COPYRIGHT PROTECTION OF DIGITAL IMAGES OF CULTURAL HERITAGE. CASE STUDY: "ULYSSES", THE OFFICIAL WEB PORTAL OF THE HELLENIC CULTURE

Dimitrios K. Tsolis[1], Vassilios Fotopoulos[2], George K. Tsolis[1], Emmanouil G. Karatzas[1], Spiridon N. Nikolopoulos[1], Theodore S. Papatheodorou[1] and Athanasios N. Skodras[2]
[1]Dept. Comp. Eng. & Informatics, [2]Dept. Physics
University of Patras, Greece.
[1]{dkt, gkt, ekaratzas, nikolopo, tsp}@hpclab.ceid.upatras.gr,
[2]{vfotop1, skodras}@physics.upatras.gr

Abstract – The main issue addressed in this paper is the implementation of a robust watermarking method and an innovative user interface for the copyright protection of digital images of cultural heritage. The complete environment is being pilot tested through "Ulysses" the official portal of the Hellenic Ministry of Culture (HMC). The issue of copyright protection of the digital content of Cultural Organizations is emerging, especially now that this content is important to be accessible through the Internet for educational and / or financial purposes.

INTRODUCTION

Advances in technology have improved the ability to reproduce, distribute, manage and publish information [1]. Reproduction costs are much lower for both right holders (content owners) and infringers, and digital copies are perfect replicas. The average computer owner today can easily do the kind and the extent of copying that would have required a significant investment a few years ago. In addition, the computer networks have changed the economics of distribution. Networks enable sending multimedia content worldwide, cheaply and at a high speed. As a consequence, it is easier and less expensive both for a rights holder to distribute a work and for an individual to make and distribute unauthorized copies. Finally, the World Wide Web has altered at a fundamental way the publication of information, allowing everyone to be a publisher with worldwide reach. The variety of documents and multimedia content of all sorts on the Web demonstrate that many people worldwide are making use of that capability. This is affecting the Cultural Sector too.

The information structure has been integrated into everyday life, affecting directly the intellectual property legislation [1]. Today, actions that can be taken casually by the average citizen – downloading files, forwarding information found on the Web – can at times be violations of intellectual property laws. Consequently, individuals in their daily lives have the capability and the opportunity to access and copy vast amounts of digital information, yet lack a clear picture of what is acceptable or legal. On the other hand, the necessary amendments of the copyright legislation in several cases do not cope with the entirety of the problem, resulting in certain legislative weaknesses.

ULYSSES – PORTAL OF THE HELLENIC CULTURE

"Ulysses" [2] is one of the largest and most highly acclaimed cultural web sites worldwide. As the primary WWW site of the Hellenic Ministry of Culture (HMC), "Ulysses" provides bilingual information on almost every aspect of Greek culture through 15,000 web pages with rich multimedia content. Sections on Greek cultural heritage and contemporary cultural

creation are available, including hundreds of archaeological sites, museums and monuments, cultural organizations, exhibitions, special issues. The images presented through Ulysses' web pages are being selected by the archaeologists and scanned at a high quality (tiff compression). The digital images are stored simply in a logically structured file system and managed at a simple way. Metadata for the digital images are not being used and preserved. In addition, the digital images are being hard coded into HTML. The published on-line digital images are being compressed to JPEG format. The copyright of the digital images is not being protected. Access control, watermarking and encryption are being applied neither to the high quality digital images nor to their compressed surrogates. Finally, a copyright management framework does not exist.

The requirements above prove that the architecture used for the "Ulysses" portal especially for the digital images management and copyright protection surely does not reflect to an efficient and adequate infrastructure. The main weaknesses could be summarized to the next important points:

- The images should be digitized using custom made scanning tools and interfaces that should utilize the existing network infrastructure.
- The digital images should be stored in a relational and distributed database system.
- International metadata standards should be used for the description of the digital images.
- The copyright of the digital images should be protected using robust watermarking.
- The watermarking method should be robust mainly at Jpeg compression, the print - scan attack (digital to analog) and the usual geometrical attacks (rotation, cropping, etc.).
- The copyright management should be based on metadata standards for IPR.

Based on the above important points a system dealing with the above issues should be designed and implemented.

WATERMARKING METHOD

Watermarking principles are used whenever the cover-data is available to parties who know the existence of the hidden data and may have an interest removing it. In this framework the most popular and demanding application of watermarking is to give proof of ownership of digital data by embedding copyright statements. For this kind of application the embedded information should be robust against manipulations that may attempt to remove it. Many watermarking schemes show weaknesses in a number of attacks and specifically desynchronization is a very efficient tool against most marking techniques. This is leading to the suggestion that detection, rather than embedding, is the core problem of digital watermarking [3], [4].

According to the above the first most important step towards the implementation of the system is the selection and evaluation of the watermarking method. The method chosen is mainly based on the further elaboration of the MCWG watermarking tool, focusing on constructing a more efficient detection mechanism, which is resulting to a more robust watermarking technique. The core of the MCWG tool is a transform domain technique that is based on the use of the Subband DCT transform [5]. The marking formula is the same well known multiplicative rule used in the large majority of the existing literature. The tool has performed positively in the past in a large variety of attacks, including those that an application like Ulysses would require. It did not though provide support for geometrical attacks. Thus some improvements were considered necessary.

The basic idea that directed the implementation was to use the QBIC (Query By Image Content) techniques and construct a synchronization scheme for the attacked images. This image registration process is based upon a small amount of side information that is stored in the database and contains important characteristics for the image. There exists a one to one

connection between the images and the side information and the retrieval of the appropriate information is accomplished through an image similarity query. This scheme was integrated into the detection process of the watermarking application in the form of a portable and reusable Dynamic Link Library (DLL). Two corresponding methods were implemented which incorporate the embedding process, the attacked image synchronization and the detection process.

METADATA

The metadata set that was used for Ulysses image repository is based on two international metadata standards. The first is used for the description and dissemination of digital images of cultural heritage (Dublin Core) and the second for the management of copyright information (DIG-35). The DIG 35 Specification "Metadata for Digital Images", Version 1.1 [6] has a very important role in the selection of fields and tables, as far as the metadata for the digital images are concerned. This metadata standard is already being widely used in simple end-user devices and even to worldwide networks. The database structure has also a special focus on metadata for the Intellectual Property Rights management.

Based on the above already implemented metadata framework, which is also incorporated to the distributed digital image repository of "Ulysses", a technical study has been completed referring to the future interoperability of the final system with tools for digital rights management. From a more technical point of view the image similarity query returns an image id that is used to retrieve the corresponding side information. Based on this information the detection process applies the synchronization scheme and checks for the watermark existence. This retrieved id could be used as a very powerful information component that assists the watermark detection and also supports a digital rights management strategy. The id interconnects the digital image with an important amount of information that can be structured upon a digital rights management prototype (XrML), validated by an appropriate software (content guard) and perform limitations to the use of the digital object.

The combination of a robust watermarking technique with an image similarity based synchronization scheme and a complete image metadata set with the prototype xml templates and the validating software, results to a very sophisticated and powerful application appropriate for a significant image repository like Ulysses.

USER INTERFACE

The User – Interface implementation is based on an integrated platform, which incorporates the scalable and distributed digital image repository. The digital repository and the whole integrated platform are based upon previous experience gained by the already completed project PRAXITELIS [8]. The functions supplied to the end user are briefly presented bellow:

- Authorization procedures for the database administration and the personal settings management for every user of the system. The user has the authority to construct his own collections and keep a history file for his past actions.
- Import and export high quality digital images. Management and search capabilities for the digital images and their metadata. The digital images are being directly inserted into the database through the system via a path filename, if the image is already digitized, a digital camera or a scanner (a customized scanner preview is available).
- Watermarking: The watermark is embedded transparently into the high quality digital image as well as into the lowest quality replicas generated for the system needs. An insertion of a digital surrogate into the central database causes the automatic embedding of

a digital watermark. The user has the capability of initiating the detection process of a watermark in order to check the copyright ownership.

- For this project the usage of XML [7] is supporting the interoperability of the image repository with equivalent ones. The application creates XML documents containing information from the advanced digital image repository, based on the end user driven queries to the database. The end user is able to save his query results as XML files which can be reused e.g. for the web based presentation of part of the digital content.
- Automated, batch insert and export of digital images from and to the database. This function aims at the mass processing of digital images stored in a CD-ROM.

EVALUATION

The image repository of the Official Web Portal is open to public via the Internet. As a result the jpeg compression, rescaling and analog to digital attacks are the most common. The watermarking robustness evaluation against those attacks is more critical than towards any other kind of attack. In order to make our tests a 220 – image sample from the repository was used as long with the Imaging Kodak control so as to perform the necessary attacks. The results are shown in the following table.

Compression Name	**Down sampling**	**Miss-Rate**
High comp, Low qual.	1:96	18%
High comp, Med. qual.	1:56	11%
High comp, High qual.	1:20	1%
Med comp, Low qual	1:85	12%
Med comp, Med qual	1:57	5%
Med comp, High qual	1:18	0%

It is important to mention that except from the third and the sixth compression type in all other cases (1:96, 1:56, 1:85, 1:57) the images were severely damaged and their further commercial exploitation is by all means impossible. It is very important for the proposed watermarking application to be able to counter rescaling attacks. In the above sample of pictures a downscaling of 2 and 3 times was performed and the results were quite encouraging (Table 1).

Tiff Image format	
Down Sampling	**Miss Rate**
1:2	3%
1:3	5%

Table 1

Image File	**Image Compression**	**Print Quality**	**Result**
Tiff	None	Normal	Detected
Jpeg	Medium Comp. High Quality	Normal	Detected
Jpeg	Medium Comp. Medium Quality	Best	Detected
Jpeg	Medium Comp Medium Quality	Normal	Missed

Table 2

A very useful and powerful feature for a watermarking application is to be able to detect the watermark, even if the image has suffered the digital to analog attack. A small random sample was selected for the experiment. In order to have a more realistic attack except from the digital to analog attack the jpeg compression was performed additionally. Some results are shown in Table 2.

CONCLUSIONS - FUTURE PERSPECTIVES

Designing and implementing a watermarking application for “Ulysses” the official portal of the Hellenic Culture is a demanding and complicated task. Most of effort was dedicated to meet the watermarking functional and technical requirements and to deal with the lack of basic infrastructure as far as the copyright protection is concerned. The result is a powerful watermarking application that plays the role of an integrated platform for the storage and management of the digital images, their metadata and most of all for the copyright protection via digital watermarking. The MCWG watermarking method was further elaborated mainly towards a more powerful detection method and reconstructed as a portable and reusable Dynamic Link Library. The watermarking application was structured upon this method, benefiting by its robustness. The main contributions of this scheme is the improvement of robustness against common attacks and the ability of proving that a random watermarked image, probably downloaded through the Internet, corresponds to a specific image stored in the “Ulysses” repository. Consequently, a) it is proved that the HMC is the copyright owner of the specific image and b) the corresponding key for retrieving all the copyright related information (copyright plate, important dates, contact points, etc.) is now available.

Based on the above, in order to optimize the robustness of the watermarking application a more thorough research on the image similarity query should be initiated. In addition, the user – interface (client) should be developed focusing on the copyright information retrieval and management towards a more complete digital rights management system.

References

[1] Computer Science and Telecommunications Board, National Research Council, “The Digital Dilemma: Intellectual Property in the Information Age”, *National Academy Press*, Washington, 1999, pp. ES 2-3.

[2] The Hellenic Cultural Heritage Portal, http://www.culture.gr.

[3] S. Katzenbeisser, F. A. P. Petitcolas, “Information Hiding – techniques for steganography and digital watermarking”, *Artech House, Computer Security Series*, 2000, pp. 95-172.

[4] P. Wayner, “Disappearing Cryptography – Information Hiding: Steganography and Watermarking”, Second Edition, Morgan Kaufmann, pp. 291-318.

[5] V. Fotopoulos, A.N. Skodras, “A Subband DCT Approach to Image Watermarking”, *Proc. X European Signal Processing Conference (EUSIPCO-2000)*, Tampere, Finland, 5 - 8 Sept. 2000.

[6] Digital Imaging Group, “Inc: DIG35 Specification - Metadata for Digital Images. Version 1.0”, 2000.

[7] World Wide Web Consortium. Extensible Markup Language (XML). http://www.w3.org/XML.

[8] Dimitrios K. Tsolis, George K. Tsolis, Emmanouil G. Karatzas, Dimitrios A. Koutsomitropoulos and Theodore S. Papatheodorou, “Copyright Protection & Exploitation of Digital Cultural Heritage”, *Proceedings of EVA 2002 Florence*, pp. 163-167.