INTERNATIONAL JOURNAL OF MERGING TECHNOLOGY AND ADVANCED RESEARCH IN COMPUTING

ISSN: 2320-1363

VIDEO RETRIEVAL FROM CORRUPTED SOURCE BASED ON K++MEANS CLASSIFICATION

R.Shameli [*1], G. Bharathi [*2]

M.E Scholar, Communication Systems, Dept. of ECE, Ranippettai Engineering College, Walajah, India

Asst. Professor, Dept. of ECE, Ranippettai Engineering College, Walajah, India

ABSTRACT

This project presents a recovery technique of a corrupted video file using the specifications of a codec used to encode the video data. In case of storage, the video files are stored as per the optimum, software architecture plans. Hence it is proposed to use, frame based approach to reconstruct a corrupted video. The proposed approach uses an intelligent method of using K++ clustering algorithm, to segment the reconstructed frame. In case of bad reconstruction, probably it may give more number of clusters. Hence number of clusters obtained is kept as threshold. Moreover, addresses how to extract video frames from a portion of video to be restored is done through a linear regression based analysis. Also the scheme may be used as how to connect extracted video frames together according to the codec specifications. In forensics, recovery of a damaged or altered video file plays a crucial role in searching for evidences to resolve a criminal case. Videos are played based on the fast running of individual images at the rate of minimum 17 frames per second. Experiment results are to be verified with MATLAB 7.12 with video and image processing tool box.

INTRODUCTION

Recently, a large amount of video contents have been produced in line with wide spread of surveillance cameras and mobile devices with built-in cameras, digital video recorders, and automobile black boxes. Recovery of corrupted or damaged video files has played a crucial role in role in digital forensics [1]–[3]. In criminal investigations, video data recorded on storage media often provide an important evidence of a case. As an effort to search for video data recorded about criminal, video data restoration and video file carving has been actively studied [4]–[6].

Most existing video data restoration techniques attempt to restore the source data using meta-information recorded in the header of a file system [7], [8]. The meta-information of file system contains file information such as file name, time of modification, physical location, link, etc. When the operator deletes a file, the corresponding file information in the meta information of file system is updated as deleted although the video contents

ISSN: 2320-1363

physically remain in the medium. Even though a video content exists in the media, it is challenging to recover the video data if the relevant meta-information is removed or altered. Conventional file restoration techniques find the meta information of the deleted files to search for physical locations containing the actual file contents. However, the file cannot be restored if not all the file links are connected. Since a video file typically has a large volume of the data, it is highly likely to be fragmented although the meta-information remains in the file header [9], [10], [11]. When part of the file was overwritten, restoration of a video file with meta-information only may not be successful in most situations [12]. To tackle these problems, various techniques have been proposed by which if the file start markers and end markers are discovered based on the file signature, relevant data are collected to restore the video data [12]–[15].

Signature-based file restoration techniques search for the start marker (header) and the end marker (footer) to find a valid connection of the regions containing the header and the footer [16]. To increase the accuracy of the connection of the header and the footer regions, they used other information such as maximum size, embedded length recorded in the header. The analysis of the signature may offer a low success rate in video file restoration, when there are many file fragments and when some of them are overwritten. Especially, in the case a portion of a video file is overwritten, restoration of the video data using the file unit can be almost impossible because validation of restored file is failed by partially overwritten of restored file [17].

In this view point, we extend technique from conventional signature-based file restoration technique.1 This paper proposes a technique to restore the video data on a frame-by frame basis from its corrupted versions where the video data has been significantly fragmented or partly overwritten in the storage media. A video data consists of a sequence of video frames as the minimum meaningful unit of video file. The proposed method identifies, collects, and connects isolated video frames using the video codec specifications from non over written portions of the video data to restore a corrupted video file.

The proposed technique restores the video data in a frame unit, not in a file unit. This is a simple, yet powerful video data restoration method that can recover a portion of the file even when a complete restoration of the file is not possible. The proposed frame-based video data restoration scheme can restore the video regardless of a file system. This approach can restore a video data from fragmented data stored on a corrupted or damaged video file. Since large size multimedia file tend to have a large amount of fragments [13], a file based restoration technique may not be successful. File-based restoration of conventional methods is extremely difficult if the physical locations of all fragmented data are unknown or a part of file is overwritten.

The proposed method restores a corrupted or damaged video file using each video frame, the minimum unit of video file, using the index data on the disk area. In the region to restore, we extract the part of the data that can possibly be frame to do decoding. Then we collect the frames that can be connected after decoding to restore the video data. When a large amount of fragments exist and

ISSN: 2320-1363

even when a part of file is overwritten, we can collect and connect remaining video frame to restore a video data. The technique consists of extraction phase and connection phase of relevant video frames. The extraction phase uses the video codec specifications to extract a set of video frames from the storage media. In the connection phase, the restored video frames are used to group and connect relevant video frames using the specifications of the video file used.

K++ means clustering

Over half a century old and showing no signs of aging, k-means remains one of the most popular data processing algorithms. As is well-known, a proper initialization of k-means is crucial for obtaining a good solution.

The recently proposed k-means++ initialization algorithm achieves this, obtaining an initial set of centers that is provably close to the optimum solution. A major downside of the k-means++ is its inherent sequential nature, which limits its applicability to massive data: one must make k passes over the data and a good initial set of centers. In this work we show how to drastically reduce the number of passes needed to obtain, in parallel, a good initialization. We prove that our proposed initialization algorithm k-means|| obtains a nearly optimal solution after a logarithmic number of passes, and then show that in practice a constant number of passes success. Clustering is a central problem in data management and has a rich and illustrious history with literally hundreds of different

algorithms published on the subject. Even so, a single method | k-means | remains the most popular clustering method; in fact, it was identified as one of the top 10 algorithms in data mining [34]. The advantage of k-means is its simplicity: starting with a set of randomly chosen initial centers, one repeatedly assigns each input point to its nearest center, and then re-computes the centers given the point assignment. This local search, called Lloyd's iteration, continues until the solution does not change between two consecutive rounds. The k-means algorithm has maintained its popularity even as datasets have grown in size. Scaling k-means to massive data is relatively easy due to its simple iterative nature. Given a set of cluster centers, each point can independently decide which center is closest to it and, given an assignment of points to clusters, computing the optimum center can be done by simply averaging the points. Indeed parallel implementations of k-means are readily available (see, for example, cwiki.apache.org/MAHOUT/k-meansclustering.html). From a theoretical standpoint, k-means is not a good clustering algorithm in terms of efficiency or quality: the running time can be exponential in the worst case [32, 4] and even though the final solution is locally optimal, it can be very far away from the global optimum (even under repeated random initializations). Nevertheless, in practice the speed and simplicity of k-means cannot be beat. Therefore, recent work has focused on improving the initialization procedure: deciding on a better way to initialize the clustering dramatically changes the performance of the Lloyd's iteration, both in terms of quality and convergence properties.

A important step in this direction was taken by Ostrovsky et al. [30] and Arthur and Vassilvitskii [5], who showed a simple procedure that both leads to good theoretical guarantees for the quality of the solution,

ISSN: 2320-1363

and, by virtue of a good starting point, improves upon the running time of Lloyd's iteration in practice. Dubbed k-means++, the algorithm selects only the firrst center uniformly at random from the data.

Each subsequent center is selected with a probability proportional to its contribution to the overall error given the previous selections (we make this statement precise in Section 3). Intuitively, the initialization algorithm exploits the fact that a good clustering is relatively spread out, thus when selecting a new cluster center, preference should be given to those further away from the previously selected centers.

Algorithm

Algorithm 1 k-**means++**(k) initialization.

```
C ← sample a point uniformly at random from X
while |C| < k do
    Sample x ∈ X with probability d²(x,C) / φ_X(C)
    C ← C ∪ {x}
end while
```

SUMMARY OF PREVIOUS WORKS

Recovery of damaged or corrupted video files obtained from a crime scene or a disaster site has provided a key evidence to resolve the cause. Conventional techniques for video file restoration use the meta-information of the file system to recover a video file stored in a storage medium such as a hard drive or a memory card [8]. The file system meta-information contains the information such as the address and the link of a video file that can be used for file restoration. Carrier [8] proposes a file restoration tool based on the file system, which was implemented in a software toolkit, The Sleuth Kit [20]. This program is based on the information from the file and directory structure of a storage filesystem. Video file restoration may not be possible with such techniques, however, when the file system meta-information is not available. Thus, attempts have been made to restore the video data from video contents, rather than the meta-information of a file system.

The earlier works presents a technique to restore damaged or corrupted video files irrespective of a file system. The signature-based video restoration technique proposes File Carver [16] to address this problem. This method creates a database of the file header (beginning mark of file) and footer (the end mark of file), and define a set of rules for a specific file type. Signature-based file recovery techniques

do not require file system information, which can be applied to a video file with no meta-information because of file system change and reformatting of a storage medium. Signature-based file recovery techniques identify the fragments from the bytesequence (or magic bytes) containing file header or footer. Scalpel [16] does not rely on a file system to restore a video file. This technique requires an indexing step to find the file header and footer from a whole disc as well as a restoration step to recover indexed header and footer. We do note use file system metadata to restore the data between the header and footer to a file. This method is limited to the cases when the files are unfragmented. This method does not recover partially overwritten video files. Garfinkel [13] utilizes additional information stored in the file to extend the idea to signature-based restoration techniques. For some files, file header may contain the information of file size or length. When the file footer does not exist, they can use this information to extract a file. A video

ISSN: 2320-1363

file can be restored using Bifragment Gap Carving [13].

. In general, the signature-based file carving techniques mentioned above consist of the following three steps [2]. 1) Identification Phase: To identify a video fragment in a storage medium and to connect it to the previous fragment.

2) Validation Phase: To validate if all connected video fragments successfully form a playable video file.

3) Validate by Human Expert: To sort out false positive video segments by human expert.

The validation step checks if a restored video file is a playable video file. Conventional file-based video restoration techniques may fail to validate a restored video when a part of video is overwritten [17]. On the other hand, the proposed frame-based method carry out video restoration frame by frame, and is therefore applicable to restoration of partially overwritten video file.

PROPOSED SYSTEM

The major aim is to convert the data in sequence into a readable video format. The process includes first, the construction of a frame. Here a method regression is followed to linearise the image content as per the content available. To perform this, a median value is filled in the place of missing bytes. Once the frame is extracted, the total frames available are now subject to adaline algorithm to verify the truthness of sequential content. The cluter results of each frame are compared with their cross correlation coefficients. The matching cluster, or slightly varying frames are arranged in sequential manner and finally the image frames are written as a video format file. Initially an image frame is first constructed using linear filtering(after a matrix is constructed). The linear filtering based on regression is shown below.

Video Settings:

BitsPerPixel = 24
FrameRate = 25
Height = 480
NumberOfFrames = 84
VideoFormat = RGB24
Width = 640

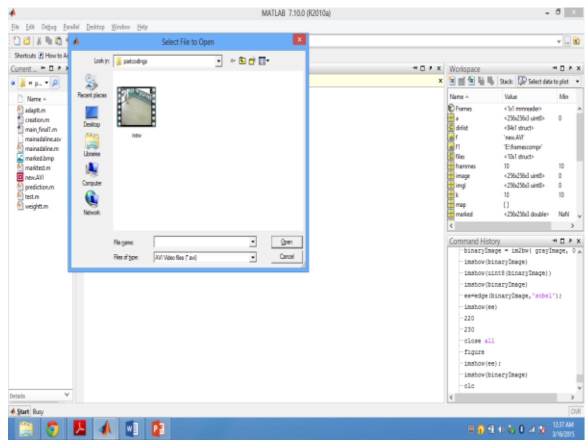

Fig.2 Snapshot of video selection

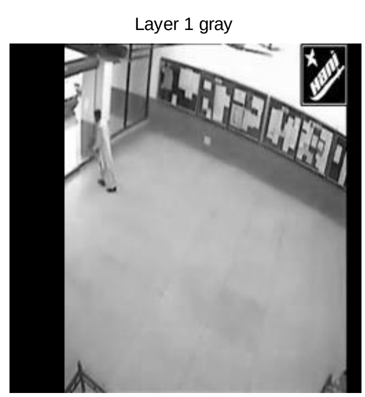

Fig.3a.Gray converted Red layer

ISSN: 2320-1363

Fig.3b.Gray converted Green layer

Fig.3c.Gray converted Blue

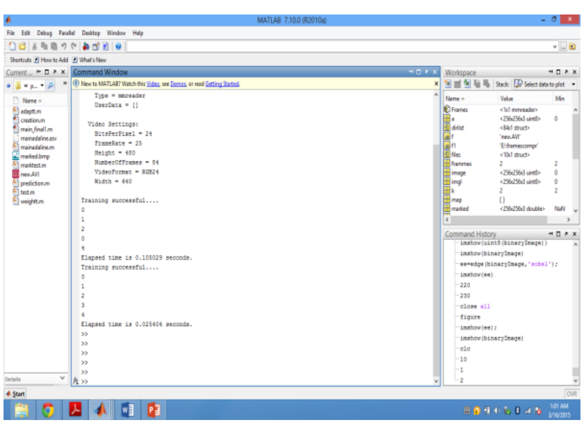

Fig.4 Command window output

MERITS IN PROPOSED WORK

- INNOVATIVE ALGORITHM IS USED
- RETRIEVAL POSSIBILITY IS INCREASED
- ADALINE ALGORITHM ACTS AS VIDEO ENHANCEMENT TOOL.
- REGRESSION METHODS ACTS AS IMAGE ENHANCEMENT TOOL

CONCLUSION

This project presents a video restoration technique for fragmented and partially overwritten video files. The proposed technique guarantees the integrity of the restored frames because video files have the minimum number of frames to offer evidence. Large-size video files are often fragmented and overwritten. Many existing file-based techniques could not restore partially overwritten video files. Unlike most existing methods that use file format or file system meta-information, the proposed technique restores the data according to the minimum meaningful frame unit. Therefore, the proposed method restores almost frames in damaged or corrupted video files without being affected by the number of fragmentations.

REFERENCES

[1] K. Medaris and R. Mislan. (2008). Expert: Digital Evidence Just as Important as DNA in Solving Crimes [Online] Available:

ISSN: 2320-1363

http://news.uns.purdue.edu/x/2008a/080425T-MislanPhones.html
[2] R. Poisel and S. Tjoa, "Forensics investigations of multimedia data: A review of the state-of-the-art," in Proc. 6th Int. Conf. IT Security Incident Manag. IT Forensics, May 2011, pp. 48–61.
[3] H. T. Sencar and N. Memon, "Overview of state-of-the-art in digital image forensics," Algorithms, Archit. Inf. Syst. Security, vol. 3, pp. 325–348, Nov. 2008.
[4] L. Huston, R. Sukthankar, J. Campbell, and P. Pillai, "Forensic video reconstruction," in Proc. ACM 2nd Int. Workshop Video Surveill. Sensor Netw., 2004, pp. 20–28.
[5] A. B. Lewis, "Reconstructing compressed photo and video data," Comput. Lab., Univ. Cambridge, Cambridge, U.K., Tech. Rep. 813, 2012.
[6] R. Poisel and S. Tjoa, "Roadmap to approaches for carving of fragmented multimedia files," in Proc. 6th Int. Conf. ARES, Aug. 2011, pp. 752–757.
[7] L. Aronson and J. Van Den Bos, "Towards an engineering approach to file carver construction," in Proc. IEEE 35th Annu. OMPSACW, Jul. 2011, pp. 368–373.
[8] B. Carrier, File System Forensic Analysis, vol. 3. Boston, MA, USA: Addison-Wesley, 2005.
[9] D. Billard and R. Hauri, "Making sense of unstructured flash-memory dumps," in Proc. ACM Symp. Appl. Comput., 2010, pp. 1579–1583.
[9] D. Billard and R. Hauri, "Making sense of unstructured flash-memory dumps," in Proc. ACM Symp. Appl. Comput., 2010, pp. 1579–1583.
[10] G. Richard, III, V. Roussev, and L. Marziale, "In-place file carving," in Advances in Digital Forensics III. New York, NY, USA: Springer-Verlag, Jan. 2007, pp. 217–230.
[11] V. L. L. Thing, T.-W. Chua, and M.-L. Cheong, "Design of a digital forensics evidence reconstruction system for complex and obscure fragmented file carving," in Proc. 7th Int. Conf. CIS, Dec. 2011, pp. 793–797.
[12] A. Pal and N. Memon, "The evolution of file carving," IEEE Signal Process. Mag., vol. 26, no. 2, pp. 59–71, Mar. 2009.
[13] S. L. Garfinkel, "Carving contiguous and fragmented files with fast object validation," Digit. Invest., vol. 4, pp. 2–12, Sep. 2007.
[14] N. Memon and A. Pal, "Automated reassembly of file fragmented images using greedy algorithms," IEEE Trans. Image Process., vol. 15, no. 2, pp. 385–393, Feb. 2006.
[15] R. Poisel, S. Tjoa, and P. Tavolato, "Advanced file carving approaches for multimedia files," J. Wireless Mobile Netw., Ubiquitous Comput., Dependable Appl., vol. 2, no. 4, pp. 42–58, 2011.
[16] G. G. Richard and V. Roussev, "Scalpel: A frugal, high performance file carver," in Proc. DFRWS, 2005, pp. 1–10.
[17] T. Laurenson, "Performance analysis of file carving tools," in Security and Privacy Protection in Information Processing Systems. New York, NY, USA: Springer-Verlag, 2013, pp. 419–433.
[18] Infornation Technology—Coding of Audio-Visual Objects—Part 2: Visual, ISO/IEC Standard 14496-2:2004, 2004.
[19] Information Technology—Coding of Audio-Visual Objects—Part 10: Advanced Video Coding, ISO/IEC Standard 14496-10:2009, 2009.