



Article

An Authorship Protection Technology for Electronic Documents Based on Image Watermarking

Anna Melman ¹, Oleg Evsutin ^{2,*} and Alexander Shelupanov ¹

¹ Department of Complex Information Security of Computer Systems, Tomsk State University of Control Systems and Radioelectronics, 634050 Tomsk, Russia; anessa.kas@gmail.com (A.M.); saa@fb.tusur.ru (A.S.)

² Department of Cyber-Physical Systems Information Security, National Research University Higher School of Economics, 123458 Moscow, Russia

* Correspondence: evsutin.oo@gmail.com

Received: 15 October 2020; Accepted: 17 December 2020; Published: 20 December 2020



Abstract: In the field of information technology, information security technologies hold a special place. They ensure the security of the use of information technology. One of the urgent tasks is the protection of electronic documents during their transfer in information systems, including smart systems. This paper proposes a technology for protecting electronic documents containing digital images. The main idea is that the electronic document authorship can be protected by embedding digital watermarks in the images that are contained in this document. The paper considers three cases of using the proposed technology: full copying of an electronic document, copying of images contained in the document, and copying of text. It is shown that in all three cases, the authorship confirmation can be successfully implemented. Some areas of the proposed technology application are described, including augmented reality applications. Computational experiments are conducted with robust watermarking algorithms that can be used within the technology. A scenario of technology implementation is proposed which provides for the joint use of different class algorithms.

Keywords: information security technologies; authorship protection; electronic documents; digital watermarking; digital images

1. Introduction

The widespread introduction of information technology in all spheres of life and the growing popularity of intelligent data-processing systems and immersive technologies lead to the fact that paper media are gradually losing their popularity. This means that the number of documents submitted in electronic form is constantly increasing. Many private and public institutions use electronic document management systems. A large number of people refuse to buy hard-copy books, newspapers and magazines in favor of their electronic versions. Various advertising materials, posters and announcements are also successfully distributed in electronic form. Even if the document is stored as a hard copy, it can be easily scanned using any modern smartphone for further processing or activation of the augmented reality application [1,2]. It is highly convenient for users, but at the same time, it is highly convenient for intruders who can easily appropriate other people's intellectual work. Therefore, at present, the problem of protecting the authorship of electronic documents is relevant and requires close attention.

The main way to protect the authorship of electronic documents is the use of a digital signature. A digital signature is an alternative to a handwritten signature and is used when organizing a secure electronic document flow. The digital signature allows to confirm the fact of changing an electronic document after it was signed. However, it does not allow to establish the fact of illegal copying of information, including digital images, into other documents. In some cases, it may also be necessary to

protect individual document fragments. A digital signature can be applied to individual fragments of an electronic document, but the simultaneous transmission and storage of a large number of digital signatures causes inconvenience.

Along with the use of cryptographic methods for data protection, it is effective to embed additional data into digital covers, mainly into multimedia content. There are two directions for data hiding in digital objects: steganography and watermarking.

Steganography [3–5] aims to protect the confidentiality of information. Protected data are embedded in some cover object and become invisible to third parties. Steganography is an alternative to encryption. Encryption makes information unreadable in the absence of a secret key. However, encryption does not hide the fact that protected information exists. Steganography conceals the very fact of secret message existence.

Digital watermarks [6–8] are designed to protect the authorship or integrity of the cover object itself. The extraction of embedded information allows to confirm, for example, that a given digital object was created by a specific person or software and has not undergone any changes. To do this, the extracted watermark is usually compared to the original one. The extracted watermark and the original watermark are expected to match each other. The match can be complete or partial. It depends on the specific use case for digital watermarks. A watermark is an alternative to a digital signature. Sometimes, these two technologies are used together [9,10].

The main idea of our research is to protect electronic documents by embedding digital watermarks in digital images contained in these documents. Unlike text, which is often processed when copied, images are usually copied unchanged, since the reproduction of such an image can be time consuming or even impossible. Minor processing of the copied image does not destroy the embedded watermark. This allows to successfully prove authorship, if necessary. This technology can provide authorship protection for hard copies of documents when they are scanned using special smartphone applications.

It is obvious that protected electronic documents must contain some kind of graphic object to implement this approach. However, at present, this is not a problem, since many electronic documents are accompanied by various illustrations, diagrams, photographs, logos and other graphic elements.

Thus, the paper proposes a new technology for protecting the authorship of electronic documents using digital watermarks. The contributions of our research are as follows:

- A technology for protecting the authorship of electronic documents by digital watermark embedding into images contained in electronic documents is proposed. An important advantage of our technology is the use of a set of watermarking algorithms when dealing with images of different types.
- All possible scenarios for the implementation of this authorship protection technology are described and analyzed depending on which part of the document (full document, only text or only images) is copied by the plagiarist.
- The limits of applicability of the proposed technology are investigated using several watermarking algorithms of different classes. Experiments are performed with both classic detailed images and poorly detailed synthesized images (illustrations).

The rest of the paper is organized as follows. Section 2 contains the literature review. In Section 3, we propose the authorship protection technology for electronic documents and we also describe its application scenarios. Section 4 shows the results of computing experiments with three different robust watermarking algorithms. Section 5 contains a discussion of the results. Section 6 sums up our research.

2. Related Work

The authorship protection of electronic documents without the use of cryptographic techniques is being investigated by many authors, but a universal solution still has not been found. An actively developing area of authorship protection of the electronic documents is the embedding of digital

watermarks into documents. Watermarks can either be generated based on the protected text itself, or they can contain some information about the author.

Text watermarking methods are divided into two large classes: linguistic and structural [11,12].

Linguistic methods are based on changing the syntactic or semantic structure of the text. For example, the authors of [13] propose a method for hiding watermark bits by making lexical or syntactic changes in texts in German. The changes are based on grammar rules related to negation, letter doubling, adjective formation, etc. The paper [14] presents a method for embedding digital watermarks in texts in English. The authors use grammatical rules of the language and the most commonly used words, such as conjunctions, pronouns and modal verbs, to form watermarks. These watermarks are then embedded in web documents.

Such methods are resistant to text formatting as well as retyping attacks. However, embedded watermarks are destroyed when the text is reworked—for example, by replacing the word order or using synonyms. Another disadvantage is the very fact of changing the content of the protected text, which in some cases is unacceptable.

Structural methods deal with the formatting of the text. They involve various manipulations with the appearance and the arrangement of text characters. For example, the authors of [15] propose to embed a watermark based on biometric characteristics into a document by shifting lines relative to each other. The paper [16] describes a method for protecting text in Arabic, which includes modifying the lengths of spaces and using a special character of the language. At the first stage of the algorithm presented in [17], a watermark is generated based on the user's password and the original text using a hash function, and at the second stage, the created watermark is embedded into the text using homoglyph characters. In [18], it is proposed to change the line spacing when embedding a watermark to ensure resistance to printing and scanning attacks. A common drawback of such methods is that the text formatting can be easily changed using any modern text editor. As a result of editing, the embedded information is lost.

The popularity and widespread use of the PDF format for storing electronic documents has attracted the attention of watermarking method researchers. A common approach to embedding digital watermarks in PDF documents is character position manipulation. Some embedding schemes change the length of spaces between words while others change the length of spaces between characters within words. The watermarking scheme presented in [19] is based on the first method. Furthermore, its distinctive feature is the introduction of changes not in the length of spaces themselves but in the frequency spectrum of the corresponding vectors, obtained using the discrete cosine transform (DCT). The scheme in [20] is based on the second method. The authors point out that in PDF documents, each character has a coordinate pair, x and y , that locates the character horizontally and vertically within two-dimensional coordinate space. The authors propose to change x -coordinates when embedding a digital watermark. The development of this scheme is presented in a later work [21]. These methods allow to protect a PDF file from unauthorized distribution. However, when copying text content, all information about character positions is lost and embedded digital watermarks become useless.

A separate class of methods corresponds to embedding information into text images. Text images are images that contain text, such as scanned copies of paper documents. The author of [22] proposes a cloud-based approach to embedding digital watermarks in text images. Another example of text image watermarking is presented in [23], where the watermark is embedded in pixels using the linear interpolation technique. The embedding method described in paper [24] combines two frequency transforms: integer wavelet transform (IWT) and DCT. The authors demonstrate the effectiveness of this method using the example of images containing text in Arabic. The authors of the study [25] propose to change the boundaries of text characters in an imperceptible way that ensures resistance to printing and scanning.

In [26], a method is proposed that combines PDF document protection and image watermarking technology. This method segments a PDF document page into blocks and divides them into two types: texture blocks and non-texture blocks. The first type includes blocks containing graphic objects or text

fragments. The second type of blocks are homogeneous blocks that contain no information at all or contain a small amount of information. The watermark is embedded in texture blocks with adaptive selection of the embedding strength according to some robust image watermarking algorithm.

In [27], the authors propose to embed additional information into images contained in Microsoft Word documents. The authors propose to hide secret messages in this way for their subsequent transmission. However, in this case, embedding is not used to protect the authorship of the document but to ensure the confidentiality of the embedded information. The resistance of the message to any typical transform is not investigated.

It should be noted that research in the field of image data hiding has been actively developing in recent years. Steganographic embedding provides an invisible transfer of information within the image [28–30]. Watermarks are designed to control the integrity, authenticity, and authorship protection of images. Watermarking in digital images contained in electronic documents uses the same techniques as watermarking in individual images. We also give a brief overview of the relevant research area current state.

Digital watermarks differ in the level of resistance to distortion. Fragile watermarks are destroyed whenever a cover image is changed. Semi-fragile watermarks can withstand some attacks, such as moderate JPEG compression. Robust watermarks are also detectable after more significant distortion of the cover image. A digital watermark itself is usually either some kind of user information, such as an image or text, or it is a bit sequence of limited size, generated based on user data or the cover image content. The embedding of a large amount of information as watermarks is not a promising area of research, since it usually does not allow to develop robust embedding algorithms. Therefore, the papers presented here describe the embedding of small watermarks.

Image watermarking methods are divided into spatial and frequency domain methods. Spatial domain methods work mainly with the pixel values of images. An example of a recent spatial domain method with a declared high efficiency is [31]. The authors of this study propose embedding a watermark by quantizing the pixels of a digital image. A binary image is used as a watermark. Arnold transform is previously applied to it to improve security. An embedding path is determined by a hash pseudorandom scrambling algorithm. A feature of the algorithm is calculating DC coefficients of two-dimensional discrete Fourier transform (DFT) without performing a real 2D DFT to reduce the running time of the algorithm.

Frequency domain embedding uses different frequency transforms, for example, DCT, DFT, discrete wavelet transform (DWT) and others. Many methods, especially early ones, used DFT, for example, [32]. There, a specially generated circular watermark is embedded in the amplitude spectrum in an additive and multiplicative manner. In [33], the watermark bits are first embedded in the DFT amplitude spectrum, and then the 2D histogram of the chromatic components Cb and Cr is modified.

In recent years, most researchers dealt with DCT and DWT. For example, the authors of [34] hide a watermark in the mid-frequency DCT coefficients of digital images. The authors embed watermark bits in 8x8 blocks. One bit of the watermark is embedded in one image block by changing the difference between adjacent blocks. The authors divide the possible values of the difference into intervals that correspond to 0 or 1. They change the frequency coefficients so that their difference falls within the corresponding interval. Another example of a DCT-based watermarking algorithm is presented in [35]. This study proposes a watermark-embedding algorithm for e-government document images combining DCT, the singular value decomposition and the genetic algorithm-based optimization.

An example of DWT-based frequency domain embedding is described in paper [36]. A key feature of this study is the use of a differential evolution algorithm to find the best location for watermark blocks.

In some studies, authors use compositions of transforms. An example of such an approach is described in [37]. The authors use a hybrid embedding scheme that combines DCT and DWT. First, DCT is applied to the pixels of the cover image, then one level of Haar DWT is applied to obtain four

frequency sub-bands. Fragments of the digital watermark are embedded to these four sub-bands in an additive manner after the preliminary application of Arnold transform and DCT. At the extraction stage, the original cover image is needed to restore the watermark.

There are also algorithms based on other less common transforms. For example, in [38], the contourlet transform and the Fresnel transform are used. A QR code is used as a watermark, and the optimal frequency coefficients for embedding are selected using optimization algorithms. The authors of [39] use the Walsh–Hadamard transform to embed watermarks in digital images. The embedding algorithm is developed using a linear prediction function.

Thus, various authors have developed many different watermarking methods. In particular, there are many methods for hiding watermarks in digital images. Many of them are claimed to be quite effective. We have selected examples of different class methods to study their applicability within the proposed technology. The proposed technology itself and the results of the experiments are presented below.

3. The Proposed Technology

In this section, we propose an electronic documents authorship protection technology based on image watermarking. We discuss scenarios for using this technology and options for its implementation.

3.1. The Main Concept

Section 2 presented examples of various methods for digital watermark embedding in electronic documents. Each class of these methods has both advantages and disadvantages that limit their practical applicability.

We propose a technology for protecting electronic documents based on hiding digital watermarks in images accompanying the text of the document. Unlike the methods presented in Section 2, our technology does not imply any changes to the text or its formatting. Instead, invisible watermarks are embedded in graphics contained in an electronic document.

Before inserting images into the document created in a text editor, the content author must embed watermarks (the same or different) into all these images according to the chosen embedding algorithm. It should be especially noted that in order to ensure a high level of protection, robust watermarks that are not destroyed under various destructive effects on the image should be used. After placing watermarked images in the document, this document can be stored or transmitted in its original form or after conversion to another format, such as PDF. The general scheme of this process is shown in Figure 1.

The proposed protection technology allows us to implement an electronic document authorship protection without changing the content and structure of the text. A well-chosen watermarking algorithm ensures a high level of robustness. Therefore, the images copied along with the text will contain fragments of the watermark even after distorting influences such as compression, brightness change, cropping, scaling, etc.

Obviously, the proposed technology has a limitation associated with the fact that the protected electronic document must contain a graphic object. However, despite this limitation, this technology can be used to protect a large number of electronic documents, since many documents nowadays contain some kind of graphic element. Even corporate documents often contain a logo, which can be used as a cover image for a watermark. We also note that the problem of authorship protection for digital content is more typical for electronic documents related to creative and intellectual work. The results of such work are books, papers, blog posts and other publications, which are usually accompanied by diagrams and illustrations. Therefore, the target audience of the proposed technology is mainly authors of original content in different spheres.

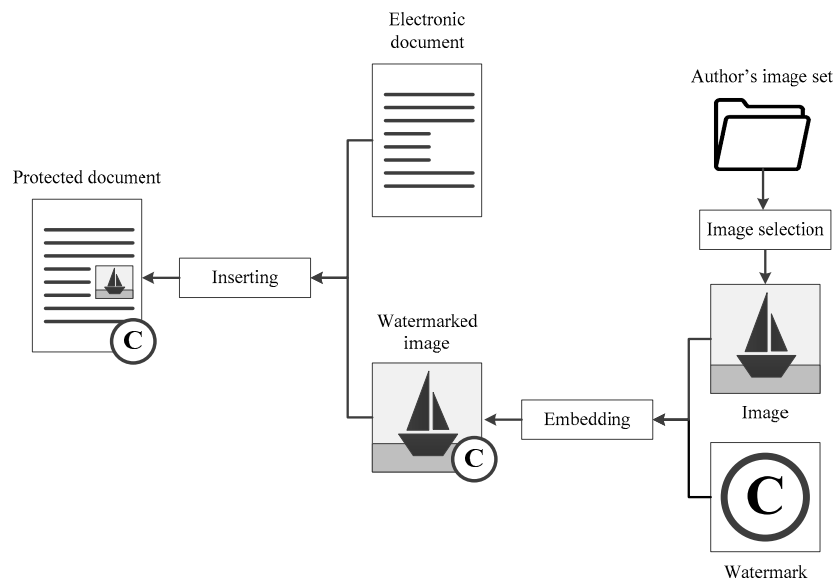


Figure 1. General scheme of document protection according to the proposed technology.

Unlike software that completely prohibits copying information, our technology allows to copy text and graphic objects from a document. We mean cases where an author publishes a document in the public domain and does not object to the use of its fragments if such use does not infringe copyright. Regardless of the purpose for copying a document, an embedded watermark can successfully identify the original author of the content.

The author of the content can use their identification data and brief information about the protected document as a digital watermark. For the most effective authorship protection of an electronic document, it is necessary to ensure the link between the image and its context in the document. In this case, to generate a watermark, the identification data of the author, the name of the document and a fragment of its text are required. The association of a digital watermark with the text of a document can be used to prove authorship even if the text of the protected document is copied without a graphic object.

Here is a formal description of the process of protecting an electronic document.

A model of an electronic document belonging to a certain user is a set D of the following form:

$$D = T \cup I = \{t_1, t_2, \dots, t_k\} \cup \{i_1, i_2, \dots, i_l\}, \quad (1)$$

where $T = \{t_1, t_2, \dots, t_k\}$ is a set of finite sequences of characters written in some finite alphabet; $I = \{i_1, i_2, \dots, i_l\}$ is a non-empty set of digital images contained in a document, which is a subset of the set of all kinds of finite-resolution images $I \subset J$.

Let us denote the set of digital watermarks associated with the owner of the document as:

$$W = \{w_1, w_2, \dots, w_r\}. \quad (2)$$

The set of watermarking algorithms available to the author of the document is denoted as:

$$A = \{a_1, a_2, \dots, a_r\}, a_j : I \times W \rightarrow \tilde{I}, j = \overline{1, r}, \quad (3)$$

where $I \neq \tilde{I} \subset J$.

In addition, we introduce two mappings.

$$\varphi : I \rightarrow A. \quad (4)$$

The mapping (4) matches each image contained in the text to a watermarking algorithm.

$$v: I \rightarrow W. \quad (5)$$

The mapping (5) assigns each image contained in the text to a digital watermark from the set W .

Then, the mathematical model of the electronic document protecting process can be described as the following set:

$$E = (D, W, A, \varphi, v). \quad (6)$$

The model (6) defines the case when the watermark is not based on the image context. If the watermark depends on the image context, then the mathematical model of the electronic document protecting process has the following form:

$$E = (D, W, B, \varphi, v), \quad (7)$$

where the set of digital watermarking algorithms available to the owner of the document is denoted as:

$$B = \{b_1, b_2, \dots, b_r\}, b_j: I \times W \times T \rightarrow \widetilde{I}, j = \overline{1, r}. \quad (8)$$

Let us describe the embedding options corresponding to the given models in more detail.

The case when the watermark is specified by the user and does not depend on the context of the image in the document is illustrated in Figure 2a. In this case, the author of digital content chooses some information that is used as a watermark. This can be text information about the author or a personal logo. This data can be used as a watermark without any changes or after applying additional transforms. The author embeds the selected watermark into the image, which is then placed in the document.

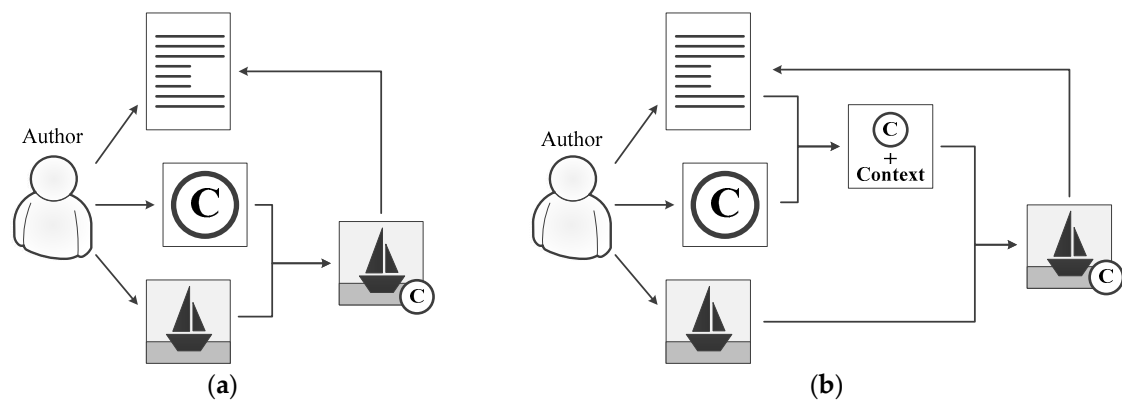


Figure 2. Embedding scheme: (a) The watermark is not based on the image context; (b) the watermark is based on the image context.

This scheme hardly differs from the classic case of embedding digital watermarks into digital images, especially if the resulting document is stored in its original format. The situation when the final text document is converted into any graphic format or into PDF format is of greater interest. In this case, the images contained in the document are distorted. Therefore, to protect authorship, it is necessary to use robust watermarks that are resistant to image compression.

The advantage of using this embedding scheme is convenience for the user since no additional steps are required, except for choosing a watermark for embedding. However, this scheme has a significant drawback. If a plagiarist copies the text of a document without images, then the link between the images and the document is lost. Subsequently, it will be difficult to prove authorship for the entire document, and not just for an individual image.

In another possible scheme, the watermark is generated based on user information as well as the context of the image. Image context is the text fragments or other objects surrounding the image in an electronic document or information about the location of the image (page number, paragraph, line, etc.) in the original document. In this case, the step of watermark embedding is preceded by the step of watermark generating. User information is combined with context information, for example, by concatenation. Hashing can be used to obtain a watermark of a fixed small size. The described embedding scheme is illustrated in Figure 2b.

This embedding scheme is more difficult for the user than the previous one because it is necessary to set the context for each specific image. However, this process can be successfully automated. At the same time, the use of the context significantly increases the security of the proposed technology. Even if the text of the document is copied without images, the author can easily prove the authorship by demonstrating the link between the digital watermark and the text. The use of context will help to prove the authorship in a situation when the text copied by a plagiarist was significantly processed. The watermark shows the link with the original document even after significant text changes. This variant of the proposed technology application is recommended for practical use.

It should be noted that the proposed technology can be used in combination with other technologies for protecting electronic documents if an increased level of security is required. The combination of this technology with text watermarking methods provides reliable protection of authorship for an electronic document at several levels at once.

The proposed technology can be applied in a variety of systems dealing with electronic documents. In the simplest case, the protected content is published on the Internet as a file for download or as an element of a web page after embedding a watermark on the user's device side. Another use case may be associated with electronic document management systems. In this case, an additional watermarking module is needed.

The use of the proposed technology in smart systems and augmented reality systems is of particular interest. In this case, the authorship of digital content can be confirmed even if the electronic document is printed. The following idea looks the most impressive. After scanning a hard copy of a document using a smartphone, the elements of graphics and text are recognized using appropriate computer vision algorithms. Then, a search for the correspondence of these elements to previously published materials is performed using standard search services. As a result of the coincidence of elements of the printed document and the previously published original, the system extracts the watermark from the original version of the document and displays it or the original document over the scanned copy.

The difficulty in implementation of such a system lies in the fact that for its correct functioning, authors of digital content must use the same watermarking algorithms. However, at present, there is no single watermarking infrastructure, and each application uses its own embedding algorithm.

However, there are much more realistic scenarios for the application of this technology. These scenarios are suitable for a separate corporate infrastructure which uses a single watermarking standard. In case the watermark is print-resistant, the smart application can recognize the image and immediately extract the watermark. The application then displays it on the user's screen over the scanned document. In the case that the watermark is less robust, the scanned document must be compared with the original document stored, for example, in the cloud. This idea is illustrated in Figure 3. After that, the watermark is extracted from the original document and displayed on the smartphone screen. Similarly, authorship confirmation can be organized when scanning an electronic document from a computer screen using a smartphone.

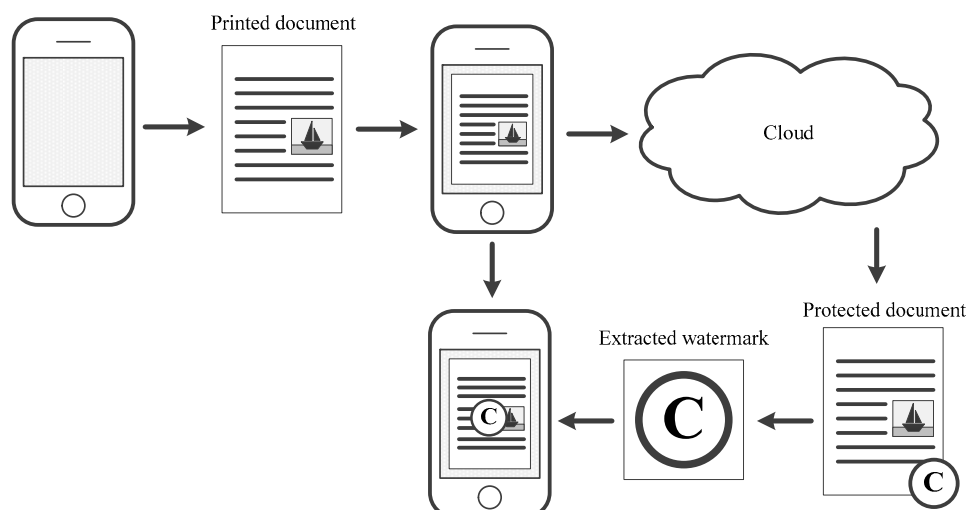


Figure 3. Authorship proof of a printed document using a smart application.

3.2. Application Scenarios

In practice, when using the proposed technology, three main scenarios can be carried out. The application scenario is the situation of copying an electronic document when authorship proof becomes relevant. The purpose of copying is not analyzed. In terms of the proposed technology, it does not matter who the plagiarist is and why he or she copies the information. The main scenarios include copying of an entire electronic document (full copying) as well as partial copying. Partial document copying can be divided into image copying and text copying. A detailed description of each scenario in terms of the effectiveness of the proposed technology is given below.

3.2.1. Full Copying

Full copying of a document means copying the entire content of an electronic document, including text and images. Most likely, the text and images are logically related and the meaning of the document is distorted in the absence of text or graphics. In terms of the proposed technology, it does not matter whether the plagiarist copies the document file or copies its content to another file.

Since watermarked images were copied along with the text of the document, you only need to extract the watermarks from them in order to prove authorship. If the author of the original content used a robust watermarking algorithm, authorship confirmation is successful even if certain typical image processing techniques were applied to the protected graphic objects. The text processing does not reduce the effectiveness of authorship verification as watermarks are embedded in images.

In this scenario, a high level of efficiency can be achieved by embedding both an independent watermark and a context-based watermark. However, using a context-based watermark provides an additional link between the image and the text. This option is preferable, especially if the document consists of several pages and images are not located on every page. The described scenario is illustrated in Figure 4.

3.2.2. Text-Only Copying

A plagiarist can copy only the text of a protected document without copying images. This scenario is most likely when images are mostly decorative and do not affect the quality of information perception. Since the text of an electronic document is not protected by a watermark, the authorship proof procedure requires the original version of an electronic document to contain watermarked graphic objects. If proof of authorship is required, the author of the original content must provide the original version of the electronic document and extract the watermarks from the images contained in the document. In this case, high efficiency of the proposed technology is ensured if the image context in the document is used

when generating the watermark. If a document contains several pages and some of these pages do not contain images, a plagiarist can copy only the latter. In this case, the watermark must be generated based on the entire document content, for example, using hashing. This ensures that the watermark is associated with the entire document, including text-only pages. The robustness of the watermark is optional. The described scenario is illustrated in Figure 5.

3.2.3. Images-Only Copying

The plagiarist can copy only graphic objects (photographs, illustrations, diagrams, etc.) into another document. They can also copy these graphic objects without placing them in another electronic document. This is the most likely scenario if the document consists mainly of images or if images duplicate the meaning of the text (for example, diagrams, infographics). This case is completely analogous to the classic scenario of embedding digital watermarks into images. The invisible watermark must be extracted from the image to prove authorship. If the author of the original content used a robust watermarking algorithm, authorship proof will be successfully implemented even after distorting the watermarked image. The effectiveness of the proposed technology is high regardless of whether the author uses an independent watermark or a context-based watermark. This scenario is illustrated in Figure 6.

All three scenarios show that the authorship proof procedure can be successfully implemented. Since any of these scenarios can potentially be carried out in practice, it is recommended to use robust watermarking algorithms and linking to the electronic document context to ensure a high level of authorship protection.

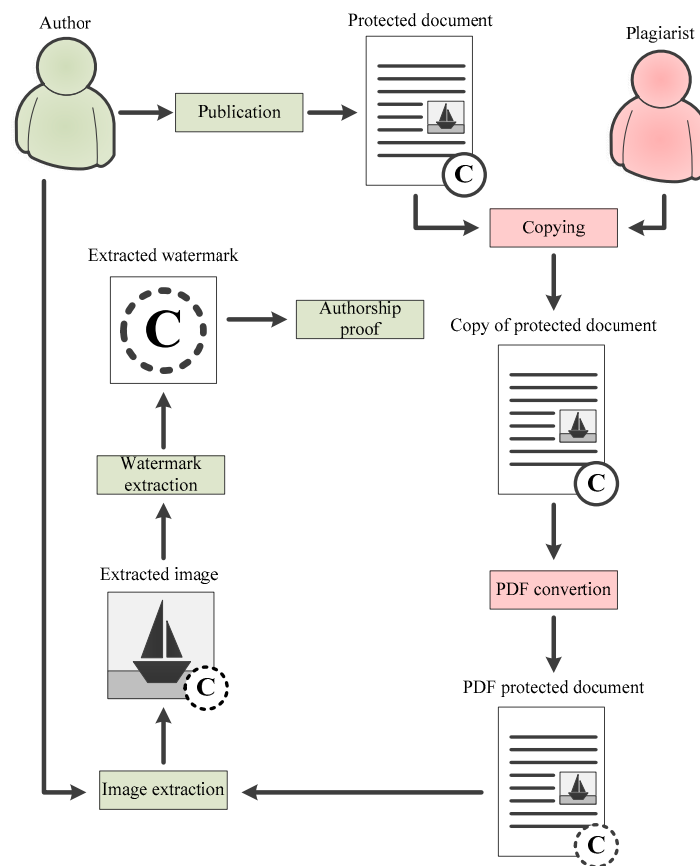


Figure 4. Authorship proof scenario when copying a full document.

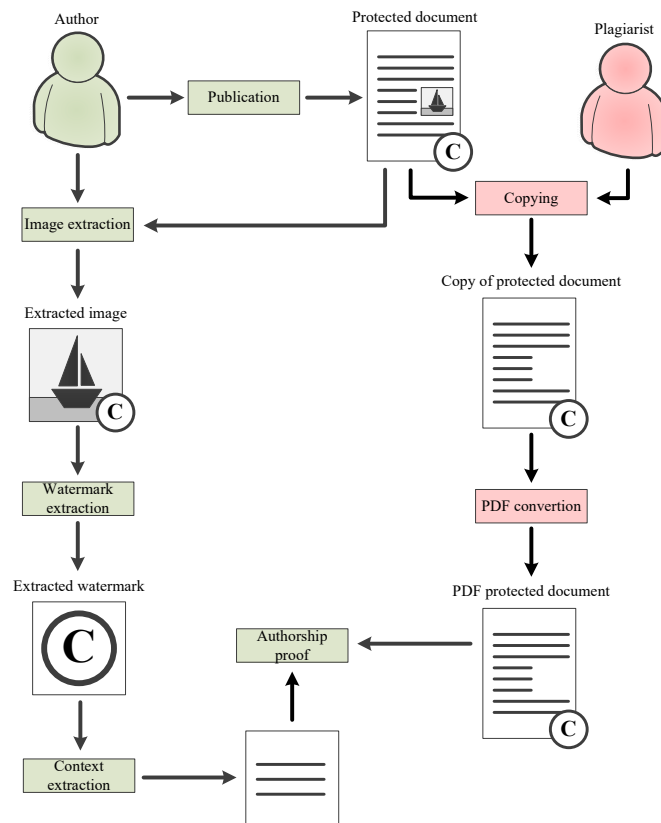


Figure 5. Authorship proof scenario when copying text only.

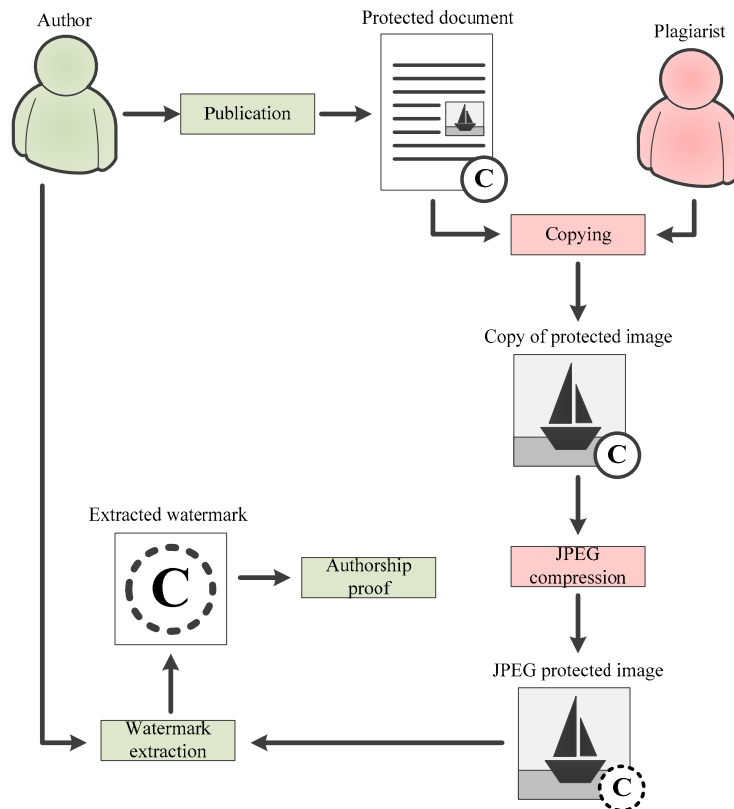


Figure 6. Authorship proof scenario when copying images only.

4. Experimental Results

To carry out the experiments, we created electronic documents containing text and halftone 512×512 images using Microsoft Word text editor. We used 10 images in our experiments. Half of the images were taken from the USC-SIPI database [40]. These are classic images such as “Airplane”, “Baboon”, “Lena”, “Goldhill” and “Peppers”. Five more images included are stock images from pixabay.com [41]. These are illustrations containing a large number of smooth areas. This set of images was chosen due to the fact that electronic documents can contain images of different types. These can be photographs containing many objects and textures, as well as various illustrations and schemes. The cover images are shown in Figure 7.

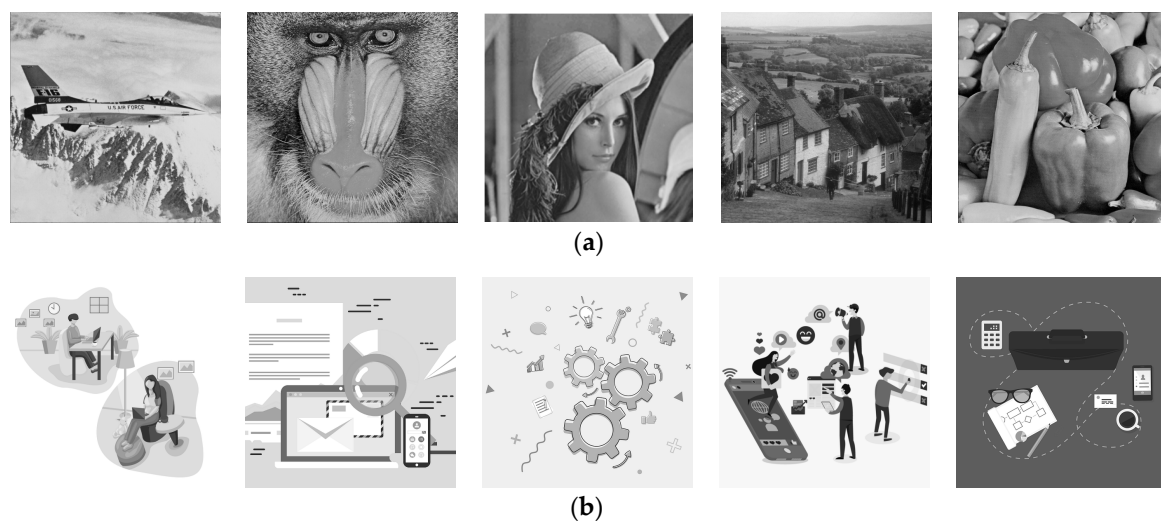


Figure 7. Cover images: (a) from the USC-SIPI database; (b) from pixabay.com.

A 64×64 binary watermark was embedded in each of the images. In these experiments, we did not use the context for watermark generation. The purpose of the experiment was to test the resistance of the watermarked images to document file conversion to PDF format. We chose converting to PDF because it is one of the most common conversions for electronic documents.

We used the algorithms described in papers [31,34,37] for embedding. These are state-of-the-art robust watermarking algorithms, characterized by increased resistance to JPEG compression, which is most commonly used when converting documents to PDF. The algorithm from [31] is an example of spatial domain embedding, and [34] and [37] are examples of frequency domain embedding. A more detailed description of these works is given in Section 2. A watermark and watermarked images are shown in Figure 8. The peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) [42,43] were used to assess the quality of watermarked images.

The resulting electronic documents were converted to PDF files with various graphics compression quality settings. Then, watermarked images were extracted from the PDF document, and watermarks were extracted from images.

To quantify the compression quality, we measured the difference between an original cover image and an extracted cover image after conversion to PDF using PSNR and SSIM metrics. We also measured the difference between an original cover image and an extracted watermarked image after conversion to PDF. The averaged results for all images are presented in Table 1.

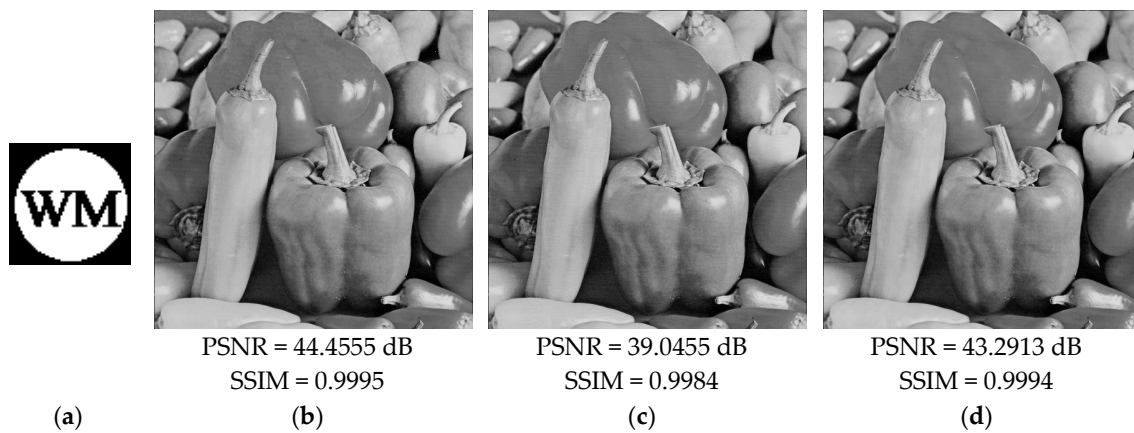


Figure 8. Watermark (a) and watermarked images obtained using (b) algorithm [31]; (c) algorithm [34]; (d) algorithm [37].

Table 1. Compressed images quality metrics.

Compression Quality	Compressed Images Quality			
	Cover Image	Watermarked [31]	Watermarked [34]	Watermarked [37]
No	PSNR = INF SSIM = 1	PSNR = 46.3677 dB SSIM = 0.9997	PSNR = 38.4450 dB SSIM = 0.9980	PSNR = 43.2915 dB SSIM = 0.9993
Maximum	PSNR = 42.9381 dB SSIM = 0.9993	PSNR = 41.2843 dB SSIM = 0.9989	PSNR = 36.0020 dB SSIM = 0.9964	PSNR = 40.0725 dB SSIM = 0.9984
High	PSNR = 38.4946 dB SSIM = 0.9977	PSNR = 37.7989 dB SSIM = 0.9973	PSNR = 34.6790 dB SSIM = 0.9949	PSNR = 37.1789 dB SSIM = 0.9970
Medium	PSNR = 36.1193 dB SSIM = 0.9957	PSNR = 35.6825 dB SSIM = 0.9953	PSNR = 33.3799 dB SSIM = 0.9927	PSNR = 35.2881 dB SSIM = 0.9950
Low	PSNR = 34.2754 dB SSIM = 0.9931	PSNR = 33.9960 dB SSIM = 0.9929	PSNR = 32.3243 dB SSIM = 0.9904	PSNR = 33.7073 dB SSIM = 0.9925
Minimum	PSNR = 31.0073 dB SSIM = 0.9850	PSNR = 30.8568 dB SSIM = 0.9847	PSNR = 29.4452 dB SSIM = 0.9806	PSNR = 30.7325 dB SSIM = 0.9843

The bit error rate (BER) and normalized cross-correlation (NCC) were used to assess the quality of the extracted watermarks. BER can be expressed by the formula:

$$\text{BER} = \frac{B_e}{B} \quad (9)$$

where B is the number of watermark bits; B_e is the number of errors (changed bits) that occurred during extraction.


























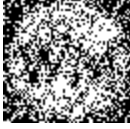




NCC between the original watermark W and the extracted watermark W_{ext} is computed using the formula:

$$\text{NCC} = \frac{\sum_{x=1}^M \sum_{y=1}^N (W(x,y) \times W_{ext}(x,y))}{\sqrt{\sum_{x=1}^M \sum_{y=1}^N (W^2(x,y))} \sqrt{\sum_{x=1}^M \sum_{y=1}^N (W_{ext}^2(x,y))}} \quad (10)$$

where $M \times N$ is the size of an original watermark.

Tables 2 and 3 show the results of watermark extraction from images using the spatial domain embedding algorithm [31]. The tables show the compression quality, BER and NCC values and demonstrate the extracted watermark.

Table 2. Watermarks extracted from highly detailed watermarked images obtained by the algorithm [31].

Compression Quality	Airplane	Baboon	Lena	Goldhill	Peppers
No	 BER = 0 NCC = 1	 BER = 0 NCC = 1	 BER = 0 NCC = 1	 BER = 0 NCC = 1	 BER = 0 NCC = 1
Maximum	 BER = 0 NCC = 1	 BER = 0 NCC = 1	 BER = 0 NCC = 1	 BER = 0 NCC = 1	 BER = 0 NCC = 1
High	 BER = 0.0002 NCC = 0.9998	 BER = 0.0005 NCC = 0.9996	 BER = 0.0007 NCC = 0.9993	 BER = 0.0005 NCC = 0.9996	 BER = 0.0007 NCC = 0.9993
Medium	 BER = 0.0173 NCC = 0.9841	 BER = 0.0439 NCC = 0.6596	 BER = 0.0225 NCC = 0.9794	 BER = 0.0231 NCC = 0.9787	 BER = 0.0232 NCC = 0.9787
Low	 BER = 0.1104 NCC = 0.8970	 BER = 0.1543 NCC = 0.8580	 BER = 0.1172 NCC = 0.8916	 BER = 0.1228 NCC = 0.8861	 BER = 0.1129 NCC = 0.8952
Minimum	 BER = 0.3018 NCC = 0.7177	 BER = 0.4116 NCC = 0.6078	 BER = 0.3059 NCC = 0.7227	 BER = 0.3303 NCC = 0.6832	 BER = 0.3105 NCC = 0.7142

The experimental results confirm the high resistance of the algorithm from [31] to image compression. The watermark was extracted in its original form or with minor distortion after maximum and high-quality compression. Even after minimum quality compression, fragments of the watermark can be found in the protected image. It should be mentioned that watermark embedding in poorly detailed images (synthesized images) shows higher resistance to compression than watermark embedding in highly detailed images. High robustness of embedding makes it possible to effectively protect the authorship of electronic documents containing images.

Table 3. Watermarks extracted from poorly detailed synthesized watermarked images obtained by the algorithm in [31].































Compression Quality	Work1	Email	Gears	Social media	Work2
No					
	BER = 0.0054 NCC = 0.9951	BER = 0.0081 NCC = 0.9926	BER = 0.0034 NCC = 0.9968	BER = 0 NCC = 1	BER = 0 NCC = 1
Maximum					
	BER = 0.0068 NCC = 0.9937	BER = 0.0088 NCC = 0.9918	BER = 0.0034 NCC = 0.9968	BER = 0 NCC = 1	BER = 0 NCC = 1
High					
	BER = 0.0115 NCC = 0.9895	BER = 0.0102 NCC = 0.9905	BER = 0.0042 NCC = 0.9962	BER = 0.0002 NCC = 0.9997	BER = 0.0002 NCC = 0.9997
Medium					
	BER = 0.0187 NCC = 0.9828	BER = 0.0163 NCC = 0.9849	BER = 0.0139 NCC = 0.9872	BER = 0.0019 NCC = 0.9982	BER = 0.0022 NCC = 0.9979
Low					
	BER = 0.1359 NCC = 0.8751	BER = 0.1519 NCC = 0.8645	BER = 0.1313 NCC = 0.8792	BER = 0.1021 NCC = 0.9045	BER = 0.1174 NCC = 0.8957
Minimum					
	BER = 0.2851 NCC = 0.7354	BER = 0.1953 NCC = 0.8161	BER = 0.2456 NCC = 0.8028	BER = 0.3161 NCC = 0.7855	BER = 0.1716 NCC = 0.8409

Table 4 shows the results of the experiment with the algorithm in [34]. Unlike the previous experiment, in this case, the watermark is embedded in the DCT frequency domain. The algorithm is based on changing the differences in the DCT coefficients, and this makes it unsuitable for poorly detailed images because the difference between the coefficients of adjacent blocks is often zero. Therefore, experiments on embedding the watermark in synthesized images were not carried out.

Experiments have shown that the watermark is completely destroyed after minimum quality compression. However, when compressed at a higher quality, the watermark can be easily detected in the image. Therefore, this algorithm also provides effective protection for electronic documents.

Table 4. Watermarks extracted from highly detailed watermarked images obtained by the algorithm in [34].





























































Compression Quality	Airplane	Baboon	Lena	Goldhill	Peppers
No	 BER = 0 NCC = 1	 BER = 0 NCC = 1	 BER = 0 NCC = 1	 BER = 0 NCC = 1	 BER = 0 NCC = 1
Maximum	 BER = 0 NCC = 1	 BER = 0 NCC = 1	 BER = 0 NCC = 1	 BER = 0 NCC = 1	 BER = 0 NCC = 1
High	 BER = 0.0044 NCC = 0.9959	 BER = 0.0032 NCC = 0.9942	 BER = 0.0024 NCC = 0.9978	 BER = 0.0068 NCC = 0.9937	 BER = 0.0046 NCC = 0.9957
Medium	 BER = 0.1243 NCC = 0.8842	 BER = 0.1445 NCC = 0.8661	 BER = 0.1301 NCC = 0.8796	 BER = 0.1406 NCC = 0.8686	 BER = 0.1421 NCC = 0.8677
Low	 BER = 0.3129 NCC = 0.6976	 BER = 0.2605 NCC = 0.7552	 BER = 0.3289 NCC = 0.6881	 BER = 0.3113 NCC = 0.7036	 BER = 0.3142 NCC = 0.6977
Minimum	 BER = 0.5398 NCC = 0.4235	 BER = 0.5259 NCC = 0.4835	 BER = 0.5325 NCC = 0.4269	 BER = 0.5376 NCC = 0.4623	 BER = 0.5374 NCC = 0.4322

Table 5 shows the results of the experiment using the algorithm in [37]. This hybrid algorithm is based on a combination of DCT and DWT. It differs from the algorithms from [31] and [34] by non-blind watermark extraction. This means that the cover image is required for watermark extraction. This is not always convenient, which is a disadvantage. However, experimental results show improved compression resistance. Even with minimum compression quality, the watermark can be easily detected in the image. Therefore, this algorithm can also be used for effective protection of electronic documents containing images. It is worth noting that when watermarks were embedded in synthesized images, noticeable distortions of the extracted watermark occurred even at maximum compression quality. According to this, we can conclude that the algorithm in [37] is suitable only for highly detailed images, as well as the algorithm in [34].

Table 5. Watermarks extracted from highly detailed watermarked images obtained by the algorithm in [37].

Compression Quality	Airplane	Baboon	Lena	Goldhill	Peppers
No	 BER = 0 NCC = 1	 BER = 0 NCC = 1	 BER = 0 NCC = 1	 BER = 0 NCC = 1	 BER = 0 NCC = 1
Maximum	 BER = 0.0012 NCC = 0.9989	 BER = 0.0005 NCC = 0.9996	 BER = 0.0015 NCC = 0.9987	 BER = 0.0009 NCC = 0.9991	 BER = 0.0012 NCC = 0.9989
High	 BER = 0.0261 NCC = 0.9760	 BER = 0.0209 NCC = 0.9807	 BER = 0.0183 NCC = 0.9832	 BER = 0.0186 NCC = 0.9830	 BER = 0.0227 NCC = 0.9791
Medium	 BER = 0.0537 NCC = 0.9504	 BER = 0.0601 NCC = 0.9445	 BER = 0.0601 NCC = 0.9446	 BER = 0.0674 NCC = 0.9381	 BER = 0.0588 NCC = 0.9456
Low	 BER = 0.1047 NCC = 0.9027	 BER = 0.1111 NCC = 0.8967	 BER = 0.1000 NCC = 0.9069	 BER = 0.1089 NCC = 0.8993	 BER = 0.1089 NCC = 0.8993
Minimum	 BER = 0.2522 NCC = 0.7652	 BER = 0.2488 NCC = 0.7672	 BER = 0.2446 NCC = 0.7686	 BER = 0.2515 NCC = 0.7627	 BER = 0.2456 NCC = 0.7688

5. Discussion

In this section, we discuss the results of the experiments and draw conclusions on the practical value of the proposed protection technology. We also present a comparison with the state of the art and discuss the advantages of our solution.

5.1. Discussion of Experimental Results

The experiments presented in Section 4 were carried out to confirm the practical value of the proposed technology for protecting electronic documents with the example of converting to PDF. Another purpose was to evaluate the effectiveness of various embedding algorithms that can be used to implement this technology.

Digital watermarking algorithms can be divided into two large classes based on the embedding domain. Spatial domain algorithms work with image pixels. Frequency domain algorithms work

with the spectrum of an image obtained using some frequency transform. The choice of frequency transform has a significant impact on the efficiency of embedding. Frequency domain algorithms most often use DFT, DCT and DWT for embedding digital watermarks. There are also algorithms based on the joint use of several transforms.

For the experiments, we selected algorithms that represent different classes. We chose a spatial domain embedding algorithm from [31] and frequency domain embedding algorithms from [34] and [37]. The algorithm in [34] is based on DCT. The algorithm in [37] is based on DCT and DWT and is distinguished by non-blind watermark extraction. The choice of these algorithms from the set of possible ones was determined by the high efficiency of embedding, which was declared by the authors of the relevant studies. We did not use DFT-based algorithms to conduct experiments. Such algorithms provide good resistance to geometric attacks. This is due to the properties of the DFT. However, JPEG compression does not belong to this class of attacks. Therefore, DFT-based algorithms are less useful for our technology than other frequency algorithms.

The experiments have shown that none of the algorithms have an absolute advantage over the rest of the algorithms. A comparison of these algorithms in terms of their resistance to converting documents to PDF is shown in Table 6. To quantify distortions, we used average BER and NCC values (BER_{avg} and NCC_{avg} , respectively).

Table 6. Comparison of compression resistance of the selected algorithms.

Algorithm	Highly Detailed Images	Synthesized Images
[31]	<ul style="list-style-type: none"> Noticeable distortions of the digital watermark ($BER_{avg} = 0.1235$, $NCC_{avg} = 0.8856$) appear when the quality of the compressed images is low. The watermark is almost destroyed ($BER_{avg} = 0.3320$, $NCC_{avg} = 0.6891$) when the quality of the compressed images is minimal. The pattern of digital watermark distortion is the same for different images. 	<ul style="list-style-type: none"> Noticeable distortions of the digital watermark ($BER_{avg} = 0.1277$, $NCC_{avg} = 0.8838$) appear when the quality of the compressed images is low. The watermark is significantly distorted but not destroyed ($BER_{avg} = 0.2427$, $NCC_{avg} = 0.7961$) when the quality of the compressed images is minimal. The pattern of digital watermark distortion is noticeably different for different images.
[34]	<ul style="list-style-type: none"> Noticeable distortions of the digital watermark ($BER_{avg} = 0.1363$, $NCC_{avg} = 0.8732$) appear when the quality of the compressed images is medium. The watermark is almost destroyed ($BER_{avg} = 0.3056$, $NCC_{avg} = 0.7084$) when the quality of the compressed images is low. The watermark is completely destroyed ($BER_{avg} = 0.5346$, $NCC_{avg} = 0.4457^*$) when the quality of the compressed images is minimal. The pattern of digital watermark distortion is the same for different images. 	Embedding was not carried out.
[37]	<ul style="list-style-type: none"> Noticeable distortions of the digital watermark ($BER_{avg} = 0.1067$, $NCC_{avg} = 0.9010$) appear when the quality of the compressed images is low. The watermark is significantly distorted but not destroyed ($BER_{avg} = 0.2485$, $NCC_{avg} = 0.7665$) when the quality of the compressed images is minimal. The pattern of digital watermark distortion is the same for different images. 	Embedding was not carried out.

The advantage of the spatial domain embedding algorithm is that it is applicable to both highly detailed and poorly detailed synthesized images. Moreover, in the case of embedding of digital watermarks in synthesized images, this algorithm shows very high resistance to compression.

Frequency domain algorithms are hardly applicable to synthesized images. We can conclude that embedding in such images should be carried out using spatial domain algorithms. Special frequency algorithms can also be developed for this purpose.

When embedding digital watermarks into highly detailed images, the best result was shown by the non-blind algorithm based on a combination of DCT and DWT. This advantage is due to the use of the original image when extracting the digital watermark.

The experimental results lead to the conclusion that the use of algorithms belonging to one class within the proposed technology for protecting electronic documents is not advisable. A good solution is to use them together, depending on the characteristics of a particular electronic document.

Therefore, we propose the following scenario for the implementation of our technology:

1. To embed digital watermarks into images in electronic documents, a pool of algorithms should be formed that covers various groups of images.
2. The analysis and classification of the images contained in the document should be performed before the embedding procedure. The level of image detail should be used as a classification criterion.
3. Digital watermarks can be generated with or without the context of the document.
4. The generated digital watermarks should be embedded into images of the document. The choice of the embedding algorithm for each image will be performed based on the image class.

5.2. Comparison of the State of the Art

In this subsection, we compare the proposed technology with known solutions for electronic documents authorship protection using watermarking techniques. We selected five methods of different classes described in Section 2 for comparison. The study [14] belongs to the class of linguistic methods of text watermarking, and [18] is an example of the structural method of text watermarking. The study [21] presents a PDF document watermarking method that modifies the coordinates of symbols. In [24], a watermarking scheme is proposed for protecting text images, for example, document scans. The paper [26] describes an approach to protecting PDF document pages using image watermarking algorithms. Table 7 demonstrates the advantages and disadvantages of methods belonging to different classes using the examples of selected studies.

Table 7. Comparison of the state of the art.

Ref No.	Document Type	Embedding Method	Authorship Protection		
			Text Copying	Images Copying	Format Changing
[14]	Any type	Making lexical or syntactic changes in texts	Yes	No	Any format changing
[18]	RTF, DOC, DOCX, PDF	Changing the value of line spacing	No	No	Printing, scanning, conversion to image
[21]	PDF	Character coordinate modification	No	No	Printing, scanning, conversion to image
[24]	Text image	Changing the frequency coefficients of images	No (separate copying of text is not possible)	Yes	Conversion to JPEG format
[26]	PDF	Robust image watermarking (any algorithm)	No (separate copying of text is not possible)	Yes	Printing, scanning, conversion to image
Proposed	Any document containing an image	Robust image watermarking (any algorithm)	Yes, if the watermark is created using context (for example, hashing)	Yes	Any format changing

The advantage of linguistic methods of text watermarking, such as in [14], is the protection of the text from being copied into another document. Another advantage is the preservation of the embedded watermark when converting a document into other electronic document formats, copying and scanning. However, the images contained in the document are not protected.

Structural methods, such as in [18], protect the authorship of a specific document file as well as its copy obtained by printing, scanning or converting to an image. However, copying text to any other document destroys the watermark completely. The method in [21] designed to protect PDF documents has similar advantages and disadvantages.

Text image watermarking methods [24] protect a specific copy of a scanned document or a document converted to an image. The advantage of such methods is image authorship protection since the image is a part of the protected page. The image contains a watermark fragment after copying. However, these methods do not allow to copy the text, which can be inconvenient in some cases. When recognizing text using special programs, the watermark is not preserved. The method proposed in the paper [26] is designed to protect PDF documents, but it works in a similar way; therefore, it has similar features.

The proposed technology can be used to protect the authorship of electronic documents in any format. Since the watermark is embedded in images, the authorship for images is successfully protected. If the watermark is created using the content of the document, for example, by hashing, then it can be also used to confirm the authorship for the text of the document. A robust watermarking algorithm allows to preserve the watermark when converting a document to another format, for example, when converting from DOC to PDF and vice versa, when saving the document as an image and even when printing and scanning.

6. Conclusions

The paper presents a new technology for protecting the authorship of electronic documents. To ensure authorship protection, it is proposed to embed invisible watermarks in the images contained in the document. The key feature of the proposed technology is that the content and structure of the document do not change in the watermarking process. Our technology is useful for different classes of documents containing images. It can also be used for various tasks in practice: for document publication on the web; for electronic document management systems; or for applications based on immersive technologies.

Our research demonstrates all possible scenarios for the implementation of the proposed protection technology: full copying of an electronic document, text-only copying, and images-only copying. The proposed technology effectively provides proof of authorship both when copying an entire document and when copying parts of it.

To ensure a high level of protection, it is recommended to use robust watermarking algorithms. We have conducted a study of the applicability of the proposed technology using some robust watermarking algorithms. At the same time, we have carried out experiments with both photographic images and synthesized graphics (illustrations). The experimental results demonstrate that the proposed technology can be successfully applied in practice. It is advisable to use a combination of spatial and frequency domain embedding algorithms, depending on the level of image detail.

Author Contributions: Conceptualization, methodology, writing—original draft preparation and review and editing by O.E.; methodology, software, validation, investigation, writing—original draft preparation and review and editing by A.M.; funding acquisition, formal analysis, writing—original draft preparation by A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Ministry of Science and Higher Education of Russia, Government Order for 2020–2022, project no. FEWM-2020-0037 (TUSUR).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ryu, H.-S.; Park, H. A system for supporting paper-based augmented reality. *Multimedia Tools Appl.* **2015**, *75*, 3375–3390. [[CrossRef](#)]
2. Raso, R.; Cucerca, S.; Werth, D.; Loos, P. Automated Augmented Reality Content Creation for Print Media. In *Information Systems and Management in Media and Entertainment Industries*; Springer Science and Business Media LLC: Berlin, Germany, 2016; pp. 245–261.
3. Hussain, M.; Wahab, A.W.A.; Bin Idris, Y.I.; Ho, A.T.; Jung, K.-H. Image steganography in spatial domain: A survey. *Signal Process. Image Commun.* **2018**, *65*, 46–66. [[CrossRef](#)]
4. Kadhim, I.J.; Premaratne, P.; Vial, P.J.; Halloran, B. Comprehensive survey of image steganography: Techniques, Evaluations, and trends in future research. *Neurocomputing* **2019**, *335*, 299–326. [[CrossRef](#)]
5. McAteer, I.; Ibrahim, A.; Zheng, G.; Yang, W.; Valli, C. Integration of Biometrics and Steganography: A Comprehensive Review. *Technologies* **2019**, *7*, 34. [[CrossRef](#)]
6. Qasim, A.F.; Meziane, F.; Aspin, R. Digital watermarking: Applicability for developing trust in medical imaging workflows state of the art review. *Comput. Sci. Rev.* **2018**, *27*, 45–60. [[CrossRef](#)]
7. Kumar, S.; Singh, B.K.; Yadav, M. A Recent Survey on Multimedia and Database Watermarking. *Multimedia Tools Appl.* **2020**, *79*, 20149–20197. [[CrossRef](#)]
8. Yu, X.; Wang, C.; Zhou, X. Review on Semi-Fragile Watermarking Algorithms for Content Authentication of Digital Images. *Futur. Internet* **2017**, *9*, 56. [[CrossRef](#)]
9. Lin, W.-H.; Horng, S.-J.; Kao, T.-W.; Chen, R.-J.; Chen, Y.-H.; Lee, C.-L.; Terano, T. Image copyright protection with forward error correction. *Expert Syst. Appl.* **2009**, *36*, 11888–11894. [[CrossRef](#)]
10. Rawat, S.; Raman, B. A publicly verifiable lossless watermarking scheme for copyright protection and ownership assertion. *AEU Int. J. Electron. Commun.* **2012**, *66*, 955–962. [[CrossRef](#)]
11. Ahvanooy, M.T.; Li, Q.; Shim, H.J.; Huang, Y. A Comparative Analysis of Information Hiding Techniques for Copyright Protection of Text Documents. *Secur. Commun. Networks* **2018**, *2018*, 1–22. [[CrossRef](#)]
12. Ahvanooy, M.T.; Li, Q.; Hou, J.; Rajput, A.R.; Yini, C. Modern Text Hiding, Text Steganalysis, and Applications: A Comparative Analysis. *Entropy* **2019**, *21*, 355. [[CrossRef](#)] [[PubMed](#)]
13. Halvani, O.; Steinebach, M.; Wolf, P.; Zimmermann, R. Natural language watermarking for german texts. In Proceedings of the First ACM Workshop on Information Hiding and Multimedia Security—IH&MMSec’13, Montpellier, France, 17–19 June 2013; Association for Computing Machinery (ACM): New York, NY, USA, 2013; pp. 193–202.
14. Mali, M.L.; Patil, N.N.; Patil, J.B. Implementation of Text Watermarking Technique Using Natural Language Watermarks. In Proceedings of the 2013 International Conference on Communication Systems and Network Technologies, Gwalior, India, 6–8 April 2013; Institute of Electrical and Electronics Engineers (IEEE): Washington, DC, USA, 2013; pp. 482–486.
15. Lozhnikov, P.S.; Sulavko, A.E.; Eremenko, A.V.; Volkov, D. Method of protecting paper and electronic text documents through a hidden biometric identifier based on a signature. In Proceedings of the 2016 Dynamics of Systems, Mechanisms and Machines (Dynamics), Omsk, Russia, 15–17 November 2016; Institute of Electrical and Electronics Engineers (IEEE): Washington, DC, USA, 2016; pp. 1–5.
16. Taha, A.; Hammad, A.S.; Selim, M.M. A high capacity algorithm for information hiding in Arabic text. *J. King Saud Univ. Comput. Inf. Sci.* **2020**, *32*, 658–665. [[CrossRef](#)]
17. Rizzo, S.G.; Bertini, F.; Montesi, D. Fine-grain watermarking for intellectual property protection. *EURASIP J. Inf. Secur.* **2019**, *2019*, 10. [[CrossRef](#)]
18. Kozachok, A.V.; Kopylov, S.A.; Shelupanov, A.A.; Evsutin, O.O. Text marking approach for data leakage prevention. *J. Comput. Virol. Hacking Tech.* **2019**, *15*, 219–232. [[CrossRef](#)]
19. Kuribayashi, M.; Wong, K. Improved DM-QIM Watermarking Scheme for PDF Document. In Proceedings of the 18th International Workshop on Digital Watermarking—IWDW 2019, Chengdu, China, 2–4 November 2019; Springer: Cham, Switzerland, 2019; pp. 171–183.
20. Bitar, A.W.; Darazi, R.; Couchot, J.-F.; Couturier, R. Blind digital watermarking in PDF documents using Spread Transform Dither Modulation. *Multimedia Tools Appl.* **2015**, *76*, 143–161. [[CrossRef](#)]
21. Hatoum, M. Digital Watermarking for PDF Documents and Images: Security, Robustness and AI-based attack. Ph.D. Thesis, University of Bourgogne Franche-Comte, Besancon, France, 23 September 2020.

22. Liu, Y.; Limsiroratana, S.; Choksuriwong, A. Data hiding in text document images by cloud model. In Proceedings of the 4th International Congress on Image and Signal Processing, Shanghai, China, 15–17 October 2011; Institute of Electrical and Electronics Engineers (IEEE): Washington, DC, USA, 2011; Volume 2, pp. 1025–1029.
23. Laouamer, L.; Tayan, O. Performance Evaluation of a Document Image Watermarking Approach with Enhanced Tamper Localization and Recovery. *IEEE Access* **2018**, *6*, 26144–26166. [[CrossRef](#)]
24. Alotaibi, R.A.; Elrefaei, L. Text-image watermarking based on integer wavelet transform (IWT) and discrete cosine transform (DCT). *Appl. Comput. Inform.* **2019**, *15*, 191–202. [[CrossRef](#)]
25. Tan, L.; Hu, K.; Zhou, X.; Chen, R.; Jiang, W. Print-scan invariant text image watermarking for hardcopy document authentication. *Multimedia Tools Appl.* **2018**, *78*, 13189–13211. [[CrossRef](#)]
26. Mehta, S.; Prabhakaran, B.; Nallusamy, R.; Newton, D. mPDF: Framework for Watermarking PDF Files using Image Watermarking Algorithms. *arXiv* **2016**, arXiv:1610.02443. preprint.
27. Uljarević, D.; Veinović, M.; Kunjadić, G.; Tepšić, D. A new way of covert communication by steganography via JPEG images within a Microsoft Word document. *Multimedia Syst.* **2017**, *23*, 333–341. [[CrossRef](#)]
28. Rashid, A.; Coustaty, M.; Prasath, V.B.S. An Algorithm for Data Hiding in Radiographic Images and ePHI/R Application. *Technologies* **2018**, *6*, 7. [[CrossRef](#)]
29. Wazirali, R.; Alasmary, W.; Mahmoud, M.M.E.A.; Alhindi, A. An Optimized Steganography Hiding Capacity and Imperceptibly Using Genetic Algorithms. *IEEE Access* **2019**, *7*, 133496–133508. [[CrossRef](#)]
30. Di, F.; Zhang, M.; Huang, F.; Liu, J.; Kong, Y. Reversible data hiding in JPEG images based on zero coefficients and distortion cost function. *Multimedia Tools Appl.* **2019**, *78*, 34541–34561. [[CrossRef](#)]
31. Su, Q.; Liu, D.; Yuan, Z.; Wang, G.; Zhang, X.; Chen, B.; Yao, T. New Rapid and Robust Color Image Watermarking Technique in Spatial Domain. *IEEE Access* **2019**, *7*, 30398–30409. [[CrossRef](#)]
32. Poljicak, A.; Mandic, L.; Agic, D. Discrete Fourier transform-based watermarking method with an optimal implementation radius. *J. Electron. Imaging* **2011**, *20*, 033008. [[CrossRef](#)]
33. Cedillo-Hernandez, M.; Garcia-Ugalde, F.; Nakano-Miyatake, M.; Perez-Meana, H. Robust watermarking method in DFT domain for effective management of medical imaging. *Signal Image Video Process.* **2013**, *9*, 1163–1178. [[CrossRef](#)]
34. Ko, H.-J.; Huang, C.-T.; Horng, G.; Wang, S.-J. Robust and blind image watermarking in DCT domain using inter-block coefficient correlation. *Inf. Sci.* **2020**, *517*, 128–147. [[CrossRef](#)]
35. Horng, S.-J.; Rosiyadi, D.; Li, T.; Takao, T.; Guo, M.; Khan, M.K. A blind image copyright protection scheme for e-government. *J. Vis. Commun. Image Represent.* **2013**, *24*, 1099–1105. [[CrossRef](#)]
36. Salimi, L.; Haghighi, A.; Fathi, A. A novel watermarking method based on differential evolutionary algorithm and wavelet transform. *Multimedia Tools Appl.* **2020**, *79*, 11357–11374. [[CrossRef](#)]
37. Abdulrahman, A.K.; Ozturk, S. A novel hybrid DCT and DWT based robust watermarking algorithm for color images. *Multimedia Tools Appl.* **2019**, *78*, 17027–17049. [[CrossRef](#)]
38. Maheswari, S.U.; Hemanth, D.J. Performance enhanced image steganography systems using transforms and optimization techniques. *Multimedia Tools Appl.* **2015**, *76*, 415–436. [[CrossRef](#)]
39. Pakdaman, Z.; Saryazdi, S.; Nezamabadi-Pour, H. A prediction based reversible image watermarking in Hadamard domain. *Multimedia Tools Appl.* **2016**, *76*, 8517–8545. [[CrossRef](#)]
40. SIPI Image Database. Available online: <http://sipi.usc.edu/database/> (accessed on 5 September 2020).
41. Pixabay. Available online: <https://pixabay.com/ru/> (accessed on 5 September 2020).
42. Ponomarenko, N.; Silvestri, F.; Egiazarian, K.; Carli, M.; Astola, J.; Lukin, V. On between-coefficient contrast masking of DCT basis functions. In Proceedings of the Third International Workshop on Video Processing and Quality Metrics, Scottsdale, AR, USA, 25–26 January 2007; Volume 4.
43. Kwan, C.; Larkin, J.; Budavari, B.; Chou, B.; Shang, E.; Tran, T. A Comparison of Compression Codecs for Maritime and Sonar Images in Bandwidth Constrained Applications. *Computers* **2019**, *8*, 32. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).