KEY FRAME SELECTION FROM VIDEO BASED ON WEIGHTED MINKOWSKI DISTANCE

ARUN KUMAR GAUTAM

School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi
E-mail: arungautam.cs@gmail.com

Abstract- Key frames are extracted to make video processing more efficient and effective. The key frames are those frames which are selected from the total frames of a video that can represent the information of whole video. Basically a video contains hundred, or thousands, or millions of frames depending upon the size and frame rate. The previous approaches have been either extracted a single frame or extracted those frames which are visually different. This paper have been used the Minkowski as a method to measure the distance between the two images and compares them with a threshold value to find the key frames. In this paper the standard deviation of the differences of images has been used asa threshold to determine the key frames.

Keywords- Frame selection, Minkowski distance, Shot detection, Shot selection

I. INTRODUCTION

This paper introduces the technique for the selection of key frame from video. Videos can be represented by a hierarchical structure consisting of five levels (video, scene, group, key frame, and shot) from top to bottom increasing in granularity, while a shot is the basic unit of a video. Video files consists of millions of frame or thousands of frame, depends on the compression and size of the files. The key frames are considered as the most common frame occurring in the video. In other words, frames which contains most of the information. The information may be in the form of objects, change in the corresponding images. These key frames required because of the interest of people for accessing the video on demand. The frame extracted from the video are large in number among which some frames are exactly similar to another one with a minor difference that cannot be seen. So that if it is possible to select some of the frames which act as key frames for the video and can be processed easily.

For efficient and effective video content analysis the key frame selection could be quite useful. Due to temporal nature of the video it contains redundant information among those frames. Basically a video is structured of a number of video shots and each shot is a sequence of frames. A clustering approach is been discussed in. This paper illustrate about the clustering approach which are based on local and global feature of the images. It elaborated the concept of clustering to bring out similar frame together and select the centre point of each cluster as a key frame. Two adjacent frames are compared based on some threshold value, which is quite challenging to set. The key point based approach is applied for key frame selection in. It comprises of two steps, in first step the key point of the whole shot based on visual descriptor is been collected from each frame and in second step a global pool of unique key point is created to represent the whole video shot. Finally, the frames which cover most of this global pool of unique key are selected as key frame.

In this paper first the distance of images in terms of minkowski has been measured and based on this measurement some threshold value are fixed based on which the adjacent frames are compared. The two frames which are likely to have the similar properties (distance) are omitted and the comparison will be done until the frame which will result greater in distance. That frame could be selected as key frame.

II. RELATED WORK

Multimedia mining is a major area of research. In multimedia, videos are most difficult task for processing. The video is basically the sequence of frames. These frames can be of thousand in a few minutes of video depending upon the frame rate and the size. The content of the consecutive frames are almost similar up to an extent. Here is some of the literatures which explains the key frame selection from the videos.

In, an automatic extraction of single key frame has been carried out from a sequence of video. This is comprised of three basic techniques i.e. shot boundary detection, shot detection and key frame extraction within the selected shots. Shots are basically a collection of frames within a window.
Shot Detection: Shot detection is the first stage of video analysis. First, the motion activity has been measured by pixel-wise frame difference to detect the shot in video.

Shot Selection: After calculating motion activity by frame difference FD,spatial activity is computed by entropy of pixel value distribution.
Key Frame Selection: Most of the video summaries based on selection of key frames within the shots of a

video. The final selection of frames has been carried out by calculating the difference between entropy and the frame difference i.e. score of frames.

In [3], the keypoint-based frame selection has been carried out for an efficient video content analysis. The paper proposed keypoint-based framework for the selection of key frames. The selected key frames should both be representative of video content and containing minimum redundancy. The paper has used two techniques one is coverage and second is redundancy. A video shot was first represented by a global pool of keypoints through keypoint chaining. Second, a greedy algorithm was developed to select suitable keyframes based on the two intuitive metrics of coverage and redundancy.

III. MINKOWSKI DISTANCE

The distance measures are very useful techniques that have been used in a wide range of applications such as fuzzy set theory, decision making, operational research, etc. Minkowski distance is one of the main distance measures because it generalizes a wide range of the other distances such as hamming distance, Euclidean distance and Manhattan distance[1]&[9].

$$d(i,j) = (|x_{i1} - x_{j1}|^p + |x_{i2} - x_{j2}|^p + \cdots + |x_{in} - x_{jn}|^p)^{1/p}$$

Where, p is a positive integer. It represents Manhattan distance when p=1 and Euclidean distance when p=2. If each variable is assigned with a weight according to its perceived importance, the weighted Minkowski distance will be

$$d(i,j) = (w_1|x_{i1} - x_{j1}|^p + w_2|x_{i2} - x_{j2}|^p + \cdots + w_m|x_{in} - x_{jn}|^p)^{1/p} \quad (1)$$

The generalization of the formula will be:

$$d(i,j) = \left(\sum_{k=1}^{n} w_k(x_i(k) - x_j(k))^p\right)^{1/p} \quad (2)$$

Where w is the weight associated with each object differences which reflects the relative importance of each objects.

Minkowski Distance for Images:

The frames of the videos are considered as objects. So the formula for calculating distance between two images or frames i.e. d(x,y) will be:

$$d(x,y) = \left(\sum_{i,j=1}^{mn} w_{ij}((x_i - y_i)(x_j - y_j))^p\right)^{1/p} \quad (3)$$

The metric coefficients w_{ij} where i,j= 1,2, , MN, is the weighting factor changed by heuristically and can also be calculated as [6]:

$$w_{ij} = < e_i, e_j > = \sqrt{< e_i, e_i >}\sqrt{< e_j, e_j >}.\cos\theta_{ij}$$

where <, > is the scalar product, and θ_{ij} is the angle between e_i and e_j.

Here are the steps for processing key frames in this paper:

- **A.** Extract all frames from videos.
- **B.** Load the databases i.e. collection of images
- **C.** Pre-process them to convert them grayscale images
- **D.** Calculate the difference between the images using Minkowski distance for p=3.
- **E.** Compare them with the threshold to choose the key frames.
- **F.** Calculate the reduction of frames.

IV. EXPERIMENTAL SETUP AND RESULT ANALYSIS

Experimental setup has been carried out on twelve datasets. Each dataset contains the .jpeg images of different resolutions. Dataset1 is been a movie clip from bollywood movie don, dataset2 and dataset3 is a documentary video form NASA. Dataset4, dataset5, dataset6, datset7, dataset8, dataset9, dataset10, dataset11, and dataset12 are obtained from and of different genre. The details of twelve dataset videos information is given below in table1. The whole setup and program has been implemented in Matlab software. The threshold to be compared with the differences of images is chosen by the mean and standard deviation. This method has subsequently reduced the number of frames as the original number of frames is too much. Since the number of frames is reduced, it is easier to process less number of frames rather than whole. The number of key frames extracted is given below in table 2: Figure-1 shows all the key frames extracted from the 12 databases. In this figure the first column and the last column signifies the first and last key frame detected from the databases while the other column shows the randomly picked intermediate key frames in serial, in between first and last key frame of the databases

Table1: Video information

Sr. No.	Name of video and type	Size (in MB)	Duration of videos (in seconds)	No. of frames	Frame Resolution (height × width)
1.	Nasa video (.mpg)	5.27	30	914	240 × 320
2.	Movie clip (.avi)	4.95	60	1,509	240 × 624
3.	Trackvideo (.mpg)	10.3	75	2,264	240 × 320
4.	Adam (.wmv)	9.73	138	3,375	240 × 320
5.	Exotic_flowers (.mpg)	17.4	120	3604	240 × 352
6.	Texas-short (.mpg)	17.7	132	3,972	240 × 320
7.	Indians (.mpg)	25.6	148	4461	240 × 320
8.	Indians_1(.mpg)	29.6	171	5151	240 × 320
9.	Bluelesson(.mpg)	49.9	301	7534	288 × 352
10.	Nasa 2(.mpg)	68.9	400	11,995	240 × 320
11.	Robotsmall (.avi)	8.13	480	1,2005	240 × 320
12.	Minerva (.mpg)	70.0	478	14,331	240 × 352

Table2: Total Key Frame Reduction of the databases

Sr. No.	Name of Video	Total number of frames	Selected Key frames	% reduction of frames
1.	Nasa video (.mpg)	914	92	89.93
2.	Hindi Movie clip (.avi)	1,509	233	84.55
3.	Trackvideo (.mpg)	2,264	279	87.67
4.	Adam (.wmv)	3,375	282	91.64
5.	Exotic_flowers	3604	1622	54.99
6.	Texas-short (.mpg)	3,972	488	87.71
7.	Indians (.mpg)	4461	427	90.42
8.	Indians_1 (.mpg)	5151	543	89.45
9.	Bluelesson (.mpg)	7534	622	91.74
10.	Nasa Video 2 (.mpg)	11,995	804	93.29
11.	Robotsmall (.avi)	1,2005	19	99.84
12.	Minerva (.mpg)	14,331	1039	92.75

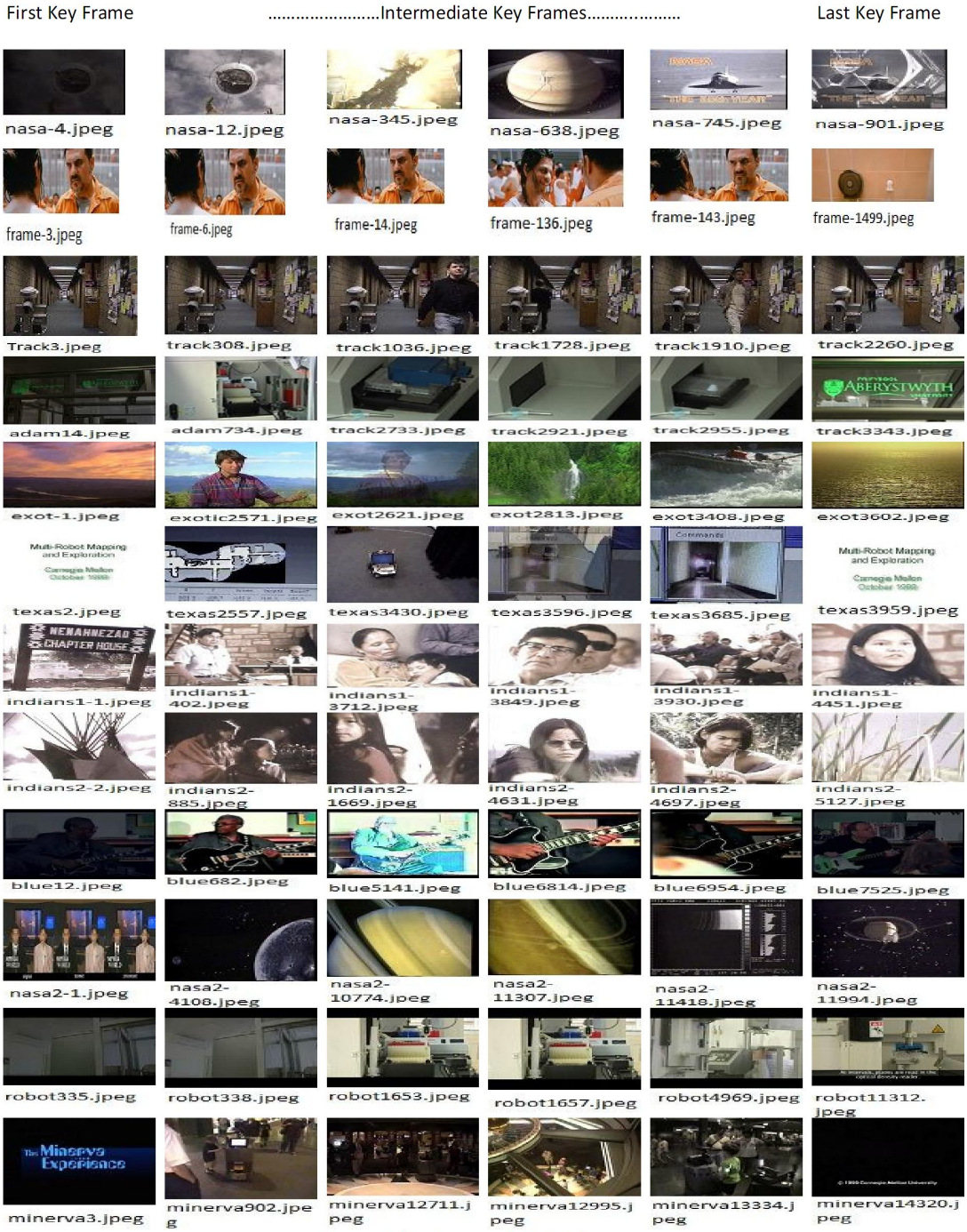

Figure-1: Key frames of 12 databases

CONCLUSION

The key frames extracted will help in processing the video files effectively and efficiently. The key frames are basically a reduction of the total number of frames. The table shows the there is no linear relation between the percentage of reduction of the frames. This means that if there is a frame or image has the sufficient information to represent itself as a key frame then it will be chosen as a key frame.

REFERENCES

[1] José M. Merigó, Montserrat Casanovas, "The Induced Minkowski Ordered Weighted Averaging Distance Operator. Spanish Congress on Fuzzy Logic and Technologies," 35-41, 2008.

[2] Frédéric Dufaux: "Key Frame Selection to Represent a Video," ICIP: 275-278, 2000.

[3] Genliang Guan, Zhiyong Wang, Shiyang Lu, Jeremiah Da Deng, David Dagan Feng: "Keypoint-Based Keyframe Selection," IEEE Trans. Circuits Syst. Video Techn. 23(4): 729-734, 2013.

[4] R. Revathi, M. Hemalatha, "Efficient method for feature extraction on video processing," CCSEIT, 539-543, 2012.

[5] Prof. A. Ardeshir Goshtasby, "Similarity and Dissimilarity Measures, Image Registration, Advances in Computer Vision and Pattern Recognition," chapter 2, pp 7-66, 2012.

[6] Liwei Wang, Yan Zhang, Jufu Feng, "On the Euclidean Distance of Images," IEEE Trans. Pattern Anal. Mach. Intell. 27(8): 1334-1339, 2005.

[7] Nishchal K. Verma, "Future image frame generation using Artificial Neural Network with selected features," AIPR: 1-8, 2012.

[8] Istvan Csapo, Brad Davis, Yundi Shi, Mar Sanchez, Martin Styner, Marc Niethammer, " Longitudinal Image Registration with Non-uniform Appearance Change," MICCAI (3): 280-288, 2012.

[9] Y.Rui, T.S.Huang, S.Mehrotra,"Constructing Table-of-Content for Videos", ACM Multimedia Systems, vol.(7), pp. 359-368, Sept. 1999.

[10] http://www.aber.ac.uk/en/cs/research/cb/projects/robotscientist/video/#adams%20videos

[11] http://robots.stanford.edu/videos.html

[12] http://www.open-video.org

[13] Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques", 2nd ed., Morgan Kaufmann Publishers, ISBN 1-55860-901-6, March 2006.

[14] Sujatha C, Uma Mudenagudi, "A Study On Keyframe Extraction Methods For Video Summary", International Conference on Computational Intelligence and Communication Systems, IEEE Computer Society, pp. 73-77, 2011.

★★★