

# Proposal and evaluation of a high-speed data erasure method for mass storage devices

Kousuke Yoshioka† Thatsuo Yoshida† Yuki Ashino†  
Tetsutaro Uehara†† and Ryoichi Sasaki†

†Graduate School of Engineering, Tokyo Denki University

††Academic Center for Computing and Media Studies, Kyoto University

**Abstract.** In recent years, leakage of personal information has become recognized as a social issue, and erasure processing of the data in a hard disk drive (HDD) prior to PC recycling has become essential. However, HDD capacities are increasing rapidly and excessive amounts of processing time are required when utilizing an existing erasure method. To solve this problem, the authors propose an effective erasing method utilizing a load sharing erasure technique that operates while the PC is in use, as well as a high speed erasure technique that takes characteristics of the file structure into consideration. This paper describes the method and introduces a developed system based on the method, and provides the results of a functionality and performance comparison between the new system and a conventional existing method.

**Keywords:** data erasure, information leakage, mass storage device, evaluation

## 1 Introduction

In the time since Japan's PC recycle law<sup>[1]</sup> was revised on October 1, 2003, numerous PC components have been recovered in usable condition and recycled for further use. As a result, the probability that a new owner could come across a used HDD recovered from a recycled PC has increased. Furthermore, if the previous owners data was not been erased completely, it is possible for secondhand users to recover such data from the HDDs using well-known recovery software. Thus, whenever personal or confidential information is included on a HDD, there is a risk of information leakage to secondhand users. To avoid this situation, data erasure, which refers to the process of eliminating data without any possibility of future data recovery, is necessary before a PC is recycled. However, if the HDD must be erased en bloc just before a PC is recycled, a considerable amount of erasure time will be required. This is because the capacity of HDDs is increasing at an accelerating rate.

To solve this problem, the authors propose an effective erasing method that utilizes a load sharing erasure technique that operates while the PC is in use, as well as a high

speed erasure technique that takes file structure characteristics into consideration. Furthermore, we have developed a system based on this method, have confirmed the functionality and utility of the method, and have compared it with a conventional existing method.

This paper describes the method, a system that has been developed based on the method, and it provides the results of the comparison between the newly-developed system and a conventional system as part of an overall evaluation<sup>[2]</sup>.

## **2 Deletion and erasure**

In this section, we define the terms “deletion” and “erasure” and discuss the problem of achieving effective erasure.

### **2.1 Structure of a data file**

"Data" exists as a "file" on a Windows system, and a data file consists of the "management information area" and the "real data area". In the management information area, the file name, its size and the information required to access the real data are stored. The bulk of the actual data that we manipulate when using files is stored in the real data area.

### **2.2 Difference between deletion and erasure**

In general, when the "Delete" menu option is selected by the user, the corresponding data is moved to the trash box in the Windows system. Then, if the "Empty the trash box" menu option is selected, the data is eliminated. However, in this case elimination means that the information in the management information area has been removed and permission to write new data on the related real data area has been given. Therefore, the actual data of the deleted file remains on the disk after deletion and can easily be recovered using well-known recovery software such as PandoraRecovery<sup>[3]</sup>.

To avoid this situation and achieve total erasure, the user must overwrite the real data area of the file with a random number or a fixed value (such as 0xFF or the NULL value) because the real data of the file has been overwritten, it becomes impossible to recover the data, even if recovery software is used. This process is called “real erasure” or just “erasure”.

While in general, real erasure implies that the entire actual data area is overwritten with a random number or a fixed value; in this paper, we also use the term “real erasure” to refer to a technique to overwrite part of the real data area in such a way that the data can only be recovered with extreme difficulty. A high speed erasure technique of this type is described in section 4.3.

### 2.3 Existing method

Using an existing method, data can be erased by overwriting all the domains of the HDD en bloc with a random or fixed value, just before recycling the PC. In order to investigate the time required for erasing with the existing method, we measured it using existing erasure freeware named "Eraser" <sup>[4]</sup>. The measurement result is shown in Fig. 1.

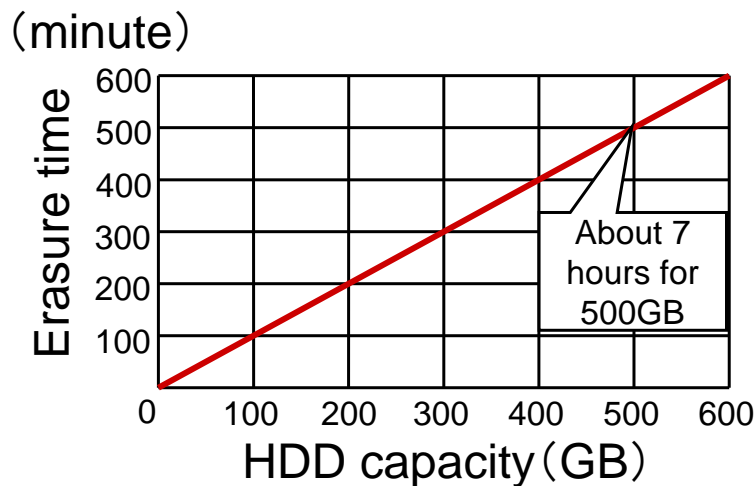


Fig. 1: Erasure time with the existing method

From this figure, we can determine that the erasure time per 1 GB data is about 1 minute and therefore the total time required to erase data from an entire 500 GB HDD would be about 7 hours. Therefore, a new method is required to decrease the time required to erase an entire HDD just before a PC is recycled.

### 2.4 Files targeted for erasure

The files that should be erased include those containing personal or confidential information. Examples of such files are provided below:

- (i) Text Files (MS Word, etc)
- (ii) Tables and Lists (MS Excel etc)
- (iii) Conference material (MS Word, MS PowerPoint etc)
- (iv) Email
- (v) Pictures, Photos (JPEG, Png, etc)
- (vi) Movies (AVI, MPEG etc)

### 3 Proposed method

#### 3.1 Basic approach

The experimental result in Section 2.3 showed that erasure takes an excessive amount of time with the conventional method. Therefore, a new method is required to decrease the time required to erase an entire HDD.

We propose a load sharing type erasure method, which does not erase data en bloc prior to recycling, but instead performs erasure when required by preset conditions, such as when the menu option “Empty the trash box” is selected. In our proposed method, the command “Erasure” is added and the user can select the pull down menu option “Erasure”. If this menu option is selected, then the system immediately overwrites the real data area of the file in question with a fixed value.

#### 3.2 New problems to be solved

##### (1) Limitations of real time erasure

Even with our method, it is time consuming to erase the data, when the file to be erased is very large. Therefore, we introduce a background erasure function in addition to the real time erasure function described above.

##### (2) Writing conflict with OS

A problem arises also with background erasure. If a file is deleted, permission to write data on the real data domain of the file is given to all programs, including the OS. If background erasure is performed just after the OS writes data to the released domain, this data will be erased, as shown in Fig. 2. If this problem cannot be solved, this method cannot be used.

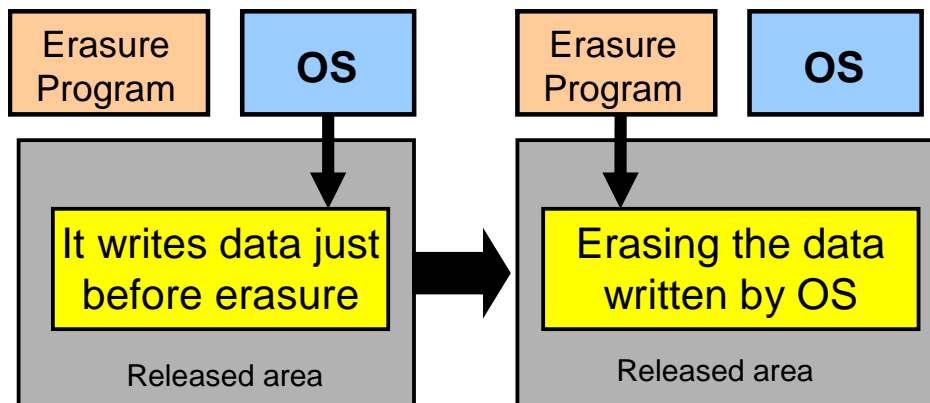


Fig 2: Example of writing conflict

As a solution, we adopted the technique whereby the file designated for background erasure is not immediately erased. Instead, the file is sent to a protected domain prepared beforehand. If the file is sent to the protected domain, it cannot be overwritten by other programs, including the OS, because the protected domain cannot be released to other programs for writing.

### **(3) Selected protected domain**

For the protected domain, we selected the trash box, which is already protected in any case.

From the beginning, the trash box in Windows was designed with a function that protects against accidental deletion. The files moved to the trash box by the “Delete” command or drag & drop are stored in the folder named recycler in the HDD<sup>[5]</sup>. Those files stored in the recycler are maintained there until the trash box accepts the command for erasure, even if the old directory for the file has been lost. Therefore, if a file in the trash box is moved to another directory, the file can be used again. We require an identical function for our protected domain.

### **(4) Improvement in erasure speed**

The data length of most of the files shown in Section 2.4 is normally less than a few dozens of MB. Therefore, the erasure time for these files is normally reasonably short.

However, movie files are very large compared with other files, even if a compression scheme is used. Therefore, it takes a long time to erase a movie file. This situation required the development of a high speed technique that could be used to erase a movie file that has been compressed with a scheme such as MPEG4.

To cope with this problem, we developed a high speed erasure technique that takes advantage of the characteristics of the MPEG4 file structure.

## **4 Development of the proposed system**

The authors developed the system proposed in Section 3. The development environment is shown in Table 1.

Table 1: Development environment

OS for development	Windows XP Professional
Development tool	VisualStudio2005
Program Language	C++

### **4.1 Structure of the system**

Figure 3 illustrates the software structure of the system used to construct the required functions for a PC running Windows.

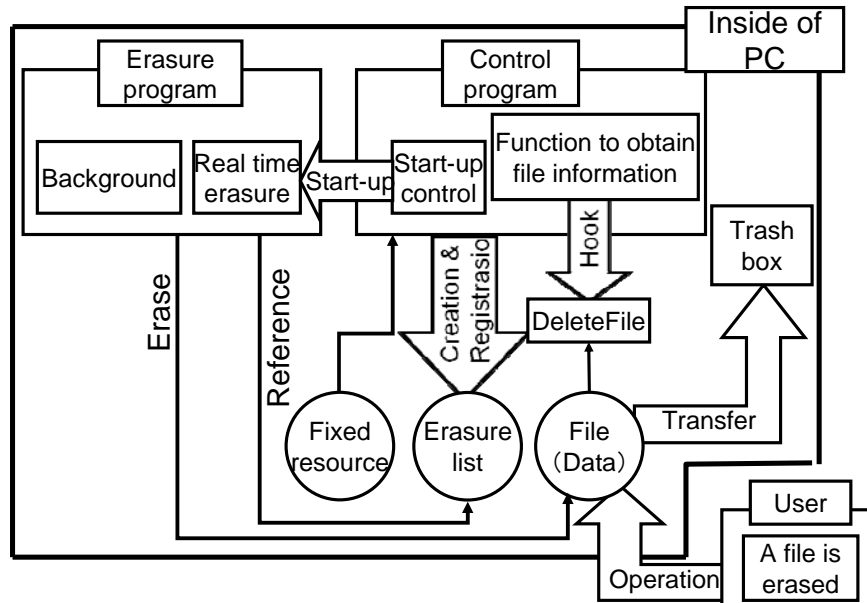


Fig. 3: Structure of the system

#### Functional requirements

The proposed system should have the following functions.

##### Control program

- a) a function to obtain file information
- b) Start-up control function
- c) Erasure list

##### Erasure program

- d) Real time erasure function
- e) Background erasure function

Each function is now explained in more detail.

##### Control program

The usual control program in the system monitors the status of PC resources, and controls the start up of these programs. The control program includes the following functions related to file erasure.

Obtain information such as the "File name", the "Extension", the "Offset", the "Size of file", and the "Time stamp".

Register the obtained file information for the erasure file.

Start up the erasure program when the computation power available is greater than a set value.

Stop the erasure program, which executes in the background, when the workload of the PC exceeds a set value.

##### A function to obtain file information

This is the function to obtain information about the file to be erased. The obtained information is used in the following situations.

Erase files immediately.

Store information on the file to be erased in the background.

Identify the appointed files for erasure from ordinal files in the Trash Box.

b) Start-up control function

The erasure program starts up only when erasure is required. The control program has a function to control the start up and shut down of the erasure program.

c) Erasure list

The erasure list is a management list in CSV file format. The information on files which are targeted for background erasure is obtained by the control program and stored in this list. When background erasure is carried out, this list is searched.

Erasure program

This program is used for erasing the file, and is carried out basically in the foreground in real time. When the file size is large, the size information is reported to the PC user. If the user agrees to change from erasure in the foreground to erasure in the background, background erasure begins. Here, background erasure is carried out by overwriting with 0xFF once on the actual data area of the file.

d) Real time erasure

In real time erasure, the appointed file is erased immediately the erasure command is given to the erasure program. The erasure command can be given by selecting the menu option "Erasure" in the context menus.

e) Background erasure

In background erasure, a large file is erased in the background using surplus resources of the PC. The user does not have to be aware of the initiation and/or termination of the program.

## **4.2 Developing the background erasure function**

Background erasure is not directly started and/or interrupted by the user as these operations are performed automatically by the control program. The initial value of the CPU rate of use in the control program is set by the PC user before the operation. As a default, the following values are set.

<Start condition of background erasure>

The CPU rate of use for starting the background erasure is 40%.

<Stop condition of background erasure >

The CPU rate of use for stopping the background erasure is 80%.

## **4.3 Development of the high speed file erasure technique**

### **High speed file erasure technique**

The high speed file erasure technique that takes account of characteristics of the file structure (referred to hereinafter as the high speed file erasure technique) overwrites a fixed value of an actual data domain (but not the entire domain), which is selected after considering the file structure. For example, when high speed erasure is carried out on a file with a "RIFF" structure in MPEG4, which is a commonly used movie file structure, only the area holding container format information is overwritten by the fixed value. Once that is done, it becomes extremely difficult to recover and

use the data. Furthermore, even if a part of the data were to be retrieved, it would be impossible to reconstruct the data, because an attacker cannot know which stream or codec should be used. Hence, it can be said that this high speed file erasure technique has sufficient strength to protect against reconstruction of the data.

#### **MPEG4**

For the following reasons, MPEG4 was selected as the movie file format from many candidates.

MPEG4 is an efficient, high-quality next generation format for movie files.

It is a widely used and well-known standard.

Improvement of MPEG4 is well advanced.

The "RIFF" <sup>[6]</sup> file structure used in MPEG4 is also used in other movie file structures such as AVI and WMA. Therefore, a technology designed for implementation with MPEG4 has general applicability.

#### **Structure of MPEG4 and the problem**

There are many data block types used in MPEG4. The number of data block types registered to the MP4 Registration Authority<sup>[7]</sup> is 124 and there are many unregistered data block types. Additionally, data block types with hierarchal structures also exist, and the rules for hierarchal structure have not been decided.

#### **Development of MPEG4 structure erase**

Although there is no standard structure in MPEG4, a rule exists for each block. In each data block, the data block size is written in the first 4 Bytes and the type of data block is provided in next 4 Bytes. Making use of this information, we develop the high speed file erasure technique described below.

- (a) The block structure of MPEG4 is extracted.
- (b) From the information extracted about the structure, the first address of each block is acquired.
- (c) The first 8 Bytes of each block are overwritten with the fixed value 0xFF.
- (d) Because, the data block "mdat" includes the information in container format, all areas in the block are overwritten.

By utilizing this technique, it is possible to expect a decrease in the computation time required for erasure. The performance achieved by the high speed erasure technique is shown in Section 5.3.

### **4.4 Consideration of operation of the proposed system**

The operational considerations of each scenario involving the introduction of this system are provided in this section. In this paper, the proposed system is assumed to have been used since the initial setup of the PC.

#### **Real time erasure**

If the user wants to erase a file immediately, this erasure type can be selected from the context menu. Although immediate erasure can be carried out easily, the user's other work may be affected due to the extra load on the PC coprocessor during the erasure. The impact resulting from this operation is evaluated, and the result is described in section 5.

### **Background erasure**

Background erasure is started with the erase program and erasure list under control of the control program. The file information in the erasure list is removed once the file targeted for background erasure has been completed.

If the background erasure is suspended due to increased PC workload, updated information about the file being erased is stored in the erasure list. Using this information, it is possible to restart erasure operations from the file position where the process was interrupted.

### **Recycle or disposal of the PC**

If a file for which erasure has not been completed remains in the trash box when the PC is recycled or disposed of, these files are first erased by the user. As a result, erasure of the file that the user selected for deletion while the machine was in use is completed. Afterwards, erasure is carried out to an arbitrary domain by the user.

### **When an HDD is lost or stolen**

When a notebook PC is lost or stolen, proper use of this system can minimize the damage, because important data but non-essential data has been erased and cannot be recovered.

## **5. Performance evaluation with experiment**

To evaluate the functionality and performance of the proposed method, the following experiments were carried out.

### **5.1 Experimental environment**

The experiment was performed using a PC with the following specifications.

#### **Experimental machine specs**

OS Windows XP Professional  
HDD 2.5 inch SATA connection 160 GB  
CPU Intel Core2 CPU U7500 1GHz  
Memory DDR2 PC4200 1GB

### **5.2 Experimental method**

#### **a) Functional experiment**

The load on the PC was measured, when the proposed system was introduced.

Experimental method

- (a) It was confirmed that all the data targeted for erasure by our proposed system had been overwritten with 0xFF.
- (b) The increase in load when erasure was being carried out was measured and compared to the load without the proposed program. Furthermore, the impact on the user when the erasure program was running was surveyed. To be more specific, the subjective influence on the user was investigated when document processing, Internet surfing and three-dimensional (3D) action games were being carried out while the proposed program was being used.

**b) Performance experiment**

The effect of erasure was investigated using the background erasure technique.

Experimental method

- (a) The erasure speed was measured when the real time erasure program was used.
- (b) Total data erased over an extended period was measured under conditions where the PC was used normally.

**c) Experiment utilizing the high speed erasure technique**

The speedup effect of the high speed erasure technique compared with the conventional erasure technique was measured.

Experimental method

- (a) A number of MPEG4 files of different sizes were created and the time required to erase each file was measured using both the high speed erasure technique and the conventional method.
- (b) We investigated whether recovery and data extraction were possible using a number of data recovery tools, including shareware.

The following three programs were used for attempted data recovery.

PandoraRecovery (Freeware)

PerfectRecovery Pro 2007

Final • data 8.0

**5.3 Experimental results**

**(a) Functionality check of file erasure**

It was confirmed that the files were overwritten with 0xFF fixed values over all domains targeted for erasure without failure. This result shows that the erasure function of the proposed system worked correctly.

**(b) Measurement of the load by the system**

As a result of the experiment, the CPU and memory use rates were obtained as illustrated in Fig. 4 and Fig.5. Here, measurements were performed over two days, in one case running the proposed system and in the other without it.

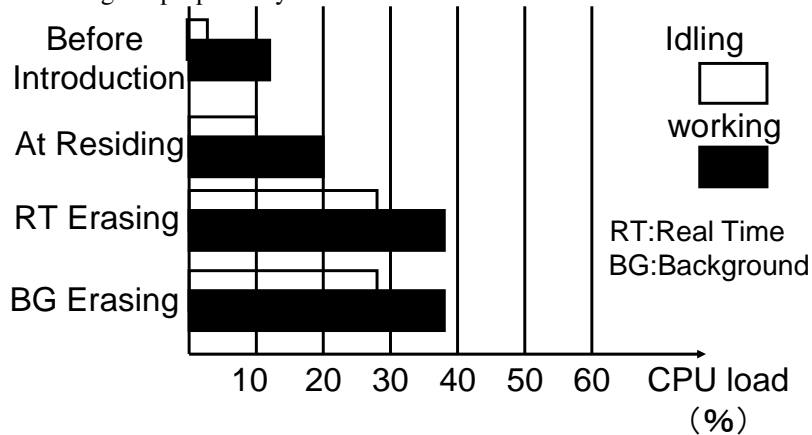


Fig. 4 CPU load in each condition

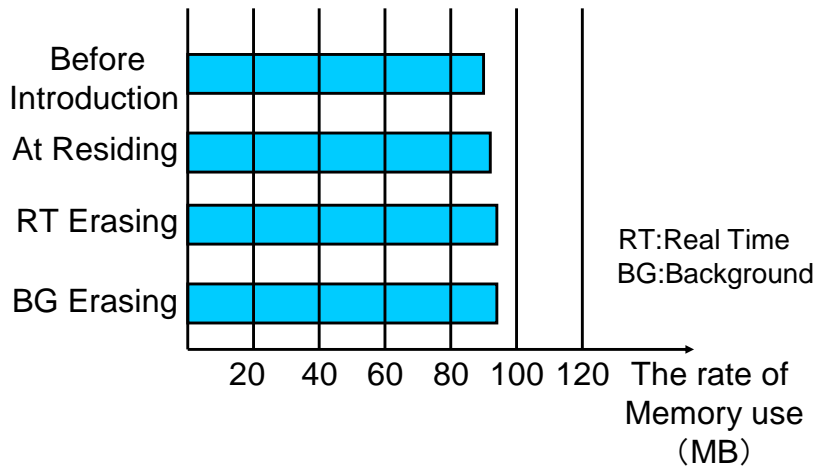


Fig. 5 Rate of memory use in each condition

As illustrated on Fig. 4, there are two kinds of bars in the graph. Here, a black bar shows the CPU rate of use in the working environment of other programs, and the white one shows the rate in an idle time environment of other programs. This experiment demonstrated the following:

- 1) The CPU usage rate increased by 8%, as shown in Figure 4, when the proposed system was introduced. Moreover, CPU utilization rate increased from 28 to 38% when the erasure was carried out.
- 2) The difference in memory usage was about 2 MB, depending on the usage environment as illustrated in Figure 5. However, the value was small compared with the measured value. Therefore, it can be said that the influence was very small.
- 3) Both of two subjects answered that they did not feel any stress when document processing, Internet surfing or playing a 3D action game, while the proposed program was operating.

For comparison of the result of (a), the CPU rate of use when general work was carried out using the PC was measured. Table 2 shows the results.

Table 2: Measurement of CPU load in PC operation

	CPU Usage Rate
Movement of Mouse Cursor	11~17
Exchange of Focus	13~19
Usage of IE	24~42
Operation of Windows	50~64

From Table 2, we can determine that the load on the CPU caused by introduction of our system is approximately the same as that caused by usage of IE. Therefore, we can say that the additional load caused by using our system is rather small.

**(c) Measurement of real time erasure speed**

The measured speed for real time erasure is shown in Table 3.

Table 3: Time for real time erasure

	1 MB	10 MB	100 MB
Erasure Time (ms)	316	3234	31537

Documents created with the MS Office software can be erased within negligible amounts of time, because the average data size of MS Office files is about 10 MB, which can be erased in about 3 seconds, as shown in Table 3.

**(d) Measurement of the effect of long term usage**

About 300 data sets per day were registered in the erasure list when the PC with our proposed system was used in the normal way. Because all data stored in the erasure list was erased within 15-16 hours, there was no overload carried forward to the next day.

**(d) Comparison of ordinary erasure and high speed erasure**

Table 4 shows the experimental results comparing the times of conventional erasure and those of high speed erasure.

Table 4: Comparison of erasure speed (ms)

	100 MB	500 MB	1000 MB
Conventional Erasure	4123	181741	523339
High Speed Erasure	499	14953	53307

From Table 4, we can determine that the processing time for MPEG4 files erased with high speed erasure is about one tenth of the ordinary erasure time. Therefore, the effect of introducing the high speed technique is substantial.

**(e) Experiment on recovery using recovery software**

The result of the experiment on recovery using recovery software is shown in Table 5.

Table 5: Result of restoration software

Used software	Restoration	Partial extraction
Pandora Recovery	×	×
Complete reconstruction PRO2007	×	×
Final data 8.0	×	×

This table shows that it was impossible to recover the file with any of the recovery software applications used.

Consideration of the possibility of recovering the file with future recovery software

Because recovery software will be improved in future, we must consider the possibility that a file might be recovered with future recovery software.

Because the data structure of MPEG4 has a hierarchical structure and there is at least one mistake in the structure, the probability of successful recovery can be

estimated as 1 part in the number of data block types raised to the power of the number of data blocks. The number of data blocks is about 100, and the number of data block types registered already is 124. Therefore, the probability of successful recovery is extremely low and this high speed erasure technique can be expected to provide sufficient safety, even in the future.

## 6 Conclusion

In this paper, we proposed an effective erasing method incorporating a load sharing erasure technique, that can be utilized while the PC is in use, as well as a high speed erasure technique that takes consideration of file structure characteristics. Furthermore, we developed a system based on this method and carried out experiments to confirm functionality and performance of the method in comparison with the conventional method. As a result of the experiments, we were able to confirm the usefulness of our proposed method.

This paper describes a situation where the proposed system was introduced at the time the PC was set up. In the future, we would like to make the system usable even when introduced during the working life of the PC.

## References

- [1] Ministry of Economy, Trade and Industry; Recycling Law for more efficient use of resources - personal computer -  
[http://www.meti.go.jp/policy/recycle/main/admin\\_info/law/02/index04.html](http://www.meti.go.jp/policy/recycle/main/admin_info/law/02/index04.html)
- [2] AOS Technologies TERMINATOR 8.0.  
[http://www.finaldata.jp/terminator/tmu\\_8.html](http://www.finaldata.jp/terminator/tmu_8.html)
- [3] Pandora Recovery <http://www.pandorarecovery.com/>
- [4] Erasure software 「Erasure」 <http://www.heidi.ie/node/6>
- [5] Microsoft : Outline of trash box.  
[http://www.microsoft.com/windows/windows2000/en/advanced/help/recycle\\_bin.htm](http://www.microsoft.com/windows/windows2000/en/advanced/help/recycle_bin.htm)
- [6] Microsoft : AVI RIFF File Reference.  
<http://msdn.microsoft.com/en-us/library/ms779636.aspx>
- [7] The MP4 Registration Authority. <http://www.mp4ra.org/atoms.html>