A Proposed Strategy for Segmenting the Video Clips into Shots and Selecting the Master Frame

Soukaena Hassan Hashem
University of Technology, Baghdad, Iraq

Received 7/7/2007 – Accepted 16/9/2007

Keywords: Master frame, video clips, clustering, genetic algorithm, and retrieval and mining system

الخلاصة

لبناء فهرسة قواعد البيانات الخاصة باسترجاع وتعدين الفيديو كليب من اكبر التحديات التي ستواجهها هذه الانظمة هي عملية تحديد مقاطع الفيديو ثم اختيار افضل لقطة صورية تمثل المقطع لكل الفيديو كليب.
هذا البحث يقترح طريقة لتقطيع الفيديو كليب الى لقطات صورية مباشرة ثم تطبيق خوارزمية العنقدة لجمع كل اللقطات الصورية المتشابهه ضمن مقطع واحد (عنقود واحد) وبهذا المقطع سوف لن يكون فقط سلسلة من اللقطات الصورية المتتابعة بل ممكن ان تحتوي على لقطات صورية من اماكن مختلفة من الفيديو كليب. هذا سوف يقلص عدد المقاطع ويجعلها اكثر اعتمادا وامثل. بعدها سيتم اختيار اللقطة الاكثر ملائمة لتمثيل كل مقطع من خلال اقتراح خوارزمية باستخدام الخوارزميات الجينية كوسيلة للاختيار من خلال تمثيل كل لقطة كنقطة.

ABSTRACT

To build indexes databases for retrieval and mining video clips. The most challenge will be the detection of shots and extraction of the master frames of each shot from video clips.

This research proposed a method to segment the video clips directly to frames, then applying clustering technique to collect the so much similar frames into shots (clusters), so the shot will not be sequential frames only but may have frames from other places of video clip, that to optimize and reduce the no. of shots. Then that research will propose a strategy to select the master frame in each shot using genetic algorithms, that by represent each frame as a point.

INTRODUCTION

Currently text-based search engines are commercially available, and they are predominant in the *World Wide Web* for search and retrieval of information. However, demand for search and mining multimedia data based on its content description is growing. Search and retrieval of contents is no longer restricted to traditional database retrieval applications. As an example, it is often required to find a video clip of a certain event in a television studio. In the future the content customers will demand to search and retrieve video clips based on content description in different forms. It is not difficult to imagine that one may want to mine and download the images or video clips containing the presence of Mother Teresa from the Internet or search and retrieve them from a video archival system. It

is even possible to demand for retrieval of a video which contains a tune of a particular song. In order to meet the demands for retrieval of audio-visual contents, there is a need for efficient solution to search, identify and filter various types of audio-visual content of interest to the user using non-text based technologies [1-5]

RELATED WORKS

Rretrieve and mine the video clips depending on MPEG, concentrate only with the metadata which has at most the general information of the feature and keywords of sound in the related video clip, the MPEG (Moving Picture Expert Group) standard committee, under the auspices of the International Standard Organization, is engaged in a work item to define a standard for multimedia audio-visual content description interface. JPEG2000 is the new standard for still picture compression and has been developed in such a way that metadata information can be stored in the file header for access and retrieval by users as well. There is a mode in JPEG2000 standard which particularly focuses on compressing moving pictures or video and its content description. All these developments will influence effective mining of video data in the near future, [6].

Traditionally, in general, the retrieval and mining system video clips temporarily segmented into video shots. A shot is a piece of a video motion (a sequential group of frames or pictures) where the video content from one frame to the adjacent frames does not change abruptly. One of these frames in a shot is considered to be a master frame. This master frame is considered to be a representative for the picture content in that shot (the selection of master frame differs from one system to another may be randomly for example). Sequence of master frames can define the sequence of events happening in the video clip. This is very useful to identify the type and content of the video, see figure (1), [7, 8].

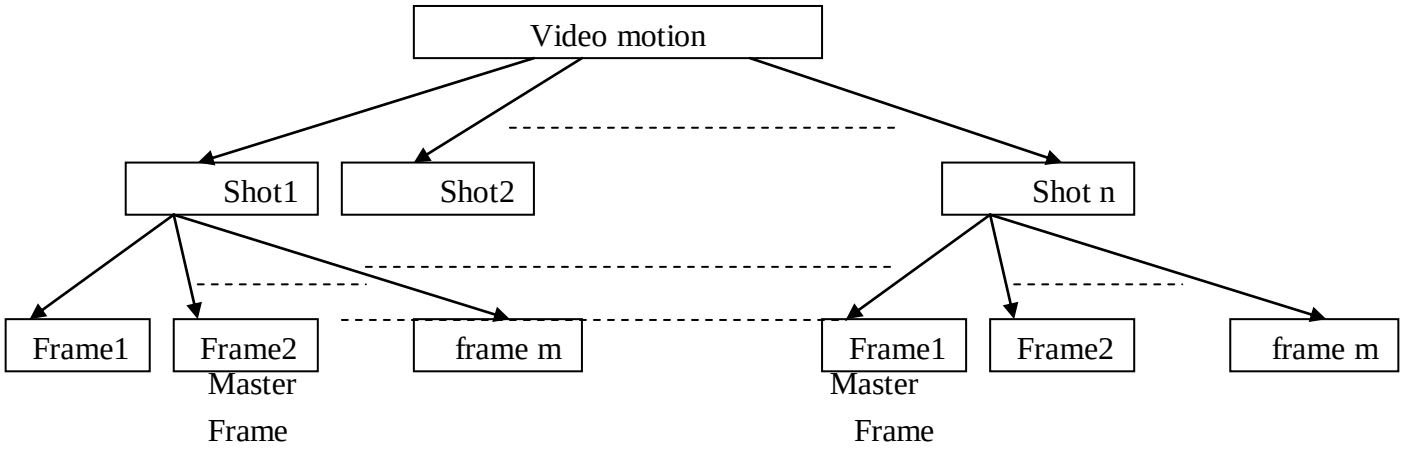

Figure -1: the main architecture for video Segmentation to shot, then frames and detecting the master frame for each shot.

THE PROPOSED SYSTEM

In this research we will propose to segment the video clips to frames directly so the video clip will has huge no. of frames each frame is a picture (image), see figure (2).

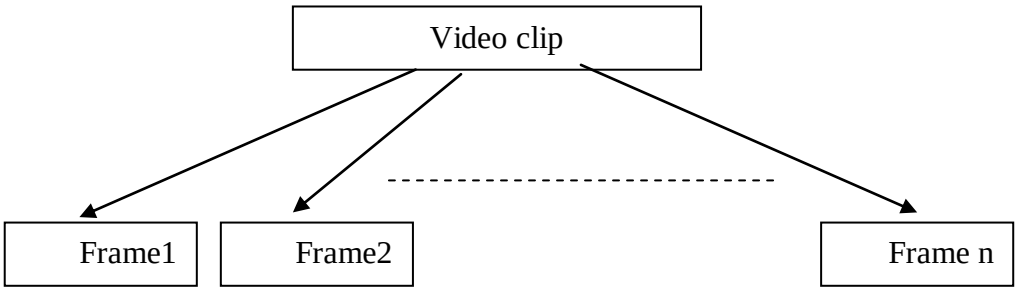

Figure -2: the proposed architecture for video Segmentation to frames directly without segmenting the clips to shot the frames.

After segmentation process we suggest to take each frame, image, and extract its feature vector, the images in an image database are indexed-based on extracted inherent visual contents (or features) such as:

1. **no. of objects: to detect the number of objects in each image we use edge detection filters (for more details see [9]).**
2. **then extract the features for each object such as position (x, y-coordinates), which mean calculate coordinates of the center for the specific object using normal image pixels coordinate from left to right and top down.**

3. **Color (color histogram, color coherence vector, color moment , and linguistic color tag), for a specific image (m x n) we use the following equations for extracting all color features:**

$$\mu_c = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij}^{c},$$

$$\sigma_c = \left[\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(p_{ij}^{c} - \mu_c)^2\right]^{\frac{1}{2}},$$

$$\theta_c = \left[\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(p_{ij}^{c} - \mu_c)^3\right]^{\frac{1}{3}},$$

where *p^* is value of the cth color component of the color pixel in the ith row and *jth* column of the image. As a result, M, N are dimension of Image, I and j are the counter of dimension is the standard deviation we need to extract only nine parameters (three moments for each of the three color planes) to characterize the color image.[2]

4. **texture, for a specific image (m x n) we use the following equations for extracting all texture features:**

$$Energy = \sum_{i}\sum_{j} C^2(i,j),$$

$$Contrast = \sum_{i}\sum_{j}(i-j)^2 C(i,j),$$

$$Homogeneity = \sum_{i}\sum_{j}\frac{C(i,j)}{1+|i-j|}.$$

5. **Object shape and topology. for a specific image (m x n) we use the following equations for extracting all texture features: The *Euler number* is defined as the difference between number of *connected components* and number of *holes* in a binary image. Hence if an image has *C* connected components and *H* number of holes, the *Euler number E* of the image can be defined as [2]**

E = C-H.

The feature vector actually acts as the *signature* of the image, figure (3) represent the main architecture to build indexes database for all frames, images, of the video clips. Table (1) represent the indexes database for each video clip and corresponding all frames of that clip.[2]

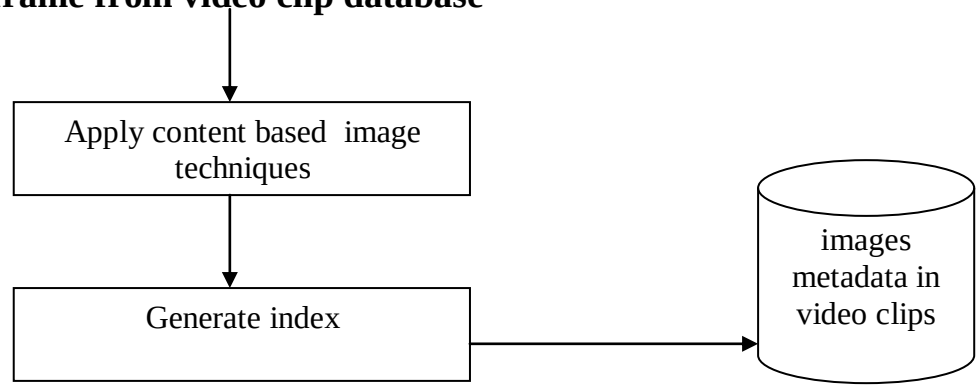

Figure -3 : the main architecture to build indexes database for the frames of the video clips.

Table -1: the indexes dbase for video clips and corresponding frames of that clip.

frame ID	No. of objects	Object ID	Features of objects x y histo mom coher tag texture shape topology
I1	4	O1	50 40 116.84 122.84 117.84 116.86 0.87 0.58 0.61
		O1	
		O2	
I2		O3	
I3			

THE PROPOSED CLUSTERING ALGORITHM

The steps of similarity algorithm are searching by no. of object in frames databases and the proprieties. The Euclidian distance *(∑(x – yi))* is used to take the minimum distance between the frame and all the frames of one cluster.

Input: Take frame from frames database and set it to first created cluster.

Output: put each frame to the cluster (shot) similar to it, see figure (4).

Step1: checking the no. of object and object properties in the taken frame with each frame in each cluster, shot, then record the degree of similarity with each cluster.

Step 2: put the frame in the shot which has a minimum degree of differential with it is frames.

Step 3: take the next frame and go to step 1.

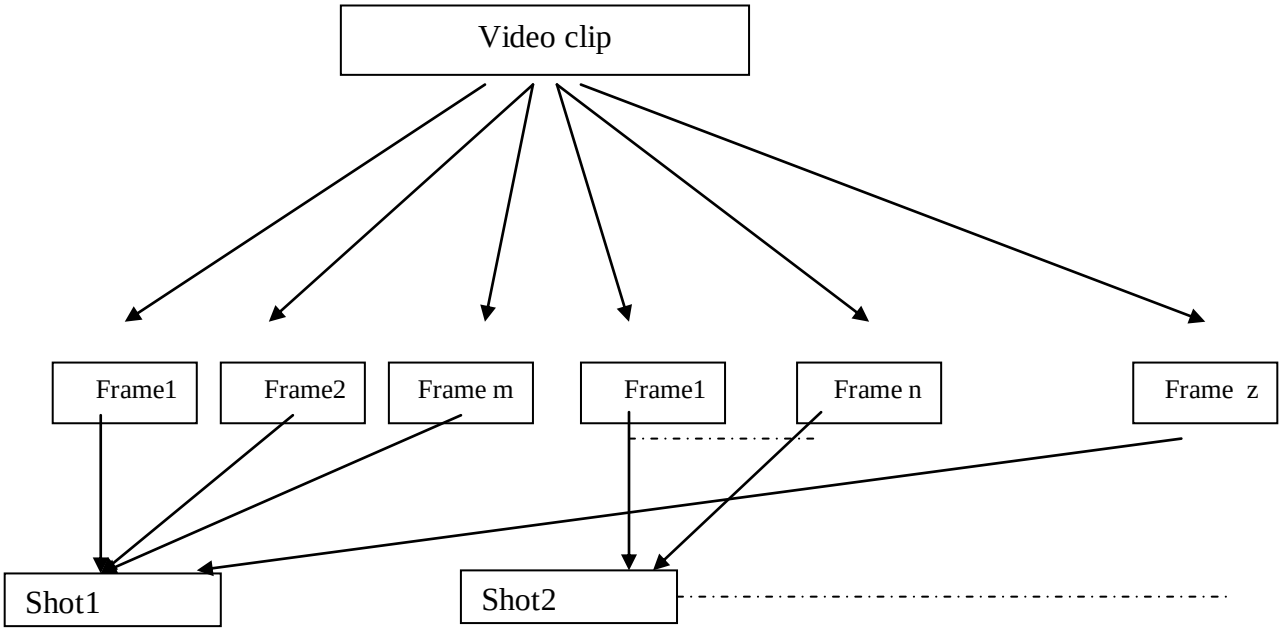

Figure -4: distributing frames to shots

THE PROPOSED GENETIC ALGORITHM

To apply a genetic algorithm for problem selecting master frame from shot, this research propose to define or to select the following components, for more explanation see the flowchart (figure 5):

Note: o represent symbol of object and oi represent the object o has the ith order.

1. **A genetic representation or encoding schema for potential solutions to the problem, here each frame will be presented as a point each point consist from the following (*no. of objects, (o1, (position (x, y coordinates), color (color histogram, color coherence vector, color moment , linguistic color tag), texture, object shape and topology), o2(…..), on(……)).* For example the frame n in shot m has the following point representation (3, *(o1, (position (450,300), color (*118.84 122.84 117.84 116.86 *),***

0.80 , 0.68 *and* 0.71*), (o2, (position (450,400), color (*116.84 123.84 117.84 116.86 *),* 0.66 , 0.58 *and* 0.61*), (o3, (position (550,300), color (*116.84 122.84 117.84 117.86 *),* 0.99 , 0.58 *and* 0.61*)).*

2. **A way to create an initial population of potential solutions, the initial population already created with clustering algorithm which established the shots. So this mean the initial population of each shot, will be all its frames represented by points.**
3. **An evaluation function that plays the role of the problem environment (best frame), rating solutions in terms of their "fitness". Here the proposed evaluation function for each frame is *f(point) = (no. of object + ∑ (features of each objects))*.**
4. **Genetic operators that alter the composition of offspring. *One-point crossover* is the most basic crossover operator, where a crossover point on the genetic code is selected at random, and two parent frames are interchanged at this point.**
5. **Crossover exploits existing frame potentials, but if the population does not contain all the encoded information needed to find the best frame, no amount of frames mixing can produce a satisfactory solution. For this reason, a mutation operator capable of spontaneously generating new frame is included. The most common way of implementing mutation is to flip a bit with a probability equal to a very low, given mutation rate (MR). A mutation operator can prevent any single bit from converging to a value through the entire population and, more important, it can prevent the population from converging and stagnating at any local optima.**
6. **Values for the various parameters that the genetic algorithm uses population size, rate of applied operators, etc..In our particular problem we use the following parameters of the genetic algorithm: Population size, *pop-size* = 400 (the parameter was already used), Probability of crossover, PC = 1, Probability of mutation, PM = 0.001 (the parameter will be used in a mutation operation).**
7. **Continue with genetic processing until obtain the optimized frame to be the master frame**

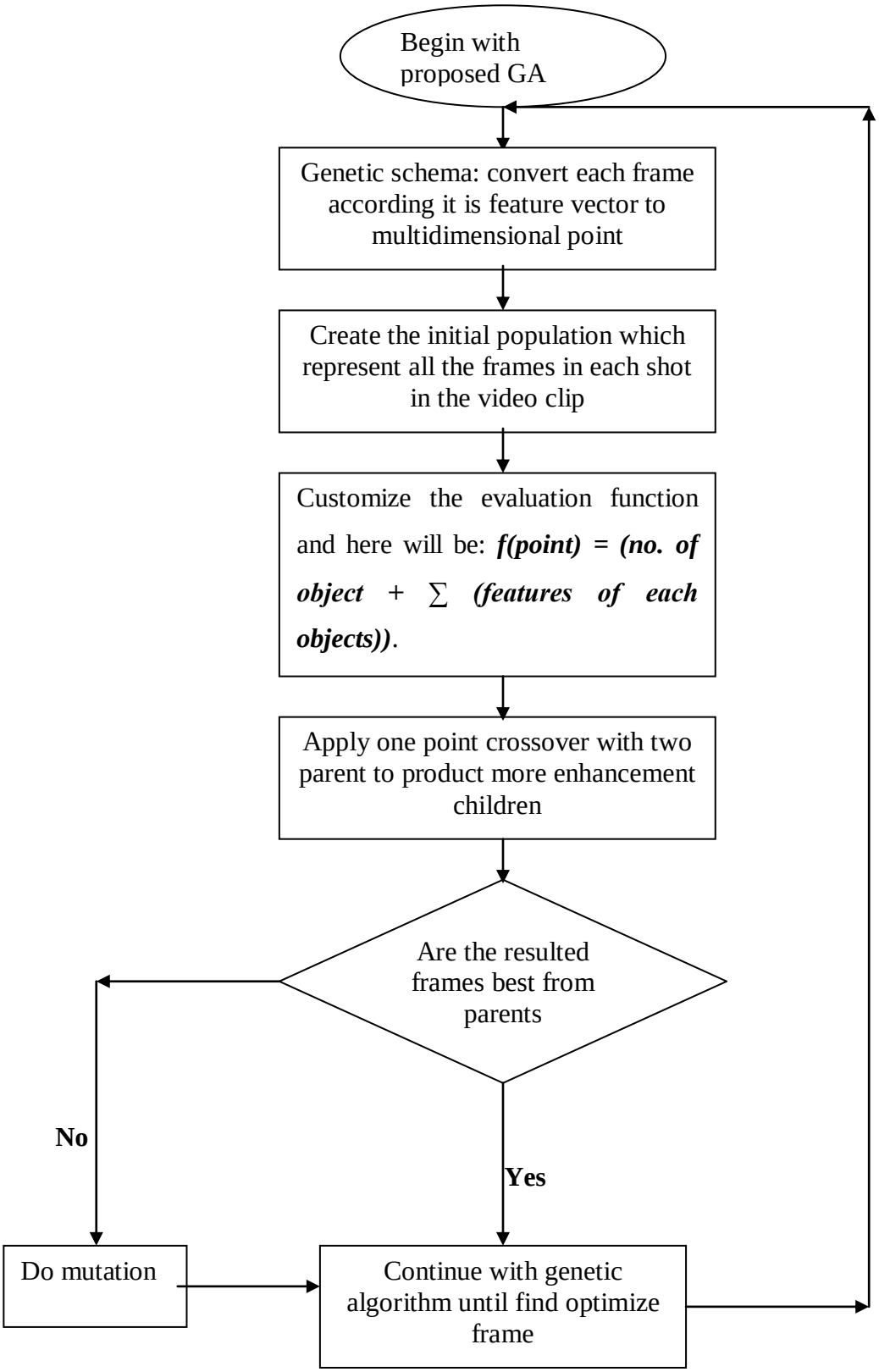

Figure-5 : the general flowchart for selecting best frame using genetic algorithms.

THE IMPLEMENTATION OF THE PROPOSED SYSTEM

The implementation of the proposed system will take each video clip will be introduced by the administrators and then analyze the video clip into collection of sequenced frame as in figure (6).

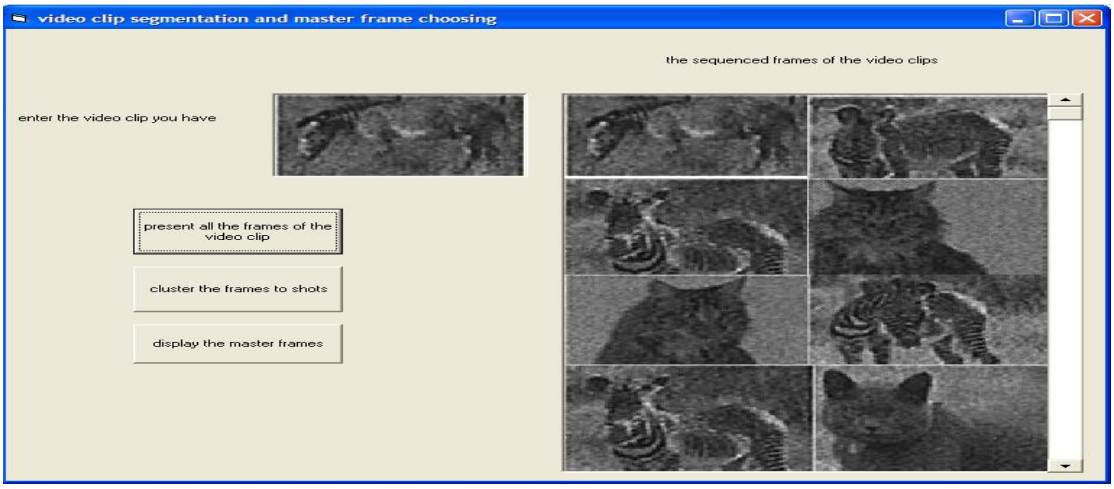

Figure -6 : The main window for introducing the video clip and analyze it.

In figure (6) there are three commands, the first one will be clicked when the administrator present the desired video clip the clicking process will introduce the sequenced frames of the clip. Where clicking the second command will display small window which introduce the results of the clustering algorithm which applied on the sequenced frames to introduce optimized shots, see figure (7).

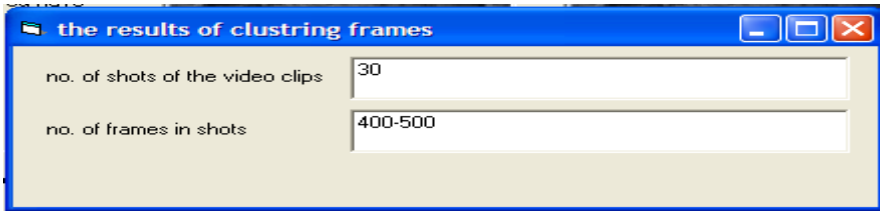

Figure -7 : The window which display the results of clustering.

Finally clicking the third command will display the results of the selecting the master frame from the shot. This window introduce the results of selected frame from shot (according the proposed genetic algorithm) for the two first shot only since all the results already stored in related database, see figure (8).

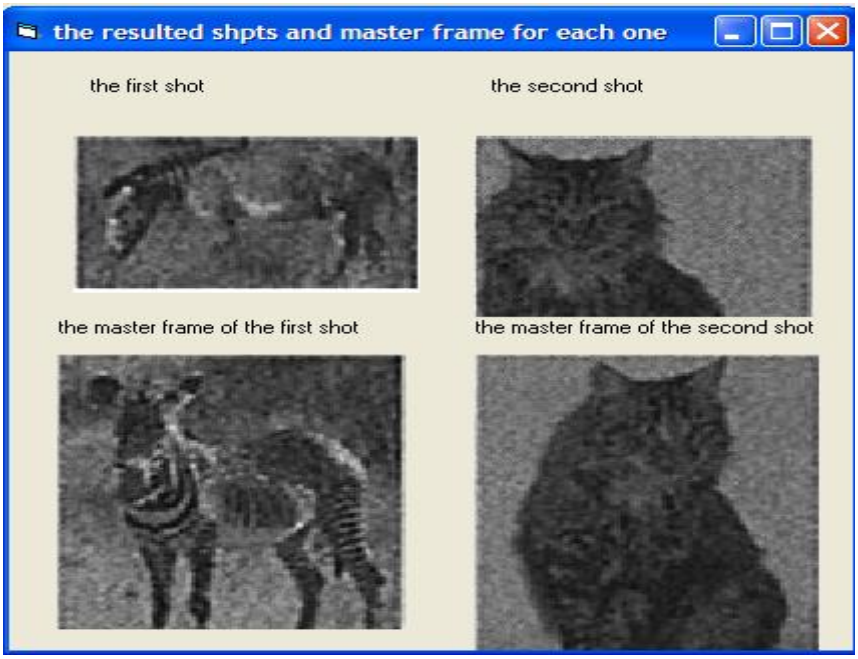

Figure 8 : The window which display the results of selected master frames for the two first shots.

CONCLUSIONS

In context with the results of the present study it can be concluded that:

1. **The proposed segmentation of the video clip directly into frames instead of the traditional method (shots then frames and some times senses then shots then frames) will optimize the no. of shots since, the shots will be resulted from applying clustering algorithm on the sequenced frame. This mean we could obtain a shot has the first 200 frame and the last 50 frame that will reduce the no. of shot and reduce the redundancy in different shots.**
2. **Using the proposed genetic algorithm to get the master frame from each shot will give the optimal result. Since the proposed fitness function deal with each frame as a point and each point represent the overall features contained in the frame so surely the best point will be the best frame.**
3. **Since GAs are parallel-search procedures that can be implemented on parallel-processing machines for massively speeding up their operations.**

REFERENCES

1. **Kantardzic M.; *"DM concepts, models, methods and algorithms"*, jhon wiley &Sons, (2003).**
2. **Vailaya A., Figueiredo A. T., Jain A. K., and Zhang H. J., *"Image Classification for Content-Based Indexing", IEEE Transactions on Image Processing*, Volume: 10 Issue: 1, pp 117 –130, (2001).**
3. **C. Saraceno and R. Leonardi, " audio as a support to scene change detection and characterization of video sequence", in proceeding of the IGASSP, IEEE computer society press, (1997).**
4. **Sakurai S., Ichimura Y., Suyama A., and Orihara R.; *"Inductive learning of a knowledge dictionary for a text mining system,"* in Proceedings of 14th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems,(2003).**
5. **Tan K. L., Ooi B.C. and Thiang L. F., *"Retrieving Similar Shapes Effectively and Efficiently", Multimedia Tools and Applications*, Kluwer Academic Publishers, The Netherlands, (2001).**
6. **Oh J., Bandi B., *"Multimedia Data Mining Framework For Raw Video Sequences"*, Proceedings Third International Workshop on Multimedia Data Mining MDM/KDD', 23rd 2002, Edmonton, Alberta, Canada.**
7. **K. Minami, A. Akutsu, and H. Hamada, "video handling with music and speech detection", IEEE multimedia, pp: 17-25, (1998).**
8. **Y. Tonomura, A. Akutsu, Y. Taniguchi, and G. Suzuki, "structured video computing", IEEE multimedia, pp: 34-43, (1994).**
9. **Gonzalez R. C. and Woods R. E., *"Digital Image Processing"*, Reading, MA: Addison-Wesley,(1993).**