

Enhanced Recovery of AVI Files Using Recuva

¹Shivaranjini, ²Mahesh Babu K, ³Maheswar Reddy K

^{1,2,3}Computer Science & Engineering (Cyber Security), Madanapalle Institute of Technology & Science, Madanapalle, India

E-mail: ranju05sb@gmail.com, maheshkollubabu@gmail.com, kingsmaheshwarreddy@gmail.com

Abstract - This research focuses on advancing the data recovery process for fragmented AVI files on NTFS file systems using enhancements to Recuva, a popular recovery tool. Audio Video Interleave (AVI) files often suffer from fragmentation due to their large file size and frequent editing, making recovery challenging. The study emphasizes improvements in three core areas: identifying fragmented file segments, reconstructing incomplete or corrupted metadata, and recovering data in scenarios where portions of the file have been overwritten. By integrating these improvements, new algorithms were developed to analyze file patterns, reconstruct lost metadata, and recover partially overwritten fragments. Rigorous testing demonstrated a significant boost in recovery success rates under diverse fragmentation and overwriting scenarios. These advancements are expected to enhance data recovery solutions for both individual users and organizations handling critical multimedia files.

Keywords: Data recovery, Recuva, AVI files, NTFS, fragmentation, metadata reconstruction, overwritten data, file carving.

I. INTRODUCTION

In today's digital era, multimedia content has become an integral part of human interaction. AVI (Audio Video Interleave) files, a widely used multimedia format, are commonly employed for storing video and audio data. However, over time, as files are modified, fragmented, and stored across disk clusters, AVI files may become fragmented, posing significant challenges for data recovery. Fragmentation occurs when a file is broken into smaller pieces and scattered across different parts of the storage medium. This can be caused by various factors such as frequent editing, resizing, or a large number of deleted files that leave gaps in storage. Fragmentation reduces the performance of storage systems and makes file recovery a complex task. File recovery tools like Recuva, which are often employed to recover lost or deleted files from storage media, face considerable challenges when dealing with fragmented files. These tools typically assume that files are stored contiguously, which is not always the case in real-world scenarios. Fragmented files, especially when they are large and dispersed, require specialized techniques to identify and reassemble their fragments. This

presents an urgent need for improvements in file recovery algorithms to handle fragmented and overwritten files effectively. The goal of this project is to enhance the functionality of Recuva for recovering fragmented AVI files, specifically focusing on AVI files stored on NTFS (New Technology File System), the most widely used file system for Windows. NTFS offers advantages over older systems, such as better file management and larger file support, but it also suffers from fragmentation. By using a combination of signature-based scanning, heuristic analysis, and metadata reconstruction algorithms, this project aims to develop a method to locate and recover fragmented AVI files more efficiently. Moreover, the project seeks to address the common problem of overwritten data, which makes it difficult to recover files when part of their original content is replaced with new information. This enhanced recovery technique focuses on detecting the fragmented parts of the AVI file, reassembling them, and ensuring that the file can be played without issues. Special attention is given to the recovery of metadata (such as headers, indices, and file structure), which is crucial for maintaining the integrity of the recovered files. Additionally, strategies to handle overwritten data and partial file recovery are employed to maximize the chances of retrieving as much of the original file as possible.

II. LITERATURE

2.1 File Fragmentation and Its Impact on Data Recovery

Fragmentation is a natural consequence of continuous file creation, modification, and deletion on disk storage systems. In an ideal scenario, files are stored in contiguous clusters, allowing the system to read and write data more efficiently. However, as files are edited or deleted, free space is created, and subsequent files are stored in non-contiguous spaces. This leads to file fragmentation, where the data is scattered across different disk locations. Fragmentation is especially problematic for large multimedia files like AVI, which require large contiguous blocks to function properly. The fragmented nature of these files makes it difficult for traditional file recovery tools to piece together the file correctly. Research has shown that fragmented files result in slower system performance, reduced data integrity, and increased time required for recovery. Several recovery methods attempt to address fragmentation by scanning the disk for file fragments

and reassembling them, but this becomes difficult when the fragments are not easily distinguishable or when the metadata (which defines the file's structure) is also fragmented or lost.

2.2 NTFS File System and Fragmentation

The NTFS file system, designed by Microsoft, supports larger file sizes and more advanced features than its predecessors, like FAT32. NTFS uses clusters, which are the smallest units of storage allocation. While NTFS is efficient for managing large files, it is also prone to fragmentation. When an AVI file becomes fragmented, the clusters that make up the file are scattered across the storage medium, resulting in inefficient data retrieval and the potential for incomplete recovery.

NTFS organizes data into a Master File Table (MFT), which holds the metadata of files, including their file names, sizes, timestamps, and locations on the disk. When a file is fragmented, the MFT contains pointers to the individual file fragments, but these pointers may become corrupted or incomplete over time. This complicates the recovery process, especially if the file has been partially overwritten or deleted.

2.3 Traditional Recovery Techniques

Conventional file recovery techniques primarily focus on undeleting files or restoring files from deleted sectors. When a file is deleted, the data is not immediately erased but marked as free space. Recovery tools typically scan the disk for these marked areas and attempt to reassemble files from the remaining fragments. However, when files are fragmented, this approach is not always sufficient. Tools like Recuva rely on sector-based scanning, which involves searching the disk for known file signatures, such as the header of an AVI file. These tools then attempt to reassemble the fragments based on the order of their discovery. While this technique can work for non-fragmented files, it struggles with fragmented files, especially when the fragments are dispersed or when the metadata is corrupted.

2.4 Fragment Detection Techniques

To overcome the limitations of traditional recovery tools, fragment detection techniques have been developed. One approach is signature-based scanning, where the recovery software looks for known patterns or signatures in the file structure to identify fragments. For AVI files, these signatures are typically found in the file header and other structural components that remain intact even when the file is fragmented.

Another method is heuristic analysis, which uses statistical models to predict where fragments are likely to be

found based on the layout of the file system and the patterns of file allocation. These techniques improve the accuracy of fragment detection, particularly when file fragments are scattered over a large area of the disk.

2.5 Metadata Reconstruction for AVI Files

The recovery of fragmented AVI files is not just about locating the fragments but also about reconstructing the file's metadata. AVI files contain specific structures like headers and indexes, which define how the video and audio streams are organized. If these parts of the file are corrupted or missing, the recovery process becomes more challenging.

Metadata reconstruction algorithms work by analyzing the recovered fragments and using existing patterns in the file structure to infer the missing pieces of metadata. This can include rebuilding the index that links video frames and audio samples, allowing the video to play correctly. Research has shown that predictive modeling can be used to reconstruct missing headers and indexes based on partial data, thus improving the chances of a successful recovery.

2.6 Overwritten Data Recovery

When a file is overwritten, the original data is replaced with new information, making recovery more difficult. However, research has suggested that data carving techniques can be employed to recover partially overwritten files. Data carving involves scanning the disk for remnants of the old data that may still reside in the storage clusters, even if they are not directly accessible through the file system. This technique is especially useful when the overwritten data is only partially replaced, leaving traces of the original file that can be reassembled.

Partial recovery methods are also critical when dealing with overwritten AVI files. In these cases, the tool may not be able to fully recover the file but can salvage significant portions of the data to restore part of the video or audio content.

2.7 Integration of Advanced Recovery Algorithms

The integration of advanced recovery algorithms into tools like Recuva aims to improve the efficiency and accuracy of recovering fragmented and overwritten files. These algorithms combine fragment detection, metadata reconstruction, and data carving techniques to maximize the chances of successful recovery, even in challenging scenarios. By incorporating these algorithms into Recuva, users can recover AVI files more efficiently and with higher accuracy, even when the files are severely fragmented or partially overwritten.

2.8 Fragmentation in AVI Files

When files are saved on a hard drive, they are usually stored in contiguous sectors. However, over time, as files are created, modified, and deleted, free space on the disk is fragmented. This results in files being split into smaller parts and stored in different, non-contiguous sectors. For instance, when a large AVI file is edited or partially deleted, its segments may be scattered across the disk. The storage locations for these segments can vary, making the file recovery process complex.

The challenge with fragmented files is that traditional recovery tools, including Recuva, are often designed to recover files in their entirety by scanning for a continuous sequence of data. When fragmentation occurs, the recovery process becomes more difficult because the file is broken into pieces that need to be reassembled in the correct order. Without this reassembly, the recovered file may be corrupted or incomplete.

2.9 NTFS File System

The NTFS (New Technology File System) is the default file system used in Windows operating systems. NTFS offers several advanced features over older file systems like FAT32, including better handling of large files and support for file permissions and metadata. However, despite these advantages, NTFS can also suffer from fragmentation, especially when large files, like AVI files, are frequently edited.

NTFS organizes data in clusters, and when files grow beyond the size of a single cluster, they are divided into multiple clusters. These clusters are then spread across the disk in non-contiguous locations when fragmentation occurs. Recuva, like most file recovery tools, uses the structure of these clusters to identify and recover files, but when files are fragmented, the tool must find all the scattered clusters and put them back together correctly.

2.10 The Data Used for Testing

For the purpose of testing and evaluating the enhanced recovery algorithms, several sets of AVI files were selected, representing a variety of real-world scenarios. These include:

Mildly Fragmented AVI Files: Files with minor fragmentation, where the fragments are still relatively close to each other on the disk.

Severely Fragmented AVI Files: Files with heavy fragmentation, where the fragments are scattered across the disk in distant sectors.

Partially Overwritten AVI Files: Files that have been partially overwritten, where new data has replaced parts of the original file, making recovery more difficult.

Combined Fragmentation and Overwriting: A challenging scenario involving both severe fragmentation and partial overwriting, which simulates the worst-case scenario for file recovery.

These files were intentionally fragmented and overwritten using a series of controlled procedures to simulate real-world scenarios where users may lose access to their AVI files.

III. METHODS

The methods used in this project are designed to enhance the recovery of fragmented and overwritten AVI files using the Recuva software. The process can be divided into several key stages:

3.1 Fragment Detection Algorithms

The first step in recovering fragmented files is detecting the scattered fragments across the disk. Fragmentation complicates this because traditional recovery tools typically scan files sequentially, assuming that the entire file will be found in one contiguous sequence. Fragmentation breaks this assumption, requiring a more advanced approach.

To address this challenge, the project integrates two main techniques for fragment detection:

Signature-Based Scanning: This technique relies on recognizing specific patterns or signatures in the file's structure. AVI files, like many multimedia files, have a well-defined structure that includes headers, indices, and data blocks. By analyzing the unique byte patterns in these parts, the recovery software can identify file fragments that may be spread across different sectors of the disk. Signature-based scanning is efficient because it targets known patterns, reducing the search space for fragments.

Heuristic Analysis: In addition to signature-based scanning, heuristic analysis is used to estimate the likely locations of fragmented file segments. This approach involves examining the disk's usage patterns and looking for clusters of unused space that may indicate fragmentation. Heuristic analysis makes use of statistical models to predict where the fragmented parts of a file are most likely located. This method is especially useful in cases where signature-based scanning alone is insufficient or where file patterns are less predictable.

By combining these two techniques, the enhanced Recuva tool is able to locate file fragments more accurately, even when they are spread out across large areas of the disk.

3.2 Metadata Reconstruction Algorithms

Metadata is essential for reconstructing files in a usable format. AVI files contain metadata, such as headers and indices, which organize and store the structural information needed to play the file. When these metadata structures are lost or corrupted—often due to fragmentation or overwriting—recovering the file correctly becomes extremely challenging.

To address this, the project integrates metadata reconstruction algorithms that aim to rebuild missing or corrupted metadata. The approach is based on predictive modeling, where the algorithm uses the available portions of the file to infer the missing metadata. The reconstruction process includes:

Header Analysis: The headers of an AVI file contain essential information about the video and audio streams, such as codec types, frame rates, and resolution. By analyzing the intact parts of the file, the tool can reconstruct missing portions of the header and ensure that the recovered file has the correct structural format.

Index Inference: AVI files use indices to store information about the location of video and audio frames. If these indices are missing, the recovery tool uses the positions of the recovered fragments to rebuild the index, ensuring that the video and audio are properly synchronized during playback.

The metadata reconstruction process ensures that even if the file's structure is fragmented, it can still be restored to a playable format.

3.3 Handling Overwritten Data

Overwriting is one of the most challenging aspects of data recovery. When a file is deleted or modified, the file system may reuse the space on the disk that was previously occupied by the file, making it difficult to recover the original data. In the case of AVI files, portions of the file may be overwritten with new data, reducing the chances of successful recovery.

To handle overwritten data, the project employs data carving techniques. Data carving is a process used to recover files when the file system metadata is missing or corrupted. The tool works by scanning the disk for known file signatures

and attempting to reconstruct the file based on the remaining data fragments.

The algorithms used in this project include:

Sector Pattern Analysis: When part of a file is overwritten, traces of the old data may still remain in the sectors. By analyzing the patterns in these sectors, the algorithm can differentiate between original data and overwritten data, allowing for partial recovery of the original file.

Partial Recovery Techniques: In situations where full recovery is not possible due to overwriting, partial recovery methods are used. These methods focus on salvaging as much data as possible, even if only fragments of the original file can be restored.

3.4 Integration with Recuva

Once the fragmentation detection, metadata reconstruction, and data recovery algorithms were developed, they were integrated into the Recuva software. The integration process included:

Updating the Scanning Engine: The scanning engine in Recuva was enhanced to incorporate the new fragment detection techniques, allowing it to search for fragmented file pieces across the disk more effectively.

Enhancing the Recovery Module: The recovery module was updated to support the reconstruction of metadata and handle cases where portions of the file were overwritten. The module also includes error correction mechanisms to ensure that recovered files are as accurate as possible.

User Interface Enhancements: The user interface (UI) was redesigned to provide users with detailed feedback during the recovery process. The new UI offers progress indicators, recovery success estimates, and information about the status of individual file fragments.

3.5 Summary of Methods

The methodology employed in this project involves a multi-step process aimed at enhancing the ability of Recuva to recover fragmented and overwritten AVI files. This includes detecting fragmented file segments through a combination of signature-based scanning and heuristic analysis, reconstructing missing metadata through predictive modeling, and handling overwritten data using sector pattern analysis and partial recovery techniques. These advancements were integrated into the existing Recuva software architecture to improve its functionality and recovery success rates.

IV. RESULT

4.1 Performance and Analysis of the SQL Injection Automation Tool

The enhanced version of Recuva was tested in various scenarios involving fragmented and partially overwritten AVI files. The effectiveness of the enhancements was evaluated by comparing recovery success rates, recovery time, and file integrity.

4.2 Test Scenarios

The tests were conducted under the following conditions:

- Mild Fragmentation: AVI files with minimal fragmentation.
- Severe Fragmentation: Files split into numerous fragments across non-contiguous disk sectors.
- Partial Overwriting: Sections of AVI files were overwritten with new data.
- Combined Fragmentation and Overwriting: Both fragmentation and overwriting occurred simultaneously.

Table 1: Performance Analysis of Recuva

Scenario	Fully Recovered (%)	Partially Recovered (%)	Not Recovered (%)
Mild Fragmentation	90%	8%	2%
Severe Fragmentation	65%	25%	10%
Partial Overwriting	55%	30%	15%
Combined Fragmentation/Overwriting	40%	35%	25%

4.3 Analysis

Mild Fragmentation: The enhanced tool successfully recovered 90% of the files. Only 2% were unrecoverable due to issues with the Master File Table (MFT).

Severe Fragmentation: The recovery rate was 65%, with 25% partially recovered and 10% unrecovered. The fragmentation detection algorithms improved recovery in challenging conditions.

Partial Overwriting: The tool recovered 55% of the files fully, and 30% partially. 15% could not be recovered due to heavy overwriting.

Combined Fragmentation and Overwriting: This complex scenario yielded a 40% full recovery and 35% partial recovery. 25% were not recovered, reflecting the difficulty of handling both fragmentation and overwriting.

4.4 File Integrity and Recovery Time

The recovered files maintained full functionality, with synchronized audio and video streams. Recovery time increased with higher fragmentation or overwriting, ranging from 5-10 minutes for mild fragmentation to 30-40 minutes for more severe cases.

4.5 Challenges and Limitations

- Computational Resources: Enhanced algorithms require more processing power, which may slow performance on lower-end systems.
- Heavily Overwritten Files: Recovery is significantly more difficult when files are heavily overwritten.
- File System Corruption: Corruption of the NTFS file system, such as damage to the MFT, can make recovery impossible.

V. CONCLUSION

This project successfully demonstrated the enhancement of Recuva's capabilities for recovering fragmented and partially overwritten AVI files, addressing significant challenges in data recovery from NTFS file systems. The integration of advanced algorithms into the Recuva framework, specifically designed to detect fragmented file segments, reconstruct metadata, and handle overwritten data, has considerably improved the software's ability to recover files that were previously deemed irrecoverable. The core of the improvement lies in the innovative methods introduced for fragmentation detection and metadata reconstruction. By using signature-based scanning combined with heuristic analysis, the enhanced tool can identify fragmented segments of AVI files spread across non-contiguous sectors, increasing the chances of successful recovery. The reconstruction of metadata using predictive modeling and heuristic approaches allows the tool to restore the structure of the recovered files, ensuring that they are not only restored but also usable. The approach for handling overwritten data ensures that even files with partially overwritten sectors can still be salvaged, though recovery success naturally diminishes as the extent of overwriting increases. The testing results confirmed that the enhanced version of Recuva significantly outperforms its original counterpart, especially in scenarios involving severe fragmentation and partial overwriting. Files that were previously difficult to recover due to fragmentation across non-contiguous sectors or overwriting by new data were now more likely to be restored, with notable improvements in both the rate of full recovery and partial recovery. This capability holds particular promise for data recovery specialists and forensic analysts who frequently deal with such challenging recovery conditions in their professional work. Despite the substantial improvements, the project also revealed several limitations that must be addressed in future work. The

enhanced algorithms require more computational resources, which can affect performance, particularly on low-end systems. While recovery rates were significantly better than the original Recuva tool, the performance of the enhanced tool still suffered in the most extreme scenarios, particularly where files had been heavily overwritten. This highlights the need for further optimization of the algorithms to reduce processing time and system resource requirements. Moreover, while the focus was primarily on AVI files, the techniques developed and tested during this project have the potential for broader applications. Future developments could explore extending the capabilities to other large multimedia file formats, such as MP4 and MKV, or even more complex file systems, including exFAT and EXT4. With additional fine-tuning, these recovery methods could become more generalized to handle different file types and environments.

The success of this project points toward the future of data recovery tools becoming more sophisticated and able to address real-world problems faced by both personal users and professionals in data retrieval. As data storage continues to evolve, especially with the increasing reliance on cloud storage and network-based systems, improving the effectiveness of tools like Recuva to recover fragmented and overwritten files will remain an essential task. This project provides a solid foundation for future research and development in the field of data recovery, potentially impacting industries ranging from digital forensics to cloud storage management.

REFERENCES

- [1] A. Putra, M. D. L. Siahaan, and A. Arpan, "Comparative analysis of data recovery using easeus data recovery wizard and recuva applications," *INFOKUM*, vol. 10, no. 03, pp. 161–165, 2022.
- [2] N. A. Abd Rahman, N. N. A. N. Suki et al., "A customized data recovery tool," in *Journal of Physics: Conference Series*, vol. 1712, no. 1. IOP Publishing, 2020, p. 012019.
- [3] S. N. Varayogula, K. Dodiya, P. Lakhalani, and A. Chawla, "Computer forensics data recovery software: A comparative study," *International Journal of Innovative Research in Computer Science & Technology*, vol. 10, no. 2, pp. 513–518, 2022.
- [4] B. K. Shibwabo and I. A. Lukandu, "Data recovery of flash memory," *Nanocrystals in Nonvolatile Memory*, p. 347, 2024.
- [5] A. Hidayat, D. Sudarmaji, D. Irawan, L. J. Susanto, and H. P. Mustika, "Comparative analysis of applications osforensics, getdataback, genius, and diskdigger on digital data recovery in the computer device," *International Journal of Technology & Engineering*, vol. 7, no. 4.7, pp. 445–448, 2018.
- [6] J. Buchanan-Wollaston, T. Storer, and W. Glisson "Comparison of the data recovery function of forensic tools," in *Advances in Digital Forensics IX: 9th IFIP WG 11.9 International Conference on Digital Forensics*, Orlando, FL, USA, January 28-30, 2013, Revised Selected Papers 9. Springer, 2013, pp. 331–347.
- [7] R. Hidayat, M. I. Marzuki, Y. Saragih, I. Ibrahim, E. U. Armin, S. Andromeda, I. B. Santoso, and S. Suroyo, "The comparative study analysis logical files recovery and low level files recovery using digital forensic methods," *TEKNOKOM*, vol. 5, no. 2, pp. 149–154, 2022.
- [8] A. Miljković, V. Stojanović, V. Jovanović, and A. Maksimović, "Possibilities for comparison of data recovery software for mobile devices," *DANI ARC IBALDA RAJSA ARCHIBALD REISS DAYS*, p. 111.
- [9] S. Al Sharif, M. Al Ali, N. Salem, F. Iqbal, M. El Barachi, and O. Alfandi, "An approach for the validation of file recovery functions in digital forensics' software tools," in *2014 6th International Conference on New Technologies, Mobility and Security (NTMS)*. IEEE, 2014, pp. 1–6.
- [10] D. Anjani, D. Novianti, and R. Ningsih, "Training to rescue and repair data on flashdisk storage for elementary school teachers of pal merah 19, west jakarta," *Mattawang: Jurnal Pengabdian Masyarakat*, vol. 1, no. 2, pp. 99–103, 2020.
- [11] V. Singh, L. Kesharwani, V. Saran, A. Gupta, E. Lal, and A. Verma, "Efficacy of open source tools for recovery of unconventionally deleted data for forensic consideration," *Int. J. Sos. Relev. Concern (IJSRC)*, vol. 3, no. 9, pp. 53–9, 2015.
- [12] W. Blakely Santiago, "An overview to digital forensics tools," *Computer Engineering*, 2010.

Citation of this Article:

Shivaranjini, Mahesh Babu K, & Maheswar Reddy K. (2025). Enhanced Recovery of AVI Files Using Recuva. In proceeding of International Conference on Sustainable Practices and Innovations in Research and Engineering (INSPIRE'25), published by *IRJIET*, Volume 9, Special Issue of INSPIRE'25, pp 308-314. Article DOI <https://doi.org/10.47001/IRJIET/2025.INSPIRE50>
