Key-Frame Extraction Algorithm using Entropy Difference

Markos Mentzelopoulos and Alexandra Psarrou
Harrow School of Computer Science, University of Westminster
Harrow, Watford Road, HA1 3TP, UK
(m.mentzelopoulos,psarroa)@wmin.ac.uk

ABSTRACT

The fast evolution of the digital video technology has opened new areas of research. The most important aspect will be to develop algorithms to perform video cataloguing, indexing and retrieval. The basic step is to find a way for video abstraction, as this will help us more for browsing a large set of video data with sufficient content representation. In this paper we present an overview of the current key-frame extraction algorithms. We propose the Entropy-Difference, an algorithm that performs spatial frame segmentation. We present evaluation of the algorithm on several video clips. Quantitative results show that the algorithm is successful in helping annotators automatically identify video key-frames.

Categories and Subject Descriptors: B.X.X [I.4.8]: Computing Methodologies , Image Processing and Computer Vision, Scene Analysis [Object recognition]

General Terms: Algorithms and Experimentation.

Keywords: entropy,semantics.

1. INTRODUCTION

The last few years' developments in software tools have made areas such as multimedia databases quite feasible. The incredible rates at which these databases are publishing have exceeded the capabilities of current text-based cataloguing [4]. New techniques and approaches and quick search algorithms have increased the potential of the media databases, which are now containing, not only text and image but video and audio as well [13]. Extensive research efforts have been made with regard to the retrieval of video and image data based on their visual content such as colour distribution, texture and shape [1, 13]. These approaches are mainly based on feature similarity measurement.

There are a number of advantages to being able to use the visual content as the basis of a search. It is often difficult to fully express a visual query in words, and yet a single image can completely describe what is being searched for. Moving from images to video adds several orders of complexity to the retrieval problem due to indexing, analysis, and browsing over the inherently temporal aspect of video. First approaches in video retrieval just added the functionality for segmentation and key-frame extraction to existing image retrieval systems [1, 14, 17, 19]. After key-frame extraction, they just apply similarity measurement on them based on low-level features. This is not satisfactory because video is temporal media, so sequencing of individual frames creates new semantics, which may not be present in any of the individual shots. Furthermore choosing the key-frames is still a challenging problem [11].

Therefore what is needed is techniques for organizing images and videos in semantic meaning [10]. The process of extracting the semantic content is very complex, because it requires domain knowledge or user interaction, while extraction of visual features can be often done automatically and it is usually domain independent [11].

In the next section we address the problem of video segmentation into shots and techniques to select representative frames for the each shots. Then in section 3 we propose a new algorithm for key-frame extraction. Key-frames provide a suitable abstraction and framework for video browsing. However our proposed algorithm will support a video segmentation based on smaller units within the key-frame. We call these smaller units "dominant objects" because we will use them to represent key-regions (salient regions) with high probability to participate in distinct actions within the shot. This proposed model is a semantic approach to video segmentation as its hypothesis is based on the relative information that the dominant objects carry in each frame. To facilitate a detailed evaluation of the retrieval methods, we tested the algorithm against on a database of 14 video clips. Our results are given in section 4. In order to demonstrate the effectiveness of the algorithm we present a graphical comparison of our proposed approach against current key-frame extraction algorithms.Finally in section 5 we discuss our final conclusions based on the evaluation of the algorithm.

2. STATE OF THE ART IN KEY-FRAME GENERATION

Early video segmentation approaches show that there are two types of video abstraction, the video skimming and the video summary. [13, 17].

MIR'04, October 15–16, 2004, New York, New York, USA.

2.1 Video Skimming

Video skimming can be divided regarding their content into two sub-categories: The Summary Sequence and the Highlight [17].

2.1.1 Video Highlight

The Highlight contains the most interesting parts of the original video, like a movie trailer. The selected scenes usually have important people and objects that contain frames with high-contrast or high-action scenes, with frames containing the largest frame differences[1] . In order to give the impression of the movies environment, scenes with basic colour composition similar to the average colour composition of the entire movie are presented. Finally all the selected scenes are organized according to their time relevance in the movie [17].

2.1.2 Summary Sequence

However video Highlight gives only the most interesting parts of a video and doesn't layout the entire video content. The Summary Sequence Model is a quick approach for the content video retrieval by speeding up the playback. Therefore the film can be watched in a shorter amount of time with no distortion.There is a limitation on the other hand depending on speed speech, beyond which speech becomes incomprehensible. The performance of this model is based on Audio Skimming by extracting key-words with their corresponding frames. This will have as a result some times the audio not to be in alignment with the scene. This is why this model is not suitable for indexing and retrieving videos with soundtracks, which contain complex audio content. On the other hand this model has very good performance for documentaries with text content [16].

2.2 Video Summary

Video Summary compared to video Skimming is much more efficient because it focuses on the retrieval of the video content. As we said video Summary is a collection of still images (key-frames) that they best underlie the content of the shots. There are five distinguished categories of constructing key-frames: the Sampling-based, Segment-based, Motion-based and Mosaic-based and Shot-based [16].

2.2.1 Sampling-based

In this class the key-frames are selected randomly or uniformly at certain time intervals. The selection is automatic and produced by an algorithm model. The drawback is that a shot might be small in time and therefore if only one key-frame is selected to represent it there is the possibility of loosing important information content [16].

2.2.2 Segment-based

In the previous class more than one key-frame/shot is selected. But this doesn't scale for long videos. A new model must be assigned for video segmentation. According to this a video segment can be a scene, an event or even an entire sequence. Segmentation measure is computed based on its length and rarity. All segments with their importance lower than a certain threshold will be discarded. The key-frame of a segment will be the frame, which is closest to the center of the qualified segment. Finally a frame-packing algorithm will be applied for pictorial summary [16]. An alternative method of clustering introduced by the National University of Singapore [15]. The video is segmented to stories and for each story unit or scenes are extracted. For each story unit a representative (R) image is selected for each component shot. All (R) images perform a "Video Poster". The series of video posters will give us the video summary. The drawback in this model is that shot clusters with low-priority may not be assigned in the poster layout and therefore semantically properties might be missed.

2.2.3 Motion-based

Motion-based is better suited for controlling number of frames based on temporal dynamics in the scene. Pixel-based image differences or optical flow computation are commonly used. Optical flow for each frame is calculated and then a simple motion metric is computed [16].

2.2.4 Mosaic-based

Mosaic-based approach can be employed to generate a synthesized panoramic image that can represent the entire content in an intuitive manner. The procedure is working in two-steps: 1) fitting a global motion model to the motion between each pair of successive frames and 2) compose images into a single panoramic image by changing the images with the estimated camera parameters. There are currently two available research areas in mosaic [9, 16]:

- *Static background Mosaic*: For background scenes.
- *Synopsis Mosaic*: Provides a visual summary of the entire dynamic foreground event in the video clip by detecting the object trajectory.

2.2.5 Shot-based

This is the most sophisticated method for video summary. Initially the first frame of each shot was used as a key-frame, but this way doesn't' provide representation of dynamic visual content. To interpret the content we need to employ some low-level visual features, such as colour and texture or shape [1, 16]. Therefore we can have video retrieval using *independent-video* features and video retrieval using *dependent-video* features.

1. *Dependent-video features*: This class focus on two types, on Environment such as a street without foreground objects and on static or moving objects such as a car or person [14]. The model is trying to distinguish the dominant objects between consecutive frames. The colour objects (top 20) with the largest number of pixels are identified as dominant objects in a frame. In order to determine if a shot is a part of a scene the algorithm perform a correlation score calculation, and if the result satisfies the condition then the shot is a part of the scene.

2. *Independent-video features*: Video retrieval using independent video features proposes a temporal segmentation method, which integrates domain-independent video features such as colour and shape of semantic objects, object locations in order to find the location of cuts and edit effects. Such features are mainly used by

[1] Frame difference: Is the difference between two consecutive frames with respect to colour, motion and audio estimation

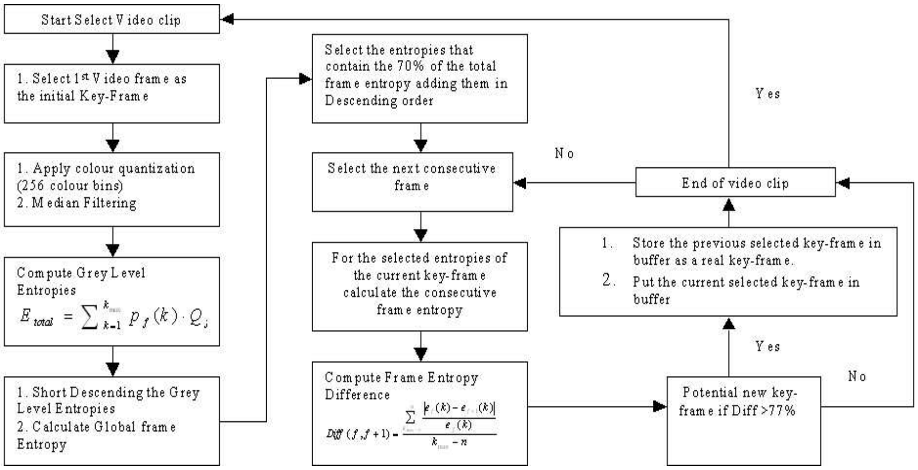
Start Select Video clip
1. Select 1st Video frame as the initial Key-Frame
1. Apply colour quantization (256 colour bins)
2. Median Filtering
Compute Grey Level Entropies
$E_{total} = \sum_{k=1}^{k_{max}} p_f(k) \cdot Q_j$
1. Short Descending the Grey Level Entropies
2. Calculate Global frame Entropy
Select the entropies that contain the 70% of the total frame entropy adding them in Descending order
Select the next consecutive frame
For the selected entropies of the current key-frame calculate the consecutive frame entropy
Compute Frame Entropy Difference
$Diff(f, f+1) = \frac{\sum \frac{|e_f(k) - e_{f+1}(k)|}{e_f(k)}}{k_{max} - n}$
Potential new key-frame if Diff >77%
Yes
No
1. Store the previous selected key-frame in buffer as a real key-frame.
2. Put the current selected key-frame in buffer
End of video clip
Yes
No

Figure 1: Flowchart of the proposed Entropy Difference Algorithm

commercial videos and are easy for retrieval. For example the logo from a TV program, the broadcasting music etc [1, 3].

2.3 Automatic Video Segmentation Into Shots

These techniques are generally based on a combination of image analysis techniques and a rule set, aiming of modeling conditions under which a certain situation occurs [1]. Most of the activity is concentrated on the detection of shot boundaries using sharp transition detection (cuts). The cut is defined as sharp transition between a shot and the one following. It is obtained by simply joining two different shots without the insertion of any photographic effect (e.g.: fades and dissolves). Cuts generally correspond to an abrupt change in the brightness pattern for two consecutive images. The principle behind this approach is that, since two consecutive frames in a shot do not change significantly in their background and object content, their overall brightness distribution differs little. On the other hand if we have a scene where there is a dramatically change in the illumination of the background it will have consequences to the brightness level of the image.

Based on sharp transition detection (cuts) a number of algorithms were implemented to extract key-frames[1]. Zhang and Smoliar [1, 6, 18] have proposed three metrics for sharp transition detection based on pairwise pixel comparison, likelihood ratio and histogram comparison. Nagasaka and Tanaka [1, 2] have proposed an algorithm for cut detection based on the normalized test χ^2, which compares the distance between colour histograms bins of two consecutive frames. Hampapur and Weymouth [1, 5, 13] have developed a cut detection by using a difference operator applied to two successive colour frames.

3. KEY-FRAME EXTRACTION USING THE ENTROPY DIFFERENCE

Describing an object to a retrieval system typically involves the use of characteristics such as texture and colour. In this algorithm we use the entropy not as global feature for the total image but as local operator. Entropy is a good way of representing the impurity or unpredictability of a set of data since it is dependent on the context in which the measurement is taken. Kadir and Brady used entropy [7]to describe parts of an image in terms of varying scales in space. In the propose model we consider that if we distribute the entropy among the image, then the higher entropy distributions will be describing the regions containing the salient objects of a video sequence. Therefore any change in the object appearance of one of this salient regions it will affect its relevant semantic information in the entire story sequence.

3.1 Entropy Algorithm

In order to eliminate as much as we can the possibility of the change of brightness during the key-frame comparison, it has been deemed necessary to quantize the image down to 256 colours and then apply a median filter for region smoothing. Let $h_f(k)$ be the histogram of frame f and k the gray level $0 \leq k \leq 2^{b-1}$ where b is the smallest number of bits, in which the image quantization levels can be represented [b=256, its frame has been quantized to 256 colours]. If the video frame is of the class M rows N columns, then the probability of appearance of this gray level in the frame will be:

$$p_f(k) = h_f(k)/(M \cdot N) \tag{1}$$

The information quantity $Q_f(k)$ transmitted by an element is equal to the log (base2) of the inverse probability of appearance $p_f(k)$

$$Q_f(k) = \log_2(\frac{1}{p_f(k)}) = -\log_2(p_f(k)) \tag{2}$$

The above information $Q_f(k)$ multiplied by its probability of appearance gives us the entropy E generated by the source for this quantization level. The sum of all the gray level entropies is the global entropy information of the frame.

$$E_{total} = \sum_{k=1}^{k_{max}} e_f(k) \tag{3}$$

Where:

$$e_f(k) = p_f(k) \cdot Q_f(k) \tag{4}$$

For each frame we first sort the entropies between all quantization levels and then we add them, starting from the highest towards the lowest entropy until we exceed the threshold

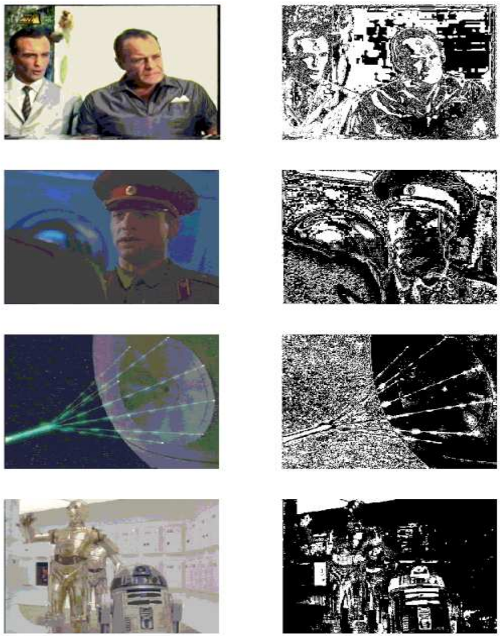

Figure 2: : On the left hand side are the original key-frames. On the right hand side are the binary image representations of original key-frames after applying the entropy algorithm. With white seeds are the regions that containing the 70% of the image information. It can be seen clearly that the algorithm has identified in all the frames the dominant objects. The frames are a sample from 4 different movies

of the 70% of the total image entropy. With this way we isolate the objects that carrying the most information in the image:

$$E_{Threshold} = \sum_{m=z}^{n} e_m \geq 0.7 \cdot E_{total} \tag{5}$$

Where n is the entropy at which the sum exceeds the 70% of the total entropy and z = Maximum entropy. For each of the gray level entropies that we used in order to reach the 70 % of the total image entropy for the first image of the sequence [key-frame = frame 1], we take the absolute difference with the relevant gray-level entropy from the next processed frame. If the sum of the normalized differences is more than 77% then we have a change in the content of the frame-sequence and therefore a new key-frame is needed (Equation 6).

$$Diff(f, f+1) = \frac{\sum_{k_{max}-1}^{n} \frac{|e_f(k) - e_{f+1}(k)|}{e_f(k)}}{k_{max} - n} \tag{6}$$

where f and f+*1* are the current frame and the consecutive to it frame.

3.2 Shot unit merging

As a post processing of shot segmentation, the shot units of which duration is less than some threshold are evaluated whether it can be merged with the neighbor shot unit [8]. Video segmentation algorithms which simply looked at changes between consecutive frames [1, 13] were unreliable because they were defeated by gradual changes which take place across a number of frames, such as fades, or by transients, such as flashes, which cause a sudden change on a single frame where no shot change is actually taking place [12]. To avoid this we evaluate the frame distance between the current key-frame and the previous key-frame. If the distance is larger than some threshold then we have a new key-frame. Figure 3 shows an instance of the post processing scheme. The frame duration of the shot unit B is less than a specified frame threshold and therefore, the shot unit B can be merged with the previous shot (shot A) .Figure 4 shows a demonstration of the proposed technique. A frame sequence has been extracted containing 4 possible shots (figure 4(a)). The merge threshold has been set to 5 frames. In figure 4(b) are the key-frames that have been extracted after applying the entropy-difference algorithm to the above segment, while in figure 4(c) are the remaining key-frames after applying and the merge pack algorithm.

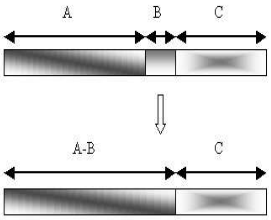

Figure 3: Post processing of Shot unit. The second shot (B) that has been identified contains less frames than a specified threshold and therefore it can be merge with the first shot (A) after which only one key-frame will be needed to semantically characterize the shot

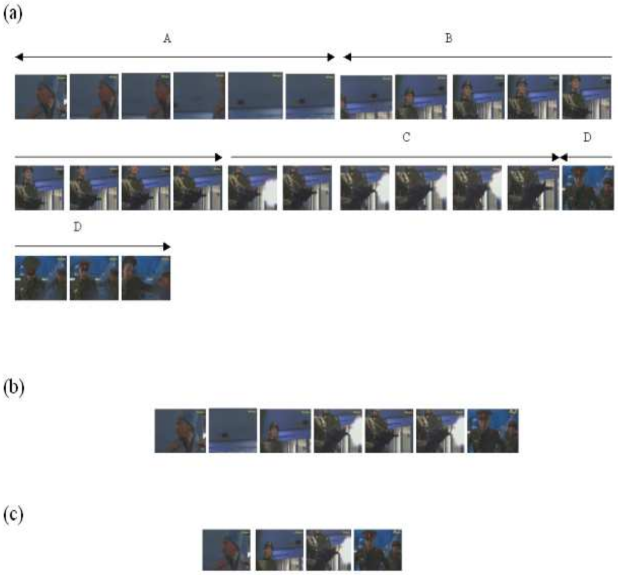

Figure 4: (a) Original frame sequence, (b) key-frames extracted after applying the entropy-difference algorithm, (c) key-frames remained after applying the post processing of shot unit

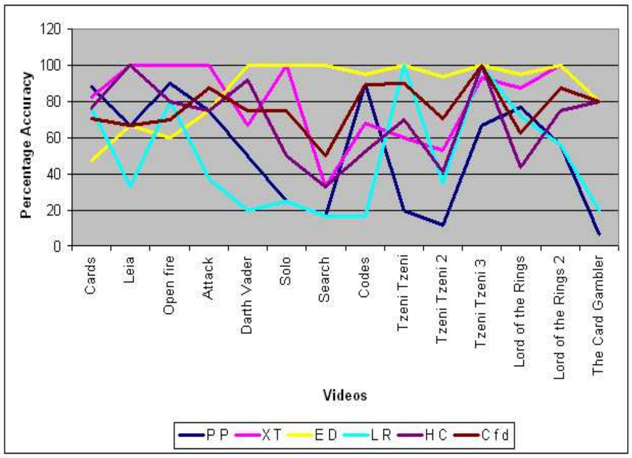

Figure 5: Graphical representation of the results from the key-frame extraction algorithms comparison regarding 14 different video clips. On the X-Axis are the video clips while in the Y-Axis is the percentage relevance of the key-frames extracted compared to the theoretical key-frames. The compared algorithms are:Pairwise Pixel(P P), x^2 Test(X T), Entropy Difference(E D), Likelihood Ratio(L R), Histogram Comparison(H C) and the Consecutive Frame Difference(C f d)

4. RESULTS

The algorithm was implemented in the Matlab workspace and compared against the 5 different key-frame extraction techniques introduced in section 2.6. In all the experiments reported in this section, the video streams are AVI format, with the digitization rate equal to 24 frames/sec. To validate the effectiveness of the proposed algorithm, representatives of various movie types are tested. The video clips that we have selected can be sorted into to three different categories: 1. High Action clips (Clips selected from Star War, Lord of the Rings and James Bond theme), 2. Conversation clips between faces (Greek Movies and Star Wars theme) and 3. Simple motion clips (Clips selected from Star War, Lord of the Rings, James Bond theme and Greek Movies). The video clip length varies from 30 sec to 4 minutes long.

Initially all the video clips have been split into simple frames in order to identify possible key-frames. These have been judged by a human who watches the entire video. Although key-frames are a semantic concept, relative agreement can be reached among different people. We used these key-frames as our theoretical database against which the six different algorithms were compared and percentage accuracy was calculated regarding to how many correct key-frames each algorithm has managed to identify. Table 1 shows a comparison of the key-frames extracted out of the sequences using the techniques from section 2.6 and the entropy-difference algorithm. It also allows to see comparison between how efficiently the key-frames are extracted in terms of redundant frames and missing frames. Table 2 contains information of the video sequences ((1)key-frames/video sequence ,(2) Total number of frames in video clip). The experimental results are shown in figure 5. We can see that the entropy-difference approach returns particularly impressive results in clips that the background is quite easily distinguishable from the dominant objects, such as conversation clips between faces (Leia) and simple motion clips (Search). On the other hand when there is transient change the performance of the algorithm is low.

Table 2: Video sequences information regarding:(a) The number of video breaks which is the ideal number of key-frames, (b)The total number of frames in the video sequence

Video Clips	K	T
Darth Vader	12	942
Solo	4	172
Leia	3	194
Search	6	157
Cards	17	1432
Codes	19	890
Open fire	10	216
Attack	8	420
Tzeni Tzeni	10	1127
Tzeni Tzeni 2	17	1450
Tzeni Tzeni 3	15	2300
Lord of the Rings	88	3992
Lord of the Rings 2	32	1439
The Card Gambler	15	4913

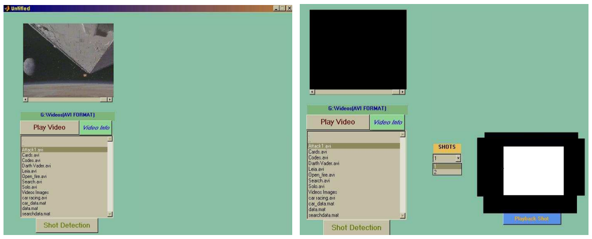

Figure 7: Graphical User Interface for key-frame extraction designed in the Matlab workspace

Figure 6, shows the key-frames extracted from two sample video sequences using the methods of (a) Consecutive frame difference [1, 6], (b) x^2 Test [1, 2], and (c) Entropy difference. Based on the shot detection created by the algorithm, we have designed a simple graphical user interface (figure 7). Clicking on the pop-up menu the user can select any available video from the current directory and apply the algorithm. At the end by clicking on the second menu there is a shot playback possibility.

5. CONCLUSION

This paper has introduced current techniques used for video segmentation. Beside standard queries algorithms used for automatic key-frame extraction we propose a new method for key-frame identification regarding to the entropy that the dominant objects contain. The algorithm performs very well when the image background is distinguishable from the objects. On the other hand when there is transient change such as flashes (e.g. explosions-shooting) the performance drops to the average algorithms performances. The main advantage of our model is that it segments with high accuracy the videos into key-frames using a semantic meaning as

Table 1: A comparison of the key-frames extracted from 14 sample video sequences using the Pairwise Pixel(P P), x^2 Test(X T), Entropy Difference(E D), Likelihood Ratio(L R), Histogram Comparison(H C) and the Consecutive Frame Difference(C f d) in terms of: (a) The number of key-frames in total(T), (b) the number of redundant frames(R),(c) the number of missing frames(M)

Video Clips	P P			X T			E D			L R			H C			C f d		
	T	R	M	T	R	M	T	R	M	T	R	M	T	R	M	T	R	M
Darth Vader	9	3	6	10	2	4	16	5	1	10	8	10	25	14	1	32	23	3
Solo	1	0	3	4	0	0	5	1	0	1	0	3	2	0	2	5	2	1
Leia	32	30	1	3	0	0	4	2	1	1	0	2	3	0	0	12	10	1
Search	1	0	5	2	0	4	14	8	0	1	0	6	2	0	4	3	0	3
Cards	22	7	2	126	112	3	88	80	9	141	137	13	50	37	4	47	35	5
Codes	38	21	2	73	60	6	61	43	1	7	6	18	93	83	9	36	19	2
Open fire	32	23	1	10	0	0	6	0	4	25	23	8	27	19	2	13	6	3
Attack	40	34	2	33	25	0	19	13	2	9	6	5	25	19	2	17	10	1
Tzeni Tzeni	2	0	8	6	0	4	22	12	0	842	832	0	11	4	3	33	24	1
Tzeni Tzeni 2	2	0	15	9	0	8	34	18	1	53	47	11	17	10	10	33	21	5
Tzeni Tzeni 3	10	0	5	51	37	1	67	52	0	816	811	0	32	17	0	67	52	0
Lord of the Rings	73	13	20	134	57	11	112	28	4	930	867	25	161	128	50	101	68	50
Lord of the Rings 2	23	5	14	229	197	0	225	193	0	180	162	14	218	104	8	38	10	4
The Card Gambler	1	0	14	12	0	3	17	5	3	21	18	12	29	17	3	75	63	3

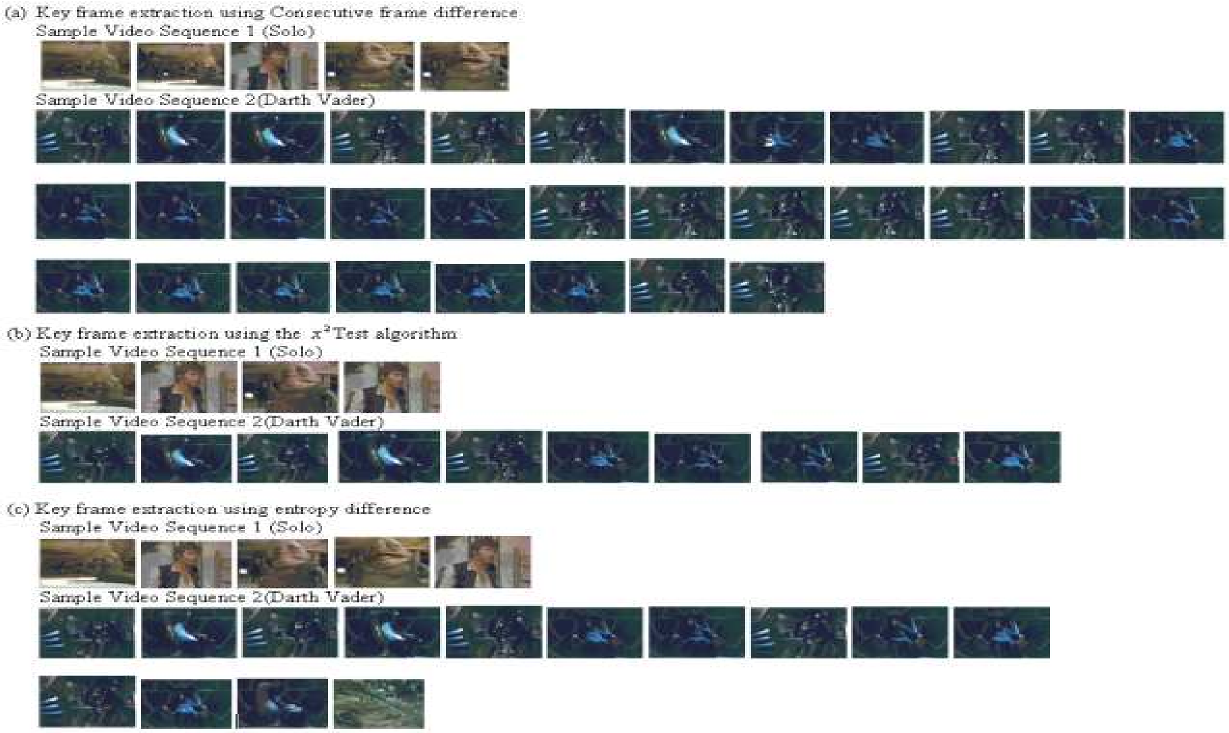
(a) Key frame extraction using Consecutive frame difference
Sample Video Sequence 1 (Solo)
Sample Video Sequence 2(Darth Vader)
(b) Key frame extraction using the x^2Test algorithm
Sample Video Sequence 1 (Solo)
Sample Video Sequence 2(Darth Vader)
(c) Key frame extraction using entropy difference
Sample Video Sequence 1 (Solo)
Sample Video Sequence 2(Darth Vader)

Figure 6: Key-frames extracted from two sample video sequences using the methods (a) Consecutive frame difference,(b) x^2 Test, and (c) Entropy difference. From the results of Table 1 it can be seen that in the first video sequence we have identified 3/4 correct key-frames using the Consecutive frame difference (2-Redundant key-frames), 4/4 correct key-frames using the x^2 Test from Nagasaka and Tanaka and 4/4 key-frames using the Entropy difference (Including 1 Redundant key-frame). In the second video clip there are 32 key-frames extracted using the Consecutive frame difference, with only 9 of them being the accurate ones (Theoretical key-frames=12) and 23 Redundant. Using the x^2 Test, 8 correct key-frames have been exracted with only 2 redundant frames including. Finally using the Entropy difference 14 key-frames have been identified with only 3 of them to be redundant.

it acknowledge that the basic information (entropy) will be concentrated in the dominant objects of the current frame. In addition, we have developed an interactive user interface to facilitate the integration of automated and human annotation, and it results in a hybrid system where computer and users work cooperatively to achieve the best retrieval performance.

6. REFERENCES

[1] A.D.Bimbo. *Visual Information retrieval.* Morgan Kaufmann Publishing, San Francisco,, 1999.

[2] A.Nagasaka and Y.Tanaka. Automatic video indexing and full-motion video search for object appearences. *Visual Database Systems II*, pages 113–127, 1992.

[3] B.Gunsel, A.M.Ferman, and A.M.Tekalp. Temporal video segmentation using unsupervised clustering and semantic object tracking. *Journal of Electronic Imaging*, 7:592–604, July 1998.

[4] D.D.Petkovic. Challenges and opportunities in search and retrieval for media databases. *IEEE Workshop on Content - Based Access of Image and Video Libraries*, pages 110–111, 1998.

[5] A. Hampapur, R. Jain, and T. Weymouth. Digital video segmentation. *ACM Multimedia*, pages 357–364, 1994.

[6] H.Zhang, J.Wu, D.Zhong, and S.W.Smoliar. An interated system for content-based video retrieval and browsing. *Pattern Recognition*, 30:643–658, 1997.

[7] T. Kadir and M. Brady. Scale, saliency and image description. *IJCV*, 45(2):83–105, November 2001.

[8] Y.-M. Kwon, C.-J. Song, and I.-J. Kim. A new approach for high level video structuring. *IEEE International Conference on Multimedia and Expo (II)*, pages 773–776, 2000.

[9] M.Iran and P.Anandan. Video indexing based on mosaic representation. *IEEE*, 5:86, May 1998.

[10] N.Sebe, M.S.Lew, X.Zhou, T.Huang, and E.M.Bakker. The state of the art in image and video retrieval. *International Conference on Image and Video Retrieval (CIVR'03)*, pages 1–8, July 2003.

[11] M. Petkovic. Content-based video retrieval. *Centre for Telematics and Information Tecnology, Univrsity of Twente*, 2001.

[12] M. J. Pickering, S. M. Ruger, and D. Sinclair. Video retrieval by feature learning in key frames. *CIVR 2002*, pages 309–317, 2002.

[13] M. S.Lew. *Principles of Visual Information Retrieval.* Springer-Verlag, London,UK, 2001.

[14] T.Lin and H.J.Zhang. Automatic video scene extraction by shot grouping. *ICPR'2000-15th International Conference on Pattern Recognition, Barcelona, Spain*, September 2000.

[15] X.Sun and M.Kankanhalli. Video summarization using r-sequences. *Real-time Imaging*, pages 449–459, December 2000.

[16] Y.Li, T.Zhang, and D.Tretter. An overview of video abstraction techniques. *HP*, July 31st 2001.

[17] Y.Rui, S.Thomas, H.Mehrota, and S.Mehrota. Exploring video structure beyond the shots. *IEEE International Conference on Multimedia Computing and Systems*, pages 237–240, June-July 1998.

[18] H. Zhang, A. Kankanhalli, and S. Smoliar. Automatic partitioning of full-motion video. *Multimedia Systems*, 3(1):10–28, November 1993.

[19] Z.Li, Q.Wei, Z.Stan, Li, S.Yang, Q.Yang, and H.J.Zhang. Key-frame extraction and shot retrieval using nearest feature line (nfl). *International Workshop on Multimedia Information Retrieval, in conjunction with ACM Multimedia Conference 2000*, November 2000.