvalue and Eigenvector of Rectangular Matrices and its Application to the Video Retrieval" submitted to the EUROSIP Journal of Mathematical Problems in Engineering, 2009.

[14] M. Radic, "Areas of Certain Polygons in Connection with Determinants of Rectangular Matrices" Beitrage zur Algebra and Geometric Contributions to Algebra and Geometry, vol.49, 2008, pp. 71-96.

[15] M. Radic, "About a Determinant of Rectangular 2×n Matrix and its Geometric Interpretation" Beitrage Zur Algebra und Geometric Contributions to Algebra and Geometry, vol.46, 2005, pp. 321-349.

[16] A. Amiri, and M. Fathy, Video Retrieval Using Attributed Relational Graphs and Singular Value Decomposition, Iranian Conference on Machine Vision and Image Processing, (2008), pp. 456-467.

[17] W. Cheng, D. Xu, Y. Jiang, and C. Lang , "Information Theoric Metrics in Shot Boundary Detection" LNAI, vol.3683, 2005, pp.388-394.

Authors

Mahmood Fathy, was born in Tehran, Iran, in 1959. He received the B.S. degree in Electronic Engineering from Iran University of Science and Technology (IUST), in 1985, M.Sc. degree in Microprocessor Engineering from Bradford, UK, 1988 and Ph.D. degree in Image Processing and Processor Design from UMIST, UK in 1991.

Since 1992, he has been with the IUST, where he is currently Associate Professor at the Department of Computer Engineering. His research interests include Computer networks, QOS , internet Engineering, Application of image processing in Traffic, Computer Architecture for image processing, video processing applications, Panorama, Supper resolution, video classification, video retrieval and summarization.

Ali Amiri was born in Zanjan, Iran, in 1982. He received the M.in S. degree in Computer Engineering from Iran University of Science and Technology (IUST), in 2006, where he is currently pursuing Ph.D. degree in the IUST. His research interests include video segmentation, video retrieval and summarization, Matrix Computation and moving object detection and tracking.