G Model
CULHER-3260; No. of Pages 8
ARTICLE IN PRESS

Journal of Cultural Heritage xxx (2017) xxx–xxx

Available online at
ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com/en

Original article

Preserving cultural heritage: A new approach to increase the life expectancy of optical discs

Goffredo Haus [a], Ciro Polizzi [a,b], Andrea Visconti [a]

[a] Università degli Studi di Milano, Department of Computer Science, via Comelico 39/41, 20135 Milano, Italy
[b] University of Bologna, Department of Industrial Chemistry "Toso Montanari", viale del Risorgimento 4, 40136 Bologna, Italy

ARTICLE INFO

Article history:
Received 10 March 2017
Accepted 11 August 2017
Available online xxx

Keywords:
Preserving cultural heritage
Optical disc
Aging test
Error distribution
Adaptive Reed–Solomon code

ABSTRACT

The past two decades have witnessed an exponential growth in the use of digital supports for data archiving. However, the expected lifetime of these supports is inadequate with respect to the actual needs of heritage institutions. In this paper, we address the problem of alleviating the effects of aging on optical discs. To solve this problem, we (a) experimentally recognize safe and critical areas of optical discs and (b) adopt an Adaptive Reed–Solomon (A-RS) code to increase their lifetime expectancy. More precisely, we reduce the error correction capability of the code in safe areas and increase it in critical areas. Interestingly, the approach adopted does not reduce the capacity of the discs but simply redistributes the error correction capability of the code itself. This adaptive approach helps to counteract the physical and chemical degradation of optical discs, thus increasing their lifetime expectancy.

1. Introduction

Preserving and exploiting Cultural Heritage (CH) by ICT requires to store a very large amounts of digital data—e.g., images, graphics, videos, audiovisual tracks, haptics, cataloging metadata, natural language, music, and so on—with extreme reliability and long-term life expectancy.

The most commonly used storage media are magnetic discs, optical discs, magnetic tapes, and solid state discs. Unfortunately, digital data storage does not have the same life expectancy of traditional analogical media such as, for example, the paper sheets books and other textual documents. Therefore it is important to adopt good practices and set early intervention measures to ensure long-term preservation of the data. Moreover, there is often confusion between the survival practices of a data archives and backup operations. Essentially, a backup is designed as a short-term insurance policy to facilitate disaster recovery, while an archive is designed to provide ongoing rapid access to great amount of long-term information. Most backups are retained only for a few days, or weeks, as later backup images supersede previous versions. Archived data records are placed outside the backup cycle for long periods of time. An effective data archiving strategy is a necessary part of every heritage archiving organization.

According to the guidelines provided by the UNESCO/PERSIST Content Task Force in 2016 [1], since the quantity of digital data will constantly increase over time, it is fundamental for heritage institutions to adapt their existing approaches to digital environments. However, digital data storage might be compromised. Therefore, we have to provide a periodic media refresh, usually by reading the digital data, checking for errors, and rewriting them on new media. Due to this, the knowledge of the mean life expectancies of different type of data storage is fundamental, and this topic has been deeply investigated [2–10].

Notice that different storage items belonging to the same typology may show different lifetimes, due to manufacturing technologies, and/or different external environment, and/or different quality and nature of the materials. Several studies have been conducted to test the longevity of optical discs [11–15]. They showed a wide variations in life expectancy, spanning from less than fifteen years (discs maintained at 25 °C and 50% Relative Humidity) to 217 years (25 °C, 40% Relative Humidity). Obviously, discs exposed to more severe conditions would be expected to have a shorter life. Therefore, it is difficult to foresee the life expectancy of this storage media. To overcome this problem it should be necessary to perform extensive statistical analyses on large numbers of specimens [16,17] with a significant investment in terms of time and economical resources.

Even if optical data storage cannot be considered the most advanced solution, optical storage media are far from obsolete, both for the great amount of already fulfilled archives, and recent developments of optical media technologies. Relevant efforts are still

E-mail addresses: goffredo.haus@unimi.it (G. Haus), ciro.polizzi@unibo.it (C. Polizzi), andrea.visconti@unimi.it (A. Visconti).

https://doi.org/10.1016/j.culher.2017.08.004
1296-2074/

focused in this field. For example, Sony, in collaboration with Panasonic, has recently presented new Optical Disc Archives based on high capacity Blu-Ray discs: Generation 1 (dated 2014) has cartridges with a storage capacity of 1.5 TB [18], Generation 2 (dated 2016) has cartridges with a storage capacity of 3.3 TB [18], and Generation 3 (still in development) should have cartridges with a storage capacity of 5.5 TB [18].

Optical discs are generally considered for heritage archives *warm* (occasionally accessed), or *cold* data (infrequently accessed). Optical discs provide backward compatibility between different types of optical discs. They are usually quite durable, water-resistant, and allow for random access. Being removable media, long-term reliability, low cost maintenance and data integrity are important factors, while access performance is not. It is generally accepted that optical discs are compliant with all these requirements, thus making them suitable storage media for warm and cold archives [18].

1.1. Motivations and contributions

The past decade has witnessed an exponential growth in the use of digital media for big data archiving. However, as raised by UNESCO [19,1], the expected lifetime of these storage media is inadequate with respect to the actual needs of cultural heritage institutions.

In this work, we address the problem of alleviating the aging effects on optical discs[1]. To this aim, we suggest a new approach that is able to contrast the physical and chemical degradation of optical discs, increasing their lifetime expectancy. Doing so, we

- investigate the degradation processes of optical discs by accelerating the long-term chemical changes that normally occur in discs;
- collect and analyze data before and after the accelerated aging test;
- identify if there are particular areas in which disc degradation is usually faster, or slower, than average;
- exploit the experimental activities to develop an Adaptive Reed–Solomon (A-RS) code which allows to improve the error correction capability in areas that degrade faster than other areas of the disc.

1.2. Organization of the paper

The remainder of the paper is organized as follows. In Section 2, we briefly introduce Reed–Solomon and Adaptive Reed–Solomon codes. In Section 3, we experimentally analyze the error distribution function in optical discs (RS code-based system) before and after an aging cycle. In Section 4, we describe a new approach based on an Adaptive Reed–Solomon code that redistributes the parity bytes of the code itself. In Section 5, we present the experimental activities performed to evaluate the new approach (A-RS code-based system). Finally, conclusions are drawn in Section 6.

2. A brief introduction to RS and Adaptive RS codes

Reed–Solomon (RS) codes are error-correcting codes introduced by Irving S. Reed and Gustave Solomon in 1960 [20]. They have been implemented in many real-world applications such as optical discs (CDs, DVDs, Blu-Ray Discs), RAID storage system, QR Codes, WiMAX, Digital Video Broadcasting, and many others. In this section we briefly introduce the main idea of Reed–Solomon codes. Readers who are not familiar with these concepts can find detailed information in [21–24].

A Reed–Solomon code is a cyclic code that can be described as (n, k) code, where n is the block length in symbols[2] and k is the number of information symbols. The redundancy used to correct errors in a data block is defined by $n-k$, also called parity bytes or p for short. This RS code can correct up to t errors, where t is equal to $\lfloor(n-k)/2\rfloor$, namely, $(n-k)/2$ (for $n-k$ even) or $(n-k-1)/2$ (for $n-k$ odd).

Notice that both information and parity symbols of a RS code are elements of a Galois Field—i.e., $GF(q^m)$ where q is prime, and in our case q is equal to 2.

A generator polynomial $g(x)$ of the RS code consists of $n-k=2t$ factors,

$$g(x)=(x-\alpha^0)(x-\alpha^1)\cdots(x-\alpha^{2t-1})$$

where α^i are elements of $GF(q^m)$.

A message of k information symbols is encoded as follows. Firstly, it is represented by a polynomial $f(x)$ of order $k-1$

$$f(x)=f_{k-1}x^{k-1}+\cdots+f_2x^2+f_1x+f_0$$

where the coefficients $f_{k-1}\cdots f_0$ are the information symbols encoded as m-bit value of $GF(q^m)$. Then, such a $f(x)$ is multiplied by x^{n-k} and divided by $g(x)$, getting a quotient $q(x)$ and a remainder $r(x)$. Finally, a code-word $w(x)$ is generated by multiplying $f(x)$ by x^{n-k} and adding remainder $r(x)$:

$$\begin{aligned} w(x)&=f(x)\times x^{n-k}+r(x)\\ &=f_{k-1}x^{n-1}+\cdots+f_0x^{n-k}+r_{n-k-1}x^{n-k-1}+\cdots+r_0 \end{aligned}$$

Notice that the code-words so defined are always divisible by the polynomial generator—i.e., $w(x)=f(x)\times x^{n-k}+r(x)=g(x)\times q'(x)$. If a received code-word is not, it means that at least one error has occurred. If the number of errors occurred is bigger than $t=\lfloor(n-k)/2\rfloor$, they exceed the maximum error correction capability of the code and are not correctable. On the other hand, if the number of errors is smaller than $t=\lfloor(n-k)/2\rfloor$, they can be corrected computing the syndromes [21,22].

On the contrary, an Adaptive Reed–Solomon (A-RS) code [23] provides the possibility to increase, or decrease, the correction capability, by modifying the redundancy of the code. Indeed, an Adaptive Reed–Solomon code can adopt two different strategies. Firstly, length of the information symbols k is kept constant and the number of parity bytes $p=n-k$ is increased or decreased, thus changing the length of the code-words n [25]. Secondly, the length of the code-word n is kept constant and the number of parity bytes $p=n-k$ is increased or decreased, thus changing the length of the information symbols k. In this paper, we will adopt this second approach.

3. Optical discs: error correction capability before and after an aging cycle

In this section we evaluate the error correction capability of optical discs before and after an accelerated aging cycle, investigating the evolution of the error distribution function over time.

A Climatic Chamber (CC) and an Automatic Testing Machine (ATM) have been designed and engineered at Department of Industrial Chemistry of the Bologna University. The CC is used to induce on the discs an accelerated aging process, while the ATM writes data and detects disc errors (before and after an accelerated aging process). The ATM consists of a sealed chamber equipped with:

[1] We analyzed only recordable discs, i.e., BD-R, DVD-R, and CD-R.

[2] $n\leq 2^m-1$, where each symbol is represented as an m-bit value.