IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 February 2024, accepted 29 February 2024, date of publication 5 March 2024, date of current version 13 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3374201

SURVEY

Digital Watermarking–A Meta-Survey and Techniques for Fake News Detection

AGNIESZKA MALANOWSKA[1], WOJCIECH MAZURCZYK[1], (Senior Member, IEEE), TANYA KOOHPAYEH ARAGHI[2], DAVID MEGÍAS[2], (Member, IEEE), AND MINORU KURIBAYASHI[3], (Senior Member, IEEE)

[1]Institute of Computer Science, Warsaw University of Technology, 00-665 Warsaw, Poland
[2]Internet Interdisciplinary Institute (IN3), Center for Cybersecurity Research of Catalonia (CYBERCAT), Universitat Oberta de Catalunya, 08018 Barcelona, Spain
[3]Center for Data-Driven Science and Artificial Intelligence, Tohoku University, Sendai 980-8576, Japan

Corresponding author: Agnieszka Malanowska (agnieszka.malanowska@pw.edu.pl)

This work was supported in part by the European Interest Group (EIG) Concert Japan Call to the Project Detection of fake newS on SocIal MedIa pLAtfoRms (DISSIMILAR) through Spanish Government under Grant PCI2020-120689-2; in part by the National Centre for Research and Development, Poland, under Grant EIG CONCERT-JAPAN/05/2021; and in part by Japan Science and Technology Agency, Japan, under Grant JPMJSC20C3. The work of Tanya Koohpayeh Araghi and David Megías was supported by the ''SECURING'' Project from Ministerio de Ciencia e Innovación, la Agencia Estatal de Investigación, and the European Regional Development Fund (ERDF) under Grant PID2021-125962OB-C31; and in part by the ARTEMISA International Chair of Cybersecurity and the DANGER Strategic Project of Cybersecurity funded by the Spanish National Institute of Cybersecurity through the European Union (EU)—NextGenerationEU and the Recovery, Transformation and Resilience Plan. The work of Minoru Kuribayashi was supported by JSPS KAKENHI under Grant 22K19777.

ABSTRACT During the past few decades, research on digital media watermarking –initially designed for digital images with the envisioned applications of copyright protection or copy control– has significantly evolved with respect to other covers (i.e., video, audio, speech) and many more potential applications, including tamper detection, broadcast monitoring, and, more recently, fake news detection. As a result, various surveys have tried to summarize certain aspects of this research field as it has grown. This has led to more than 130 survey papers being written at different points in time, describing various parts of the scientific efforts focused on digital media watermarking. Considering the above, the aim of this paper is twofold. First, we conduct a meta-survey based on 64 selected research works, in order to summarize the most notable survey papers in this field, which allows us to ''draw a map'' of this research area. Second, we focus on providing the requirements for digital watermarking techniques when applied to their most recent application: detecting fake news in multimedia content. Finally, an outline of the approach taken within the DISSIMILAR (Detection of fake newS on SocIal MedIa pLAtfoRms) project for the detection of disinformation is presented.

INDEX TERMS Digital watermarking, fake news detection, information hiding, meta-survey, signal processing.

I. INTRODUCTION

Digital watermarking is an essential part of information hiding research [1], [2], [3]. It involves the process of embedding data in the form of watermarks in carrier (host) signals, which are typically images, audio (speech) or video data [3], [4]. However, in the existing literature, the utilization of other signals has also been described, such as in 2D or 3D computer graphics [5], 2D vector maps [6], 3D printing models [7], text [5], network flows [8], databases [5], [9], various kinds of nonmedia data which can be subject to data mining process [10], or machine learning models, particularly neural

The associate editor coordinating the review of this manuscript and approving it for publication was Li He.

networks [11], [12]. Note that the embedded information can be related to the carrier signal and may be transparent or imperceptible.

Digital watermarking has many potential applications, including copyright protection, copy control, content authentication, tampering detection, and so on [13], [14]. One of the most recent applications of digital watermarking includes using these techniques for the identification and tracing of fake news, especially on social media [15]. This issue is becoming more and more pressing, as the ability to spread fake news across social media platforms has become easier, given that their users are able to create and share more information than ever before, some of which are deliberately deceitful. This has already impacted events in real, nondigital life (e.g., the 2016 presidential elections in the US). Although there have been some attempts to fight deepfakes [16] or fake news in images, they have not been proposed and analyzed for other types of digital content and have never been applied in a more complete system integrated with social media platforms.

It should be noted that, as with every innovation, digital watermarking comes with certain risks, and there are many types of potential attacks that such techniques need to withstand [13], [17], [18].

Considering the above and due to the large volume of research in digital watermarking spanning from the 1990s to present, we conducted a meta-survey (i.e., a survey of surveys) in order to arrange the existing research systematically, thus making it easier to understand the efforts, achievements, and overlaps in this area. It should be noted that existing surveys have been published specifically to describe various aspects of digital watermarking. As a result, existing works often overlap or were written in different time periods, making it hard to grasp what has been achieved in this area.

To the best of our knowledge, this paper contains the first comprehensive meta-survey of existing watermarking review papers including all types of multimedia content, namely, image, audio/speech, and video.

In more detail, the main contributions of this work are as follows:

- We conduct a meta-survey on digital watermarking in multimedia content by taking into account 133 existing surveys related to digital watermarking, of which 64 were selected to be described in this paper as we needed to leave out some of the reviews due to their low quality.
- We summarize the existing research in this area with respect to the specifics of the mentioned techniques, as well as discussed applications and attacks.
- Moreover, we holistically analyze what types of watermarking techniques can be helpful in the fake news detection process, which is the most recent application of digital watermarking.

The remainder of this article is organized as follows. Section II discusses the basics related to digital watermarking in multimedia content, in general. Next, in Section III, we describe the methodology used to select the digital watermarking surveys to create their bird-eye overview, which is presented in figure and tabular forms. Then, Sections IV-VII include the descriptions of existing surveys grouped by multimedia content, namely, image, video, audio / speech, and general articles. In Section VIII, the main conclusions drawn from the performed analyses of existing reviews are provided, while, in Section IX, the properties of digital watermarking required for fake news detection are outlined. Then, in Section X, promising research paths are characterized. Finally, Section XI concludes our work.

II. DIGITAL WATERMARKING FUNDAMENTALS

Digital watermarking [19] involves embedding some content-related data –referred to as a mark or watermark– into a digital object under certain constraints, such that the carried watermark, depending on the application, would be imperceptible and the watermark data could be extracted using a specific algorithm. The most common carriers used in digital watermarking are multimedia content, but even text and network protocols can be subject to watermark embedding. Although printed watermarks are usually intended to be designed in a visible manner, digital watermarks are mostly designed to be absolutely invisible. As such, invisible watermarking is generally utilized in digital multimedia communication systems [20].

From a visibility point of view, watermarks can be categorized into three types: visible, invisible, and dual watermarking. In a visible watermarking system, the embedded watermark is detectable by the human eye; for example, it may be a logo or text. On the contrary, in invisible watermarking, the embedded data are not observable to the human eye. Some applications of invisible watermarking include supporting data authentication and preventing data from being copied illegally. The combination of invisible and visible watermarks is called dual watermark, in which a visible watermark is first added to the carrier, following which an invisible watermark is embedded into the already visible-watermarked content [21].

A watermarking system typically involves three different steps: embedding, transmission, and extraction. In the embedding step, an algorithm inserts a digital watermark into the host with the help of a secret key to produce a watermarked signal. The watermarked signal is then broadcast or stored, and is typically sent to another entity. In such a case, any changes during transmission –either unwanted modifications stemming from noise in the transmission channels or intentional alterations by adversaries to the watermarked signal– are called attacks. In the third step, extraction of the watermark is accomplished, which legitimate users can perform (typically using a secret key and a detection algorithm). Figure 1 illustrates the general framework of the watermarking procedure.

If the detection algorithm requires the original host or cover for watermark extraction, it is known as a non-blind or

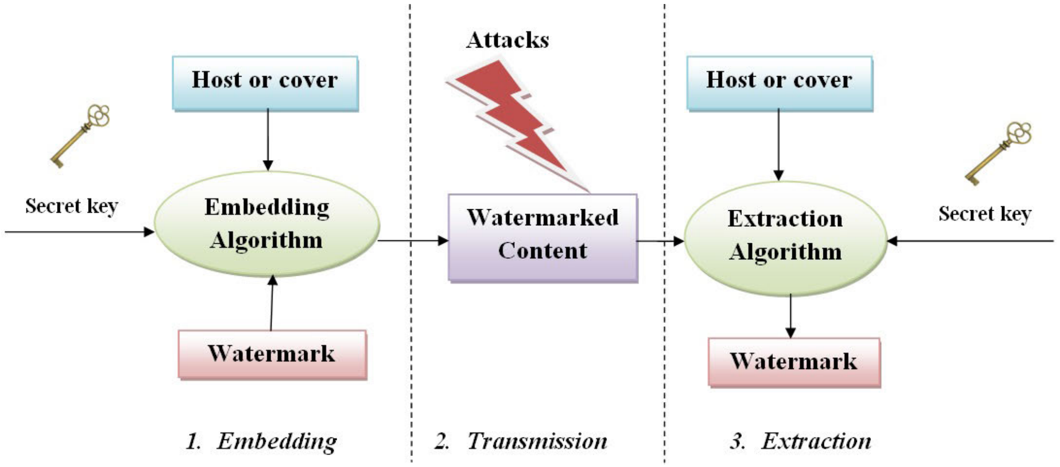

FIGURE 1. General framework of a watermarking procedure.

private watermark. In semiblind or semiprivate watermarks, the original host is not needed; however, for watermark extraction, the watermark object is required. Applications of these two types of watermarking are for offering evidence in court (to prove copy control or ownership of the information) or for fingerprinting (where the original recipients of the pirated copies can be identified). The most challenging type of watermarking is public or blind watermarking as, for watermark extraction, there is no need for either the original cover or the original watermark.

Embedding a digital watermark in a carrier may cause permanent loss for the original host. To avoid this damage, reversible digital watermarking has been proposed. In this technique, the protected data can be authenticated while the original data can be restored in a lossless manner. In other words, with this technology, small modifications to the data can be completely recovered without any additional cost. Lossless or reversible watermarking techniques are appropriate for specific applications in certain areas, including digital forensics, military, medical treatment, or smart metering systems [22].

Unlike other data hiding branches, such as steganography, in digital watermarking the embedded information is related to the carrier or cover object, and the digital object is considered to be valuable. The embedded information can be used to specify the copyright holder, grant access to the content for a limited number of users, identify the legitimate users of the content, provide data provenance information, or represent a hidden pattern that can be used to detect (or localize) tampered areas, among other possibilities.

Table 1 lists various types of applications of digital watermarking and their definitions, as well as their desired features suitable for each specific application, appropriate watermark design from the visibility perspective, and proper domain according to the watermarking usage. Bear in mind that the mentioned factors are not the only possible solutions for increasing the performance of the watermarking scheme for each particular application; for example, researchers may use a transform domain technique for tamper localization or detection, according to their own creativity.

The fundamental requirements for digital watermarking include imperceptibility, robustness, capacity, and security. The first three requirements act like a triangle; as a result, it is not possible to raise all three specifications at the same time. This means that increasing capacity and robustness will cause the degradation of the imperceptibility or quality, while the growth of imperceptibility and robustness will cause the capacity to decrease. However, the balance between these properties depends on the application. These principles are explained in the following [23].

Imperceptibility refers to the perceptual resemblance of the watermarked and original data. The watermark should be hardly noticeable, such that the end user cannot perceive any visual or audio effect from the watermarked content. Although the watermark is not supposed to degrade the quality of the content, a small amount of degradation is acceptable. The reason for this is to achieve high robustness or low cost in some applications. However, in visible watermarking, the watermark is embedded into the host such that it can be perceived without extraction [24], [25]. The embedded watermark should cause as less depreciation to the original cover/host as possible. If noticeable distortions are established in the cover, it provides a clue for attackers [26].

Robustness. The watermarked content should be robust, meaning that, despite the public principle of the watermarking algorithm, it should be impossible to remove and should resist a wide range of attacks [27]. In addition, the watermark must not be likely to be recovered or even altered without information about the secret key [28].

Capacity, or data payload, in a watermarking system refers to the amount of information that is embedded into the host and the number of bits that are encoded by the watermark. Depending on the application, the payload will be specified to be sufficient and facilitate the envisioned application [24].

TABLE 1. Applications of digital watermarking and desired features for optimum efficiency.

Application	Definition	Desired features	Watermark visibility	Suitable type of watermark	Suitable domain
Copyright protection	A watermark is used to provide information about the copyright holder of the work.	Imperceptibility, Robustness	Invisible	Robust	Transform
Content labeling	Additional information about the content is provided by an embedded watermark. The information is embedded in the content rather than being provided in the headers of the file, despite metadata.	Imperceptibility, Robustness	Invisible / visible	Robust	Transform
Copy control	The watermark is embedded to prevent the work from being copied.	Imperceptibility, Robustness	Invisible	Robust	Transform
Device control and legacy enhancement	Some hardware devices can be controlled using watermarks embedded in some content, such as television or radio transmissions.	Imperceptibility, Robustness	Invisible	Robust	Transform
Broadcast monitoring	Hidden content can be embedded in broadcast signals, so that specific programs, commercials, or copyrighted materials can be detected by an information system capturing and analyzing those signals.	Imperceptibility, Robustness	Visible	Robust	Transform
Content authentication	Authentication watermarks can be embedded in work to ensure its legitimacy.	Imperceptibility, Capacity	Invisible	Robust	Transform
Tamper detection	An embedded watermark can be used to detect whereas some digital content has been altered. In this way, if a (fragile) watermark is embedded in the content, failure to detect the same watermark in the receiver of the content can provide proof of tampering.	Imperceptibility, Capacity	Invisible	Fragile / Semi-fragile	Spatial
Tamper localization	Some fragile watermarking schemes can be used not only to detect if tampering has occurred but also to determine which areas or segments of the content have been altered by some attacker. The resolution of the tampered areas depends on the specific application.	Imperceptibility, Capacity	Invisible	Semi-fragile	Spatial
Transaction tracking or fingerprinting	In some applications, a merchant distributes copies of some content to buyers who can make legitimate use of the content. However, making identical copies of digital content is extremely easy with current technologies, and legitimate users have the possibility of redistributing the content illegally. In fingerprinting (or transaction tracking) applications, a different buyer-specific identifier is embedded in each copy of the content in such a way that the source of an illegal redistribution can be traced later on (traitor tracing).	Imperceptibility, Robustness	Invisible	Robust	Transform

Typically, capacity is measured in terms of bits per unit, such as bits per pixel for images or bits per second for audio and video.

Security. There are many ways in which attackers can extract hidden messages easily by trying an extraction algorithm or finding out the method approximately according to some known features of the watermarked content, such as Peak Signal to Noise Ratio (PSNR), that are related to the quality of the image. Therefore, it is a crucial issue to make the embedded watermark as secure as possible to prevent sensitive information from leaking [29].

In digital watermarking, it is essential to select a suitable domain for embedding the watermark based on the application, as this has a direct effect on the performance of the watermarking scheme [30]. The main reason is to fulfill the abovementioned watermarking principles, such as imperceptibility, capacity, robustness, and security, and to pay attention to the role of each domain and the specifications innately inherited by each particular selected domain with respect to the watermarked content [31].

Watermarking methods can also typically be classified based on their domain –spatial or transform. The use of each of these domains depends on the application. For example, for applications such as authentication, tamper detection, and data integrity, spatial domain methods are more compatible as they usually offer fragile watermarks; hence, the cover will be altered straightly with watermark embedding. The Least Significant Bit (LSB) and Spread Spectrum Modulation (SSM) techniques are examples of spatial domain techniques.

To the contrary, for applications requiring robustness, transform domain techniques are more suitable. In these techniques, the host data is transformed first, following which a digital watermark is embedded into the coefficients of the transformed host data. An inverse transform is needed to retrieve the original signal after embedding of the watermark. It has been proved that transform domain techniques can better resist compression and common signal processing attacks. Nevertheless, they are complex and have high computational costs while spatial domain techniques are computationally simple, fast, and straightforward [32].

Examples of the transform domain watermarking techniques include discrete cosine transforms (DCT), discrete Fourier transforms (DFT), discrete wavelet transforms (DWT), and singular value decomposition (SVD). These mentioned techniques have certain distinctions such as computational speed for real-time watermarking, but balancing between imperceptibility and robustness is not automatically possible. Generally, transform domain-based techniques offer more robustness and higher imperceptibility when the watermarked host data are faced with geometric and signal processing manipulations [33]. However, the computational costs of these techniques are higher than those of spatial domain-based methods.

Attacks on digital watermarking are classified into four categories, namely removal, geometric, cryptographic, and protocol attacks [25].

In signal processing or removal attacks, the primary aim is to remove the watermark signal without trying to endanger the security mechanisms of the proposed scheme. Therefore, in these types of attacks, there is no effort to realize the embedding technique or the encryption key. Thus, the result is damaging the watermarked content. Noise attacks, histogram equalization, and filtering attacks are included in this category [34].

In cryptographic attacks, the malicious party tries to find loopholes in the main embedding algorithm to remove the watermark information. Examples include brute force and oracle attacks. These attacks can be easily restricted if the embedding algorithm is complex [35].

In geometric attacks, the attacker attempts to distort the watermark signal geometrically. As a result, after applying these attacks, the synchronization of the embedded watermark will be demolished in the watermarked data [36]. If the geometry of the attack is devised, it is theoretically possible to detect or recover the original watermark; however, such detection or recovery procedures are complex, expensive, and slow. Rotation, scaling, translation, and cropping are examples of this type of watermark attack [25].

The main aim of the protocol attacks is to attack the entire concept of watermarked content. Considering the invertible watermark, the attacker segregates their own watermark from the watermarked data, in order to claim that they are the owner of the watermarked content. This can create ambiguity regarding the true ownership of the data.

As a result, for copyright protection purposes, the watermark needs to be non-invertible. Copy attack is the other type of protocol attack in which the goal is to estimate the watermark from watermarked image and copy it to some other image, called the target image. The copy attack is applicable when the attacker can create a legitimate watermark in the host content without knowing the algorithm of the watermarking scheme or the secret key [37]. Note that this attack is different from a copy-move attack.

Apart from the abovementioned attacks, some other attacks are identified in Tables 2 and 3, such as copy-move, collage, and vector quantization. These attacks are related to passive multimedia forgery detection, which refers to the authentication of multimedia and detection of tampered areas. For example, when forging a digital image, its statistical characteristics are also changed. Computing the statistical characteristics of each part of an image and comparing it with the intact data will result in tamper detection and localization of the manipulated areas of the data. Data forgery detection is generally categorized into passive and active detection. Tamper detection using the statistical characteristics of data is classified into the passive detection group, while active detection methods using digital watermarking or digital signatures are prevalent. In the following, we define three attacks related to passive detection [38].

Copy-move attacks, as their name implies, copy some part of the multimedia data and then paste it into another part of the same data. This kind of attack is difficult to detect, as the source and destination of the tampered area are the same and, as a result, the similarity of their statistical characteristics makes it difficult to detect tampered regions [39].

In contrast to copy-move attacks, collage attacks involve generating new watermarked data from multiple authenticated watermarked content using a combination of areas from different data and conserving their relative spatial positions. Moreover, all the watermarked data used in a collage attack are built by an identical watermarking method with the same secret keys [40].

Vector quantization (VQ) attacks are another type of tampering attack against block-wise independent watermarking schemes, in which every watermarked block relies merely on its original block. Given a set of watermarked content blocks, a VQ codebook can be generated. Based on this code book, fake data containing the sham watermark is then created [40].

Watermarking techniques based on robustness can be further divided into three main categories: robust, fragile, and semi-fragile [41]. Robustness is an important criterion which indicates the ability of the hidden watermark to resist malicious attacks. Robust watermarking is mainly designed for copyright protection, where the watermark can be detected even after exposing the watermarked data to malicious attacks. In fragile watermarking, the watermark can be simply affected by any type of manipulation, either malicious attacks or non-intentional ones. Fragile watermarking was invented for testing integrity and content authentication, where the slightest change in the watermarked data is noticeable or simply detectable. Finally, in semi-fragile watermarking, the hidden watermark can resist gentle transformations (e.g., compression), but not malevolent attacks. Semi-fragile watermarks were developed for and widely used in content verification. Semi-fragile watermarking is robust to acceptable content-preserving manipulations, while fragile to malicious distortions such as feature adding or removal, so it is suitable for proving the reliability of data [42].

In this section, the general requirements for digital watermarking were described. In the following sections, a wide range of surveys on multimedia watermarking –including digital image, video, audio, and speech watermarking, as well as other general multimedia contents– are explored.

III. BIRD-EYE OVERVIEW OF THE EXISTING SURVEYS AND PAPER SELECTION PROCESS

As mentioned above, one of the main aims of this work is to capture and review the existing literature on digital watermarking in multimedia content. We considered the existing literature covering the period 1996-2022, and in total, 64 review papers were included in the final analysis.

The selection of papers included in the meta-survey on digital watermarking for fake news detection was performed in the second half of 2021. Below, we briefly explain the methodology used for this process. Our approach was inspired by the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) [43] method, commonly used to perform literature analysis (see, e.g., [44], [45], [46], [47]).

First, we searched for potentially relevant papers using the Google Scholar search engine. It was selected as a source of papers, as it indexes publications from various databases. The goal of this study was to identify existing valuable reviews and analyses of multimedia watermarking methods. Therefore, our main focus was surveys on the watermarking of three typical types of multimedia files –that is, audio, video, and images– as well as general surveys describing watermarking techniques for various kinds of cover objects. Consequently, we used keywords formulated according to the following rule, written using Extended Backus-Naur Form (EBNF) [48] notation:

[“digital”], (“watermarking”|“fingerprinting”), [“audio” | “video” | “image” | “multimedia”], [[“meta”] “survey” | “fake news” [“detection”]];

All combinations of keywords resulting from the above rule were used as an input for Google Scholar. There was no restriction on the publication date of the obtained papers, nor were any other filters used in the search process.

Titles and abstracts of papers found on that basis were examined to determine their potential relevance. All result pages were studied until no more relevant articles were found in several subsequent pages.

In addition, all co-authors of this paper added surveys on the digital watermarking of multimedia to which they had access and which, in their opinion, could be meaningful for this meta-survey. After this initial procedure, we removed duplicates from the set of identified publications. At this stage, 167 papers remained for consideration. Each paper was assessed by exactly one of the co-authors.

At this point, some papers were excluded from further analysis, based on the following exclusion criteria. First, after a closer examination of the contents of the papers, 27 were identified as being out of the scope of this meta-survey; in particular, 19 of them turned out to be not devoted to the topic of multimedia watermarking, while the other 8 articles were only very weakly related to that topic. Moreover, another 7 papers turned out to be dedicated to the watermarking of other types of cover objects. Although they were not strictly relevant to the analysis in our study, we decided to use them as examples of other applications of the watermarking approaches, as detailed in Section I. As a result of these decisions, there were 133 publications left for in-depth analysis at the end of this step.

Next, the quality of all of the 133 relevant articles mentioned above was assessed. Papers that were declared to be surveys but consisted of only a few pages or contained only a few references were rejected at this point. The same applied to publications with obvious graphical or layout errors, as we assumed that high-quality papers should be prepared carefully in all aspects, including not only the text but also their final appearance. Unfortunately, we identified 69 such low-quality papers, meaning that, for this reason, we had to reject more than half of all relevant papers.

Consequently, in the end, we obtained 64 relevant papers of quality, which did not seem doubtful at first sight. All of these papers are analyzed and described in this meta-survey. The co-authors divided the papers by the type of cover object to which publications were dedicated (i.e., audio, video, images, or multimedia in general). Then, groups of articles determined in this way were assigned to the groups of co-authors. Each of the following Sections (IV - VII) was prepared by a separate group of co-authors. Each paper was read and described by exactly one of the co-authors.

Moreover, the described surveys are also characterized in a tabular form (see Tables 2 and 3), where the most essential features of each work are included (i.e., what kind of cover objects are considered as well as the applications and attacks analyzed). Additionally, we divided the survey articles using the type of watermark (i.e., robust or fragile/semi-fragile). It should be noted, however, that some surveys mentioned both types of watermarks. In such a case, we list them twice in Tables 2 and 3 (under both robust and fragile/semi-fragile categories). Otherwise, each survey is assigned a separate row in these tables. Finally, all works are also presented on the timeline in Figure 2 in order to present their distribution over time. As can be seen from the figure, almost half of the surveys were published in the last five years.

In the following sections, we describe the existing survey articles in more detail. We classify these papers according to the type of content they consider, namely, images, video, audio/speech, and general content review papers.

IV. SURVEYS DEVOTED TO DIGITAL WATERMARKING IN IMAGES

Through the use of image editing software, manipulation of digital images is possible, even without any professional knowledge. Attackers can subtly tamper with digital images by misusing the human visual system (HVS). Thus, digital image watermarking can prevent illegal alteration and unlawful admission to digital images by embedding a digital signal into a host image, such that the original information is perceptually kept intact, according to the HVS. Additionally,

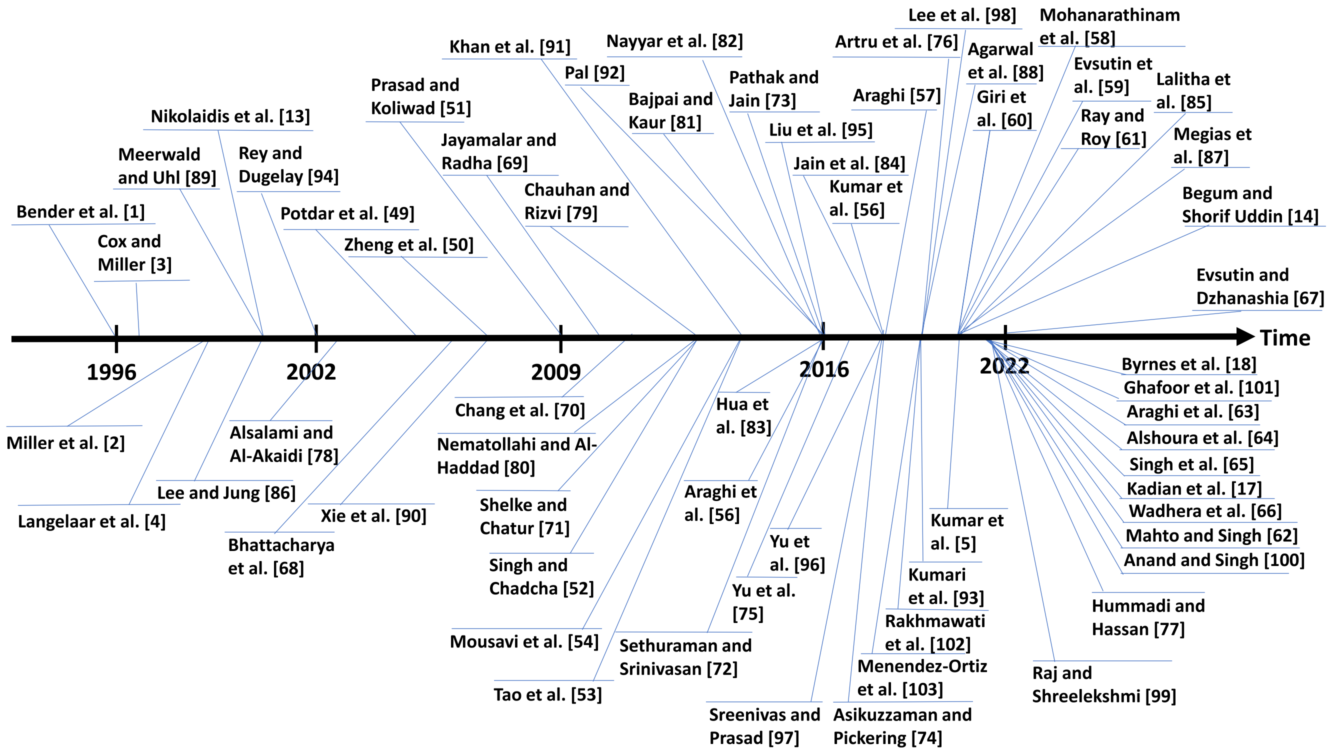

FIGURE 2. Temporal distribution of the surveys described in this paper.

the bits carrying the watermark should be scattered all over the host, such that they cannot be recognizable for manipulation. In the following, the main existing surveys on image watermarking from Tables 2 and 3 are explained, and a deep analysis of each survey is presented.

Rey and Dugelay [94] published a survey paper with the aim of introducing the notion of image authentication algorithms. They categorized these algorithms as fragile and semi-fragile watermarking. The authors also considered another alternative to classical watermarking: digital signatures. They concluded that, in general, only fragile watermarking techniques offer strict integrity, while semi-fragile and digital signature techniques assure content authentication as they can resist some manipulations, such as JPEG compression, and avoid false positive tampering. Furthermore, fragile watermarking schemes are easy to implement, while only limited methods from each category can restore the tampered areas of the images.

Podtar et al. [49] published a paper in which they mentioned digital watermarking techniques, requirements, and applications. The authors focused mostly on transform domain techniques for the purpose of copyright protection on digital images. General watermarking algorithms for DWT, DCT, and DFT techniques were discussed. Finally, the authors concluded the advantages and disadvantages of each of these transform domain techniques, which can resist rotation, scaling, and translation (RST) attacks. As a result, they can be mostly used for recovering geometric distortions. DWT has higher computational complexity, in comparison with DFT and DCT; however it also presents better compatibility with the HVS for information hiding, as there are less evident visual artifacts compared to other techniques. Moreover, a multi-resolution presentation of an image is achieved when it is decomposed by the wavelet transform, making it possible to display the image in different resolution levels, from low to high.

Zheng et al. [50] reviewed different image watermarking schemes based on transformations that are invariant against rotation, and RST operations. They classified RST-invariant image watermarking algorithms into seven different categories: RST-invariant domain-based algorithms, Radon transform-based algorithms, template-based algorithms, salient-feature-based algorithms, image decomposition-based algorithms, stochastic analysis-based algorithms, and others. The different approaches were evaluated in terms of robustness, capacity, and imperceptibility. The authors concluded that the existing RST-invariant image watermarking algorithms have pros and cons, but no ideal solution providing true robustness against RST transformations, blind watermark detection/extraction, and fast and correct watermark detection with a low error rate exists.

Prasad and Koliwad [51] presented a survey of image watermarking methods for copyright protection applications. They established the requirements for an ideal image watermarking approach for the considered type of application. The required properties are imperceptibility; robustness

TABLE 2. Summary of robust digital watermarking-related surveys. Legend: Applications - CP: copyright protection, CC: copy control, DC: device control, BM: broadcast monitoring, FP: fingerprinting; Attacks - RM: removal, GM: geometric, CR: cryptographic, PR: protocol, CM: copy/move, CG: collage, VQ: vector quantization; Symbol meaning: "√" denotes that certain aspect is mentioned in the survey; "×" that it is intentionally not mentioned in the survey; "N/A" that certain information is missing in the survey.

Survey	Application					Attacks							Cover object		
	CP	CC	DC	BM	FP	RM	GM	CR	PR	CM	CG	VQ	Image	Speech / Audio	Video
Potdar et al. [49]	√	×	×	×	×	√	√	√	√	×	×	×	√	×	×
Zheng et al. [50]	√	×	×	×	×	×	√	×	×	×	×	×	√	×	×
Prasad and Koliwad [51]	√	√	N/A	√	√	√	√	N/A	N/A	√	√	√	√	×	√
Singh and Chadcha [52]	√	√	N/A	√	√	√	√	√	√	×	√	×	√	×	×
Tao et al. [53]	√	√	N/A	N/A	√	√	√	√	√	N/A	√	N/A	√	×	√
Mousavi et al. [54]	√	N/A	N/A	N/A	√	√	√	N/A	N/A	N/A	√	N/A	√	×	×
Araghi et al. [55]	√	N/A	N/A	N/A	N/A	√	√	N/A	N/A	N/A	N/A	N/A	√	×	×
Kumar et al. [56]	√	√	N/A	√	√	N/A	N/A	N/A	N/A	N/A	N/A	N/A	√	×	×
Araghi [57]	√	N/A	N/A	N/A	N/A	√	√	N/A	N/A	N/A	N/A	N/A	√	×	×
Mohanarathinam et al. [58]	√	N/A	N/A	N/A	N/A	√	√	N/A	N/A	N/A	N/A	N/A	√	×	×
Evsutin et al. [59]	√	N/A	N/A	N/A	N/A	√	√	N/A	N/A	N/A	N/A	N/A	√	×	×
Giri et al. [60]	√	N/A	N/A	N/A	N/A	√	√	N/A	N/A	N/A	N/A	N/A	√	×	×
Ray and Roy [61]	√	N/A	N/A	N/A	N/A	√	√	×	×	×	×	×	√	×	×
Begum and Shorif Uddin [14]	√	√	√	√	√	√	√	√	√	N/A	N/A	N/A	√	×	√
Mahto and Singh [62]	√	√	N/A	N/A	N/A	√	√	N/A	√	N/A	N/A	N/A	√	×	×
Araghi et al. [63]	√	×	×	×	×	√	√	×	×	×	×	×	√	×	×
Alshoura et al. [64]	√	×	×	×	×	√	√	×	×	×	×	×	√	×	×
Singh et al. [65]	√	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	√	×	×
Kadian et al. [17]	√	√	N/A	N/A	N/A	√	√	N/A	N/A	N/A	N/A	N/A	√	×	×
Wadhera et al. [66]	√	N/A	N/A	N/A	N/A	√	√	√	√	N/A	N/A	N/A	√	×	×
Evsutin and Dzhanashia [67]	√	N/A	N/A	N/A	N/A	√	√	√	√	×	×	×	√	×	×
Bhattacharya et al. [68]	√	N/A	N/A	N/A	√	N/A	×	×	×	N/A	N/A	N/A	√	×	√
Jayamalar and Radha [69]	√	√	N/A	N/A	N/A	√	√	N/A	N/A	N/A	√	N/A	√	×	√
Chang et al. [70]	√	N/A	N/A	N/A	N/A	√	N/A	N/A	N/A	N/A	N/A	N/A	×	×	√
Shelke and Chatur [71]	√	√	√	√	√	√	√	N/A	N/A	N/A	N/A	N/A	√	×	√
Sethuraman and Srinivasan [72]	√	N/A	N/A	N/A	N/A	√	N/A	√	N/A	N/A	N/A	N/A	√	×	√
Pathak and Jain [73]	N/A	N/A	N/A	N/A	N/A	√	N/A	N/A	N/A	√	N/A	N/A	×	×	√
Asikuzzaman and Pickering [74]	√	√	N/A	√	√	√	√	N/A	N/A	N/A	N/A	N/A	√	×	√
Yu et al. [75]	√	√	√	√	√	√	√	√	N/A	N/A	N/A	N/A	√	×	√
Artru et al. [76]	√	N/A	√	√	√	√	√	√	N/A	√	N/A	N/A	√	√	√
Hummadi and Hassan [77]	√	√	N/A	√	√	√	√	√	N/A	N/A	N/A	N/A	×	×	√
Alsalami and Al-Akaidi [78]	√	√	√	N/A	√	√	√	√	×	N/A	N/A	N/A	×	√	×
Chauhan and Rizvi [79]	√	√	√	√	√	√	N/A	×	×	N/A	N/A	N/A	×	√	×
Nematollahi and Al-Haddad [80]	√	√	√	√	√	√	×	×	×	N/A	N/A	N/A	×	√	×
Bajpai and Kaur [81]	√	N/A	N/A	N/A	N/A	√	×	×	×	N/A	N/A	N/A	×	√	×
Nayyar et al. [82]	√	N/A	N/A	N/A	N/A	√	N/A	N/A	N/A	N/A	N/A	N/A	√	√	×
Hua et al. [83]	√	√	√	√	√	√	√	√	√	√	N/A	N/A	×	√	×
Jain et al. [84]	√	×	×	×	×	√	×	×	×	×	×	×	×	√	×
Lalitha et al. [85]	√	N/A	√	√	√	√	√	×	×	N/A	N/A	N/A	×	√	×
Lee and Jung [86]	√	N/A	N/A	N/A	N/A	√	√	√	N/A	N/A	N/A	N/A	√	√	√
Nikolaidis et al. [13]	√	√	N/A	√	√	√	√	N/A	√	N/A	N/A	N/A	√	N/A	√
Cox and Miller [3]	√	N/A	N/A	√	√	√	√	N/A	N/A	N/A	N/A	N/A	√	√	√
Megías et al. [87]	√	N/A	N/A	N/A	√	√	N/A	N/A	N/A	N/A	N/A	N/A	√	√	√
Langelaar et al. [4]	√	√	N/A	√	√	√	√	N/A	√	N/A	N/A	√	√	N/A	√
Agarwal et al. [88]	√	√	N/A	√	√	√	√	√	N/A	N/A	N/A	√	√	√	√
Kumar et al. [5]	√	√	√	√	√	√	√	√	√	√	N/A	N/A	√	√	√
Bender et al. [1]	√	√	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	√	√	√
Miller et al. [2]	√	√	N/A	√	√	N/A	√	N/A	√	√	N/A	N/A	√	√	√
Meerwald and Uhl [89]	√	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	√	×	√
Xie et al. [90]	√	√	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	√	√	√
Khan et al. [91]	√	√	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	√	√	√
Pal [92]	√	√	N/A	√	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	√	×	×
Kumari et al. [93]	√	√	N/A	√	√	√	√	√	√	√	N/A	N/A	√	√	√
Byrnes et al. [18]	√	√	N/A	N/A	N/A	√	√	N/A	N/A	√	N/A	N/A	√	√	√

against signal processing, geometric distortions, collusion, and forgery; universal applicability to image, video, and audio; and unambiguity in identifying the copyright holder. Their analysis of the existing works considered spatial domain methods, frequency domain methods, and others. They described different solutions published in the literature but made no comparison among the different techniques – either qualitative or quantitative– in terms of the required properties. Hence, the authors presented an overview of the different techniques but did not make any particular recommendation to select any of them. Further research challenges were not identified either.

Singh and Chadha [52] provided an overview of several digital watermarking techniques, applications, and attacks. Although their paper referred to general contents (audio, images, and video), the review and comparison of the different techniques focused on image watermarking. The reviewed schemes were classified into spatial domain

TABLE 3. Summary of fragile / semi-fragile digital watermarking-related surveys. Legend: Applications - CA: content authentication, TD: Tamper detection, AU: authentication, TL: tamper localization, CL: content labeling; Attacks - RM: removal, GM: geometric, CR: cryptographic, PR: protocol, CM: copy/move, CG: collage, VQ: vector quantization; Symbol meaning: "√" denotes that certain aspect is mentioned in the survey; "×" that it is intentionally not mentioned in the survey; "N/A" that certain information is missing in the survey.

Survey	Application					Attacks							Cover object		
	CA	TD	AU	TL	CL	RM	GM	CR	PR	CM	CG	VQ	Image	Speech / Audio	Video
Rey and Dugelay [94]	√	√	√	√	N/A	√	N/A	√	√	√	√	N/A	√	N/A	N/A
Prasad and Koliwad [51]	√	N/A	√	N/A	N/A	√	√	N/A	N/A	√	√	√	√	×	√
Singh and Chadcha [52]	√	√	√	N/A	N/A	√	√	√	√	×	√	×	√	×	×
Mousavi et al. [54]	N/A	N/A	√	N/A	N/A	√	√	N/A	N/A	N/A	√	N/A	√	×	×
Liu et al. [95]	√	√	√	√	N/A	√	√	√	√	√	√	√	√	×	×
Yu et al. [96]	√	√	√	√	N/A	√	√	√	√	√	√	√	√	×	×
Sreenivas and Prasad [97]	√	√	√	√	N/A	N/A	N/A	N/A	√	√	√	√	√	×	×
Lee et al. [98]	N/A	√	√	√	N/A	√	N/A	N/A	N/A	N/A	N/A	N/A	√	×	×
Mohanarathinam et al. [58]	√	√	√	√	N/A	√	√	N/A	N/A	N/A	N/A	N/A	√	×	×
Evsutin et al. [59]	√	N/A	N/A	N/A	N/A	√	√	N/A	N/A	N/A	N/A	N/A	√	×	×
Begum and Shorif Uddin [14]	√	√	√	√	√	√	√	√	√	N/A	N/A	N/A	√	×	√
Mahto and Singh [62]	√	√	√	N/A	N/A	√	√	N/A	√	N/A	N/A	N/A	√	×	×
Raj and Shreelekshmi [99]	N/A	√	√	√	×	√	×	√	√	√	√	√	√	×	×
Singh et al. [65]	√	√	√	√	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	√	×	×
Anand and Singh [100]	N/A	N/A	√	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	√	×	×
Ghafoor et al. [101]	√	√	√	√	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	√	×	√
Rakhmawati et al. [102]	√	√	√	√	N/A	√	√	N/A	N/A	√	√	√	√	×	×
Nematollahi and Al-Haddad [80]	√	N/A	√	N/A	N/A	√	×	×	×	N/A	N/A	N/A	×	√	×
Lee and Jung [86]	√	√	N/A	N/A	N/A	√	√	√	N/A	N/A	N/A	N/A	√	√	√
Nikolaidis et al. [13]	√	√	N/A	√	√	√	√	N/A	√	N/A	N/A	N/A	√	N/A	√
Cox and Miller [3]	×	N/A	N/A	N/A	N/A	√	√	N/A	N/A	N/A	N/A	N/A	√	√	√
Menendez-Ortiz et al. [103]	√	√	N/A	√	√	√	√	N/A	N/A	N/A	√	√	√	√	√
Langelaar et al. [4]	√	√	N/A	√	√	√	√	N/A	√	N/A	N/A	√	√	N/A	√
Agarwal et al. [88]	√	√	N/A	N/A	N/A	√	√	√	N/A	N/A	N/A	√	√	√	√
Kumar et al. [5]	√	√	N/A	√	N/A	√	√	√	√	√	N/A	N/A	√	√	√
Bender et al. [1]	N/A	√	N/A	√	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	√	√	√
Miller et al. [2]	√	√	√	N/A	N/A	N/A	√	N/A	√	√	N/A	N/A	√	√	√
Meerwald and Uhl [89]	√	√	N/A	N/A	√	N/A	N/A	N/A	N/A	N/A	N/A	N/A	√	×	√
Xie et al. [90]	√	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	√	√	√
Khan et al. [91]	√	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	√	√	√
Pal [92]	√	N/A	√	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	√	×	×
Kumari et al. [93]	√	N/A	√	N/A	√	√	√	√	√	√	N/A	N/A	√	√	√
Byrnes et al. [18]	√	√	√	N/A	N/A	√	√	N/A	N/A	√	N/A	N/A	√	√	√

schemes, additive watermarking, LSB-based schemes, SSM techniques, texture mapping coding techniques, patchwork algorithms, correlation-based techniques, and frequency domain methods (DCT, DWT, and DFT). The authors compared the different categories, regarding their advantages and disadvantages, but did not relate each category to possible application scenarios. In addition, they remarked on two challenges of existing watermarking schemes to be addressed in future research, namely, exploitation of the visual characteristics of signals to find better ways of concealing the hidden watermark, and making the best possible use of the HVS to determine the best way to embed the watermark. A detailed comparison of different schemes in terms of quantitative indicators was missing.

Tao et al. [53] analyzed different alternatives for robust image watermarking schemes. The reviewed schemes were classified into two main categories: transform domains (including SVD and the DWT) and geometric-invariant watermarking (exhaustive watermark search, template-based approach, and self-synchronizing schemes, and invariant domain). The different schemes were analyzed in terms of five relevant properties: imperceptibility, capacity, robustness, security, and false positive rate. However, no explicit comparison in terms of qualitative or quantitative indicators was provided, and no directions for future research were suggested.

Mousavi et al. [54] provided a survey of watermarking techniques focused on medical imagery. The authors reviewed and compared several medical image watermarking schemes according to the following criteria: image modality (including magnetic resonance, MRI; ultrasound; computed tomography, CT; or X-ray, among others), objective (authentication, control, integrity, data hiding, etc.), type of watermark (electronic patient record, hash, and patient information, among others), embedding region (region of interest/non-interest, ROI or RONI; or whole image), embedding technique (difference expansion, DWT, LSB, and so on), reversibility, tamper localization, tamper recovery, and fragility/robustness. The comparison among the different schemes was qualitative, reporting the values of each of the different criteria, but no quantitative analysis was provided.

In their survey paper, Liu et al. [95] presented a comprehensive report on fragile watermarking schemes for image authentication. They classified these schemes into three types –namely, spatial, frequency, and compression domains– and also divided each main type into two subtypes: recoverable and unrecoverable. Finally, they suggested a feature analysis for all schemes in order to compare them in terms of localization, reconstruction capacity, visual quality, complexity, sensitivity, and robustness. In their conclusions, they remarked on the importance of reversibility and good tamper localization while maintaining image quality in future research.

Araghi et al. [55] presented a survey on feature-based extraction techniques for image watermarking. Such methods have attracted the attention of the watermarking community and are based on including image content in the watermarking process. In feature-based detection, first, the interest points for embedding and extracting the watermark are specified. Then, selected areas are transformed to the identified size, orientation, and shape for the function of embedding and extraction. In this survey, the authors investigated chosen feature-based techniques, such as those using the Harris-Laplacian transform, affine covariant regions, binary patch, and Zernike transform. Finally, the influential factors considered in order to obtain optimum performance using feature-based extraction techniques were robustness, false positive detection, degradation of quality, and sufficient security, speed, and operational costs.

Yu et al. [96] carried out a review on semi-fragile watermarking schemes. They introduced the fundamental theory of semi-fragile watermarking technology and analyzed several types of attacks. The authors also analyzed the performance of the schemes according to their ability to recover tampered areas. They concluded that the existing semi-fragile watermarking schemes are not able to distinguish between malicious and non-malicious attacks. As many of these algorithms are based on image blocking, the detection of tampered areas is also based on the image blocks, which decreases the accuracy of such algorithms. They recommended the use of artificial intelligence-based approaches to alleviate this problem.

Kumar et al. [56] provided a survey of digital image watermarking schemes and their application to e-governance. The authors reviewed the properties and applications of image watermarking, and summarized the state-of-the-art techniques in this field, and compared them in tabular form according to the proposed objective, blind/informed extraction, domain of the embedded watermark, the size of the watermark, and the carrier work. However, they did not remark on open research challenges or suggest future research directions.

In the review by Sreenivas and Prasad [97], the authors explored different fragile watermarking schemes for the purpose of image authentication, tamper detection, and localization through recovery techniques. They also presented a framework for general fragile watermarking schemes and reviewed different metrics for watermarked and recovered images and various attacks and issues on fragile watermarking. They reviewed some fragile watermarking schemes based on pixel grouping and probability distribution functions, Hamming codes, transformed domains (DCT/DWT), and miscellaneous approaches, in order to evaluate the schemes for localizing tamper areas and their recovery. They concluded that self-embedding schemes on fragile watermarking have recently been developed for tamper localization by distributing recovery bits through the image to eliminate loss of the watermark data and tamper coincidence, as well as to guarantee tamper localization by including an image digest in the embedding procedure using several copies of the watermark. However, they pointed out that, at present, no scheme is perfect for image authentication and tamper recovery.

Araghi [57] reviewed histogram modification-based schemes in digital image watermarking against geometric and signal processing attacks to offer robustness for copyright protection. In this paper, the important aspects affecting the robustness, imperceptibility, capacity, and security of the existing algorithms, according to the strength and weaknesses of each scheme, were described. Histogram modification methods suffer from various vulnerabilities similar to the unsteadiness of the histogram shape stemming from image contrast. However, the author concluded that the use of certain techniques, such as selecting adjacent bins intelligently in watermark embedding, using secret keys, and choosing constant points of the cover images by revealing them under signal processing and geometric attacks before embedding the watermark, can make them excellent candidates in the context of image watermarking to withstand the mentioned attacks.

Lee et al. [98] focused on four diverse types of semi-fragile watermarking methods based on DWT, DCT, and VQ for image authentication. Most of their discussion was focused on watermark creation, encryption, and embedding procedures. Finally, the methods were compared based on the quality of the watermarked image, rate of tamper detection, block size, and quality of image recovery.

In the paper proposed by Mohanarathinam et al. [58], watermarking techniques were grouped based on the domain, such as spatial and transform domains, and HVS according to fragility and robustness. They also reviewed papers on reversible watermarking and watermarking with recovery. The survey aimed to introduce hybrid watermarking techniques for copyright protection, validation, and multimedia applications such as Facebook, Twitter, and WhatsApp while considering attacks and metrics for performance evaluation of the techniques.

Next, Evsutin et al. [59] classified existing data hiding techniques into the five groups of spatial, frequency, reversible, image content-based, and edge detection. In each group, they compared the different techniques according to the purpose of usage, embedding operation, key features, size of the cover, capacity, imperceptibility, robustness, and the attacks that can be countered by each technique. They concluded that the essential features to be emphasized in future research are reversibility, robustness, and resistance to steganalysis.

In the review by Giri et al. [60], the authors explored existing DWT-based color image watermarking techniques. They mainly focused on the type of wavelet filter, levels of decomposition, color space, and various optimization techniques, such as the multi-objective bees algorithm (MOBA), genetic algorithm (GA), particle swarm optimization (PSO), and so on, used to enhance the performance of the watermarking schemes. The authors claimed that the DWT is the most suitable technique for the watermarking process, as it can maintain the time/frequency decomposition characteristics that are compatible with the HVS.

Ray and Roy [61] focused on surveying watermarking techniques for copyright protection of images. They classified the research and investigations done in image watermarking techniques for copyright protection based on machine learning and deep learning techniques, a combination of encryption, cryptography, and error correction code (ECC), hybrid techniques such as DWT and SVD, biometric watermarking like iris scan and X-ray for medical images, visual cryptography, and bio-inspired algorithms such as firefly and bat optimization.

Begum and Shorif Uddin [14] explored fragile watermarking methods. Their focus is mostly on various watermarking techniques in the spatial and transform domains. They also reviewed challenges and attacks on image watermarking methods and provided a cost-effectiveness investigation based on different attack scenarios. In the end, the authors concluded that DWT is the high-quality and most robust technique for image watermarking for its multi-resolution characteristics.

Mahto and Singh [62] surveyed state-of-the-art color image watermarking schemes. They classified these schemes by RGB, YCbCr, and other model images with a detailed investigation of each group. Finally, the authors suggested some research directions such as using dual watermarking for copy protection with content integrity by adopting Zernike moments for better robustness to resist noise and geometric attacks, and usage of Redundant DWT (RDWT) instead of DWT in order to eliminate the shift-variance and down-sampling issues in transform domain-based schemes.

Raj and Shreelekshmi [99] reviewed fragile watermarking schemes for tamper detection, localization, and self-recovery. They classified tamper localization schemes into hash-based, chaos-based, and block feature-based schemes. The authors simulated each of these schemes to make a comparison between them. According to their observation, hash-based schemes offer high accuracy in detecting tampered areas, whereas chaos-based schemes are more precise in detecting untampered regions. Moreover, the calculation speed of hash-based schemes is higher than that of the other schemes. The authors also categorized self-recovery schemes into five groups: block average, block truncation code, reference sharing, singular value decomposition, and transform-based schemes. Their investigation showed that block average-based schemes perform faster than other schemes. The computational complexity for all schemes was O(N), in which N stands for the number of pixels in the cover image. However, the required time for watermark embedding differs for each scheme, depending on the number of operations required for the image blocks.

Araghi et al. [63] presented an overview of template-based watermarking methods. Templates are recognizable patterns or tiling signals which do not carry any essential information. They are added to cover images in addition to the watermark for easy retrieval of alteration of the cover images. The authors investigated these methods in both spatial and transform domains, and stated that the Pyramidal Just Noticeable Difference (PJND) has high accuracy and increases robustness against many geometric and signal processing attacks, while utilizing a curvelet transform can help to hide templates and allow the capacity of the methods to be independent of the cover image.

Alshoura et al. [64] devoted their review to robust hybrid SVD-based image watermarking schemes. These analyses included embedding methods, type of watermark, scaling factors, and watermark encryption. The authors presented a comprehensive review of false positive problems and solutions in SVD watermarking, which can be considered one of the critical issues in these schemes. Finally, some recommendations were proposed in order to increase the performance of hybrid SVD-based image watermarking. These recommendations can be summarized as encryption of the watermark before embedding; usage of SVD components as secret keys; taking advantage of optimization algorithms such as differential evolution (DE), artificial bee colony (ABC), and so on; use of a blocking strategy; and embedding the watermark into the host image redundantly.

In the survey presented by Singh et al. [65], various soft computing methodologies for image watermarking were reviewed. Soft computing techniques such as neural network (NN), GA, support vector machine (SVM), principal component analysis (PCA), Meta-Heuristic approaches, deep learning (DL), and fuzzy logic (FL) are primarily used in order to achieve an optimal balance between watermarking requirements such as capacity, robustness, and imperceptibility. The authors pointed out that, although soft computing-based watermarking can offer high-performance watermarking, some concerns need to be taken into consideration. For example, SVM classifiers, neural networks, and optimization techniques have high computational complexity while, in genetic algorithm and fuzzy logic techniques, fitness functions and inference rules must be designed based on the application.

The authors of [17] and [66] surveyed various types of digital image watermarking techniques for copyright protection based on the domain, characteristics of robustness and fragility, visibility and invisibility, and blind or non-blind schemes. They also classified the techniques based on attacks, applications, and imperceptibility. They concluded that hybrid watermarking techniques are superior, in comparison to other simple schemes.

In another survey, conducted by Anand and Singh [100], different watermarking techniques for the authentication of medical images were investigated. Digital imaging and communications in medicine (DICOM) are basically devoted to transmitting medical records over an open network. Nonetheless, there are several issues, such as protecting significant information and demand for extra bandwidth, which confine the transmission of such data over public networks. In this survey, the authors compared medical image watermarking techniques in various spatial and transform domains based on real samples of medical images such as CT, MRI, electrocardiogram (ECG), X-ray, and so on. The result of this comparison was summarized by considering a reversible blind watermarking approach for medical watermarking applications. The authors also suggested that hardware-based watermarking can be used to support real-time applications. Finally, they recommended that the hybridization of two or more watermarking techniques will enhance watermarking performance; however, this will also increase the computational complexity.

Ghafoor et al. [101] provided an overview of reversible watermarking in echocardiography applications. Reversible watermarking is an efficient tool which has uses in critical domains such as military and medical applications. The authors classified different reversible watermarking techniques with respect to spatial and frequency domains. In the spatial domain, they described four types of techniques: motion vector reversible watermarking, LSB, Harris Corner Detection and Fuzzy C-Means, and ROI-Based Tamper Detection and Recovery. Furthermore, in the frequency domain, they categorized eight types of reversible watermarking: Detection of Accurate Tamper in Region of Interest-based, Texture Quantization-Based Reversible Watermarking Scheme for Information Health Systems, Breakthrough Visibility Parameters, Reversible Watermarking Scheme for Protecting Authenticity and Integrity of Medical Images, Hybrid Cryptography-Based Watermarking Technique, Spiral Order Technique Reversible Watermarking, Hybrid Estimate Reversible Watermarking, and Reversible Watermarking for Medical Video. The authors stated that the contradictory performance parameters in data hiding have led to complications regarding the widespread acceptance of reversible watermarking techniques.

In the survey by Evsutin and Dzhanashia [67], the most critical features of watermarking schemes, attack types, and performance measures to assist in steering digital image watermarking were initially reviewed. Then, the classification of digital watermarking schemes by their robustness according to various classes of attacks was carried out. The authors stated that robustness towards a variety of attacks can be achieved in different ways, and a scheme that is robust against one kind of attack might not be robust to another type. They pointed out some considerations in order to achieve robustness according to different types of attacks. First, using redundant watermarks in the host image or utilizing scrambling techniques such as the Arnold transform can help with cropping and noise attacks. Second, the combination of different domains, such as DFT, DWT, or SVD, can improve the robustness. Third, using template techniques is beneficial for noise attacks, while paying attention to synchronization mechanisms can increase robustness against geometric attacks.

In their survey, Rakhmawati et al. [102] discussed the specifications and rules related to self-embedding fragile watermarking. They reviewed the methods used for watermark selection, generation, and embedding, as well as tamper detection, localization, and recovery algorithms. As mentioned in this paper, two watermarks are required for the authentication and recovery of data. Some bits of the cover are used for authentication, as well as tamper detection and localization, while other bits are considered for tamper recovery. The authors compared watermark embedding in terms of both spatial and transform domain techniques. They concluded that the imperceptibility obtained through spatial domain techniques is higher than that for transform domain techniques, as watermark bits are included into the image directly.

V. SURVEYS DEVOTED TO DIGITAL WATERMARKING IN VIDEO

A video stream consists of a three-dimensional signal with two dimensions in space and one dimension in time. One approach for video watermarking is to use the image watermarking techniques discussed in the previous section. However, video watermarking introduces several issues not present in image watermarking. As adjacent frames have not only high correlation but also a large amount of spatial and temporal redundancy, the watermarking algorithm must consider such characteristics in terms of capacity, robustness, and imperceptibility. In addition, video signals are highly susceptible to piracy attacks, including frame averaging, frame dropping, frame swapping, statistical analysis, and so on.

Certain types of sensitivities of the HVS have been reported in the literature. For example, the sensitivities of the human eye are not equal to RGB color channels, and in the luminance channel, the sensitivity to noise is low in regions with high brightness. It is also well-known that the changes in texture regions which retain many high-frequency components are less sensitive than those in flat regions. Based on some metrics of the HVS, such as luminance masking, spatial masking, and contrast masking, it is necessary to determine the location and strength of the watermark. In the case of a video stream, motion sensitivity makes it difficult to perceive the changes in moving blocks. If a block in a given frame is moved/changed abruptly from the previous frames, it is easy to embed the watermark imperceptibly.

Bhattacharya et al. [68] overviewed different types of watermarking methodologies depending on the embedding domain, cover media, and application area. The authors focused on video watermarking techniques, especially common video watermarking techniques based on the SSM technique. The techniques included 8 × 8 DCT components in video frames with drift compensation, code division multiple access (CDMA) modulation, changing the variable length code, region-based energy modification, and temporal properties in groups of frames. The authors also discussed the applicability to different video formats, such as MPEG-2 and H.264/AVC. A comparative analysis was summarized in terms of robustness, reliability, imperceptibility, practicality, time complexity, and synchronization recovery. The applicability of several watermarking techniques in H.264/AVC was also discussed by describing a block diagram of their operation.

Jayamalar and Radha [69] summarized the technical requirements for video watermarking and classification of threat models according to potential attacks. The requirements for robustness included signal processing operation, geometric distortions, and subterfuge attacks. Lossy compression, digital-to-analog and analog-to-digital conversion, re-sampling, re-quantization and signal enhancement of image contrast and color were regarded as common signal processing operations. From a security point of view, Kerckhoff's assumption was addressed to consider the importance of the watermarking algorithm to be public and depending solely on the choice of a secret key. Watermarking techniques in the spatial domain, frequency domain, and coding domain in MPEG compression were reviewed in the survey. Furthermore, the LSB technique and its improved methods were introduced in the spatial domain, which have advantages for real-time applications due to their low computational complexities. The consideration of the temporal synchronization was addressed in terms of vulnerability to video processing and multiple frame collusion. Some watermarking methods based on DCT, DFT, and DWT were reviewed as frequency domain approaches and embedding methods for the MPEG compressed domain were presented as extensions of DCT domain methods. The authors enumerated some typical attacks, such as adding distortions, filtering, cropping, lossy compression, geometric modifications, and multiple watermarking.

Chang et al. [70] reviewed video watermarking in terms of the characteristics of the embedding operations. Considering video formats, there are three different solutions for embedding/extracting watermarks. One is in the uncompressed domain, another is during the encoding/decoding process, and the third is in the encoded bitstream. For uncompressed video, watermarking techniques in the spatial and frequency domains, such as DCT, DFT, and DWT, were summarized. For compressed video, watermarking techniques using DCT coefficients and their variable length codes (VLCs), motion vectors, and inter-frame correlations were surveyed. In the early stages of video watermarking, the discussions focused on computational cost and watermark strength. The authors presented their conclusion that, as the relevant literature has matured, the synchronization of watermarks has become the focus of much research.

Shelke and Chatur [71] focused on robust video watermarking schemes and summarized the requirements of imperceptibility, robustness, capacity, fidelity, and computational cost. Possible applications were listed as copyright protection, source tracking, broadcast monitoring, fingerprinting, authentication, copyright protection, tamper proof, and copy control. Attack methods were summarized as lossy compression, addition of noise, filtering operation, row/column removal, cropping, rescaling, and rotation. Robust video watermarking techniques fall into three categories: spatial domain methods, transformation domain methods, and methods based on MPEG coding structures. In the spatial domain, correlation-based and LSB modulation methods were considered. Regarding transformation domain methods (e.g., DCT, DFT, DWT, SVD, and PCA), generic operations for embedding/extracting watermarks were introduced, and suitable elements were discussed in terms of imperceptibility and robustness.

Sethuraman and Srinivasan [72] described the essential tools of watermarking compared to cryptographic tools and explained the basic requirements for robustness, perceptibility, the confidentiality of the hidden message, and complexity. A classification of watermarking schemes was summarized with respect to the embedding/extraction domain, type of multimedia content, perceptibility, and application. From a cryptographic point of view, symmetric and asymmetric (public-key) cryptosystems were explained, though a detailed analysis of security aspects and requirements were not provided. They also reviewed techniques for embedding watermarks into DFT, DCT, DWT, and SVD components using the HVS. An interesting aspect of this survey is that it considered 3D signals in the video stream and the application of frequency transformations. The authors compared 2D and 3D DWTs and discussed their advantages and disadvantages.

Pathak and Jain [73] presented a report on feature-based watermarking techniques. The authors briefly characterized the embedding operation with respect to the input and output. For the embedding algorithm, the inputs are the original content, watermark, and secret key. The authors only considered a non-blind extraction algorithm which receives the watermarked and original content for the extraction of a watermark. Depending on RGB color features, hue saturation values, edges, and corner points, the different of the characteristics were explained by providing examples. It was concluded that in order to achieve high robustness, it is essential to develop new hybrid techniques to deal with a variety of attacks. In the robustness study, temporal attacks were also considered, due to the high reliance on frame sequences in many robust video watermarking schemes.

Asikuzzaman and Pickering [74] reviewed various types of video watermarking algorithms. They illustrated the unauthorized distribution of copyright-protected video, assuming that a copy of a movie file is captured from a movie theater, uploaded to the Internet, and redistributed through an internet service provider. A general classification was presented in terms of applications, including copyright protection, broadcast monitoring, copy control, authentication, fingerprinting, online location, and content filtering. The challenges faced by video watermarking schemes were characterized in terms of the requirements of imperceptibility, payload, blind detection, robustness, and security. Robustness denotes the resistance of a watermark to blind non-targeted modifications or the common media operations of regular users. The considered attacks included signal processing attacks, geometric attacks, and temporal synchronization attacks such as frame dropping, insertion, swapping, and frame rate conversion. The differences in terms of robustness and security were explicitly explained in this survey. Robustness-related attacks are common media operations that affect both video and watermark signals, where affecting the watermark might not be the main goal. On the other hand, security-related attacks are intended to gain knowledge about an embedding and/or extraction system in order to remove a watermark. Watermarking techniques in the compression domain were presented separately for different video formats such as MPEG-2, MPEG-4, H.264/AVC, and H.265, and comprehensive descriptions were provided for a review of state-of-the-art techniques. In the spatial domain, LSB modifications using pixel and block-wise methods, statistical methods such as the patchwork technique, and feature point-based methods were discussed. In the transform domain, the authors explained hybrid methods such as SVD combined with DFT, DCT, and DWT components, HVS-based approaches, and those based on feature points, such as KAZE [104]. They also described 3D watermarking techniques for protecting video streams with 3D geometric structures.

In the survey of Yu et al. [75], information hiding techniques were classified into four types: covert communication, steganography, anonymous communication, and digital watermark. A comprehensive survey was conducted to understand the divergence between steganography and watermarking. Among the information hiding techniques, this paper emphasized watermarking and focused on video watermarking. It also focused on the application of video watermarking for copyright protection. The framework of a general robust watermarking algorithm was illustrated and the algorithm is discussed in terms of various types of video formats and embedding domains. The trade-off among imperceptibility, robustness, and capacity was explained, and two commonly used quantitative evaluation metrics – namely, PSNR and Structural Similarity (SSIM) [105]– were reviewed to consider the imperceptibility. In addition to the relationships among the three terms mentioned above, real-time performance was emphasized as an important requirement for video watermarking schemes. In terms of spatial domain methods, the authors reviewed the blind and semi-blind algorithms based on LSB modification, Speed-Up Robust Features (SURF), and YUV color space components. As for frequency domain methods, DCT- and DWT-based methods were reviewed. The watermark information is often embedded into the singular values matrix to obtain good imperceptibility, which is usually adopted together with DWT in robust video watermarking methods. A quantitative comparison of the imperceptibility and robustness of several typical algorithms in the spatial, DCT, and DWT domains with SVD was summarized. The authors claimed that research on video watermarking algorithms is mature regarding the MPEG-2 and MPEG-4 formats, and the current trend is moving toward new video coding standards such as HEVC.

Artru et al. [76] reviewed the majority of watermarking algorithms and classified them according to their characteristics and evaluation indicators. The most commonly used media for watermarking are images, but videos, texts, audio files or software are also considered. The authors focused on watermarks applied to video streams as carrier signals, and discussed the applications and techniques used in the field of video watermarking. From a communication point of view, the watermarking system can be regarded as signal transmission over a noisy channel, where the signal may be distorted by unintentional (natural) and intentional (malicious) noise introduced by attackers. At the receiver side, the authors considered two different meanings for the information extractor. One is the presence of a watermark, and the other is extraction of watermark information represented by a bit sequence. Methods for the evaluation of medium fidelity or distortion include the Hamming distance, Bit Error Rate (BER), Mean Square Error (MSE), PSNR, and Normalized Correlation Coefficient (NCC). The watermark fidelity can be determined using four measurement primitives: true positive, true negative, false positive, and false negative. The main way to represent these primitives is by use of the Receiver Operating Characteristics (ROC) curve obtained by plotting the True Positive Rate (TPR) (or sensitivity) against the False Positive Rate (FPR) (or specificity). The authors called the property defined by the prior information needed by the detector to retrieve the wanted data from the carrier channel the blindness of a watermark. Private and semi-private watermarks are related to blind and semi-blind methods, respectively, while public watermarking enables users to detect or extract the watermark signal only when a secret key is available. Public key watermarking describes a system where anyone with access to the watermarked signal can observe the embedded information, but only one person could have embedded it and no-one can remove it. According to the robustness requirements, the methods were classified into fragile, semi-fragile, and robust watermarking methods. The characteristics are related to the considered applications, such as copyright protection, access control, and tamper detection. Threat models were also classified into several categories, based on the assumed capabilities of attackers in a realistic environment.

Hummadi and Hassan [77] surveyed robust video watermarking techniques. Their survey summarized the key aspects of video watermark design, such as insensitivity and robustness, and described potential applications such as fingerprinting, copy control, broadcast monitoring, video authentication, and copyright protection. The common attacks in video watermarking were listed as simple, detection disabling, ambiguity, and removal attacks. The domain for embedding the watermark was roughly classified into spatial and transform domains. The paper also covered the areas of

LSB, DFT, DCT, DWT, and SVD as embedding methods for video watermarking, discussing their advantages and limitations in terms of robustness.

VI. SURVEYS DEVOTED TO DIGITAL WATERMARKING IN AUDIO AND SPEECH

Generally, the human auditory system (HAS) is much more sensitive than the HVS, and audio signals are represented by fewer samples from time intervals. Hence, compared with image and video watermarking, it is a more challenging task to achieve imperceptibility in audio watermarking.

In audio watermarking schemes, the perceptual properties of the HAS are considered when embedding a watermark into audio sequences. The perceptual similarity between the original and watermarked audio sequence must be guaranteed, as one of the evaluation metrics. There are three conflicting requirements in audio watermarking algorithms: the payload, perceptual quality, and robustness. If the payload is increased, the number of distortions is increased and, hence, the perceptual quality decreases. Otherwise, the robustness of the embedded watermark will be decreased. Hence, the trade-off among these three requirements is an essential factor for evaluation. Many audio watermarking methods have been developed by employing various techniques, such as SSM [106], echo-hiding [107], support vector regression [108], and patchwork [1].

A survey paper on audio watermarking techniques and their requirements has been presented by Alsalami and Al-Akaidi [78]. The authors classified information hiding techniques into steganography and watermarking, according to the confidentiality and robustness of the hidden message. The authors focused on a watermarking system consisting of three modules: watermark signal generation, embedding, and detection. Relevant application areas include copyright protection, fingerprinting, content authentication, copy protection, and broadcast monitoring. The following characteristics of watermarking were discussed: embedding effect, fidelity, data payload, blind or informed detector, and false positive rate. Depending on the domain in which the watermark is embedded, audio watermarking techniques fall into four categories: frequency domain, time domain, compression domain, and wavelet domain. In frequency domain watermarking techniques, the speech masking characteristics of the HAS in the frequency domain are used; the SSM technique is an example of such a method. In time domain watermarking techniques, the watermark is embedded directly into the audio signal. However, it is difficult to maintain the fidelity and robustness of the watermark. To achieve high robustness, temporal and frequency masking approaches were introduced to illustrate the HAS-based masking effects. The authors reviewed watermarking techniques for MPEG audio bitstreams, including those that use HAS and psychoacoustic models, and those that use cyclic redundancy codes (CRCs) and the Data Encryption Standard (DES) algorithm to enhance the confidentiality and security of the hidden message. DWT domain techniques use the subbands of the signal to select a suitable region for watermarking. Benchmarks for evaluating audio watermarking were also discussed.

Chauhan and Rizvi [79] categorized information hiding technologies into the steganography and copyright masking branches. In the robust solutions branch, watermarking and fingerprinting were exemplified as different categories of information hiding technologies. The survey mentioned audio watermarking techniques and requirements. Requirements were defined in terms of perceived transparency, robustness, security, data rate, verification and reliability. The embedding and extraction processes were depicted in that paper [79] using the block diagram, and the extraction procedure is a blind scenario where the original cover object is not presented. Applications of audio watermarking include copyright protection and proof of ownership, tamper detection, copy protection, fingerprinting, broadcast monitoring, and information carrying. Audio watermarking techniques were divided into four categories: time domain, transformation domain, compression domain, and composite domain. In the time domain, the paper reviewed schemes based on LSB substitution and modification, phase coding, echo hiding, and SSM techniques. In the transformation domain, the Quantization Index Modulation (QIM) [109] and its variants, mean quantization, dither modulation, and vector quantization were described. In the compression domain, MPEG bitstreams were covered. Furthermore, schemes that combine several of the above areas are considered. A Venn diagram visually representing the relationships between various methods, according to the classification of audio watermarking techniques, was presented.

Nematollahi and Al-Haddad [80] provided an overview of speech watermarking techniques, in terms of application, capacity, robustness, and insensitivity. In terms of some signal characteristics, including bandwidth, voice/non-voice, and production model, digital speech signals differ from audio, music, and other signals. As with other multimedia content, three categories were considered: robustness against communication channel attacks, the vulnerability of the embedded watermark, and the insensitivity of the extraction module. For speech watermarking techniques, various methods were reported, including auditory masking, phase modulation, quantization, transformation, and parametric modeling. Based on the HAS, the advantages of frequency and time masking were explained by illustrating their effects in the time and frequency domains. In phase modulation, autoregressive (AR) phase models, DFT phase models including short-term Fourier transform (STFT) phase modulation, and wrapped orthogonal transforms and their variants were discussed. In cooperation with the VQ, some extensive approaches were investigated for the improvement of robustness. Due to the logarithmic order of speech signals in the frequency domain, the QIM with rational dither modulation (RDM) method and its variants were discussed as promising approaches providing higher robustness. The use of Mel-Frequency Cepstrum Coefficients (MFCC) was also explained. In addition, several speech watermarking techniques were evaluated, in terms of their capacity and imperceptibility. Moreover, speech databases and speech corpora were summarized in tables. The importance of their application in authentication was only mentioned.

Bajpai and Kaur [81] presented a literature survey paper on audio watermarking techniques in three domains: the spatial domain, frequency domain, and hybrid domain. The listed applications of audio watermarking included copyright protection, secret communication, authentication, e-commerce, e-voting, and broadcast monitoring, and the basic requirements of information hiding in audio file is named as a magic triangle composed of robustness, imperceptibility, and payload. In the spatial domain, watermark embedding algorithms based on LSB correction, SSM techniques, and the HAS were also studied. In the case of frequency transformations, machine learning methods such as SVM and artificial neural networks (ANNs) were combined with basic detection algorithm. The combinations of more than two frequency transformations and SVD method were compared in terms of imperceptibility and robustness against MP3 compression, requantization, resampling, cropping, and noise addition. In the comparison, robustness was evaluated using the NCC metric, which is valid only for the case of detection of watermarks and, hence, can be used to verify the presence of a watermark. Various existing audio watermarking algorithms were compared, in terms of their nature, capacity, imperceptibility, and robustness, though detailed experimental conditions were not provided.

Nayyar et al. [82] published an overview paper regarding audio watermarking techniques based on frequency transformations such as DCT, DWT, SVD, and DFT. The authors mentioned, in the survey, that the desired features of audio watermarking algorithms are signal processing properties including imperceptibility and robustness against intentional manipulations, security properties such as the secrecy of algorithm in terms of cryptographic perspective, and general properties such as real-time processing capability. Potential applications of audio watermarking include copyright protection, monitoring, fingerprinting, indication of content manipulation, and information carrying. From a review of several papers, they described the fundamentals, advantages, and limitations of signal processing techniques. The use of chaotic transforms for scrambling was also mentioned, although it was related to digital images. The authors summarized with a comparison of the advantages and disadvantages of some works without experimental evaluation.

A comprehensive survey of audio watermarking techniques has been published by Hua et al. [83]. The authors provided a systematic review of audio watermarking techniques published over the past twenty years, identified existing problems and difficulties, and discussed potential research directions and strategies for more advanced solutions. The authors categorized watermarking techniques and related works in terms of information theory, methodology, and counter-measures, then discussed the similarities, differences, and key features of each category. An a priori probability density function (PDF) of the host signal (usually in the transform domain) is typically assumed to be available –for example, a generalized Gaussian distribution (GGD) in the information-theoretical analysis. The watermarking methods were also categorized into time domain and transform domain methods, and they were further divided into time-aligned and echo-based methods in the time domain and SSM, QIM, and patchwork techniques in the transform domain. In the application of audio watermarking for copyright protection, the system designer usually places priority on the robustness of the system against intentional and unintentional attacks. The effectiveness of an audio watermarking system can be characterized by several performance criteria, such as imperceptibility, robustness, security, capacity, computational complexity, and so on. The authors also summarized existing attacks and applications, including a series of U.S. patents related to audio watermarking techniques. In this survey, watermarking algorithms were categorized into five primitives –time-aligned, echo-based, SSM, QIM, and patchwork– and a comparative evaluation of their robustness was provided. The trade-off between robustness and imperceptibility was also discussed for these primitives. Although watermarking techniques have been studied for more than two decades, it is still challenging to design algorithms that are robust against attacks.

Jain et al. [84] presented a brief overview of audio watermarking techniques and their challenging aspects regarding robustness, imperceptibility, and capacity. Unfortunately, the details and further research directions were not provided. Their classification of watermarking divided the working domain into spatial and transform domains, watermark extraction into blind and non-blind detection, and watermark robustness and vulnerability. The authors briefly reviewed basic watermarking algorithms, including LSB modification, the SSM technique, perceptual masking in the time domain, phase modulation methods, and replica modulation in the transform domain. Although considerations of the performance of audio watermarking were not provided in this survey, the authors summarized the classification of different audio watermarking techniques in a table.

Lalitha et al. [85] provided an overview of audio watermarking methods and the requirements that must be satisfied for robustness against attacks. The key properties of audio watermarking were enumerated as imperceptibility, robustness, payload, and security. Relevant applications included copyright protection, monitoring, fingerprinting, indication of content manipulation, information carrying, and access control. As there is a trade-off between imperceptibility, robustness, and payload, audio watermarking schemes should be developed to achieve a proper equilibrium. To evaluate the robustness, the authors explained the importance of attack models such as noise, filter, time stretch, pitch shift, conversion, ambience, dynamics, and sample perturbations. Basic embedding and extraction functions were formulated and classified into blind and semi-blind techniques, according to the input to the functions. Examples included LSB coding, echo hiding, SSM technique, QIM, and patchwork. Robustness and confidentiality evaluation criteria were also described, in terms of detection rate, accuracy, NCC, BER, and PSNR. A comprehensive review of blind and semi-blind algorithms proposed over the last two decades is presented. The authors claimed the need for a standard database for evaluation.

VII. GENERAL MULTIMEDIA CONTENT SURVEYS

Apart from those surveys that attempted to capture specific subfields of digital watermarking limited to certain capabilities of the watermark (e.g., robust or fragile) or a chosen content (e.g., image, video, or audio), there are also papers in the literature that have tried to present a broader perspective. Below, we characterize the content and main contributions of such papers.

Lee and Jung [86] prepared a short general introduction to the topic of multimedia watermarking. The authors explain the idea of embedding and detecting watermarks and the necessity of a similarity measure between the original and extracted watermarks. Invisibility, robustness, and security were indicated as the most essential properties of watermarks. Then, classification criteria for six aspects of watermarking methods were provided, and the main characteristics of each category were described. According to [86], watermarking methods can be categorized in terms of the type of media in which the watermark is embedded, the perceptibility of the watermark, robustness against attacks, the type of inserted watermark (noise or image), processing method (in spatial or frequency domain), and the data necessary for extracting the watermark (private, semi-private, or public watermarking, depending on the necessity of original media, the secret key used in the embedding process, and the embedded watermark).

In general, Lee and Jung [86] did not present an actual classification of existing methods but, rather, a framework with a set of classification criteria by which existing watermarking methods could be assigned. Specific algorithms are usually not assigned to particular categories; however, there is one exception to this rule. The authors also categorized 18 existing approaches from the literature in terms of processing methods. Algorithms that use spatial domain processing were divided into two groups, depending on the subject which is modified during watermark insertion (pixels or blocks), while methods in the frequency domain were categorized on the basis of the transformation type (DCT, DFT, and wavelet transform). Although the paper [86] provided a general introduction to the topic of watermarking and mentioned various types of media content, it was mainly described from the point of view of image watermarking (which was the major research area at the time of writing the paper, as the authors claim) and copyright protection is presented as the main application of watermarking methods. The idea and necessity of using watermarks were explained in the context of copyright protection, and other applications were only mentioned among the properties of the given subclasses in the defined framework. Similarly, possible attacks on watermarks were not directly addressed in this survey, but the paper only indicated the types of attacks to which a given category of methods is robust or weak. The main findings of that survey were the most commonly used properties in each category; that is, images as target content, invisibility and robustness of watermarks, noise as an inserted watermark type, and processing using transform domain (particularly, methods based on DCT). At the time of writing (2001), the authors also observed the growing interest in wavelet transform-based solutions and the need for development of watermarking techniques suited for audio and video.

Nikolaidis et al. [13] presented a general survey dedicated to imperceptible watermarking application scenarios. The authors focused on the description of various scenarios, their characteristics, and the desirable properties of watermarking schemes. Their aim was to classify watermarking applications and relate them to the characteristics of schemes, in order to improve the reusability of existing schemes for other applications or to facilitate the creation of new schemes for a given application [13]. Different watermarking applications were classified into three major groups: intellectual property rights (IPR) protection, content verification, and information hiding. They also indicated four general types of attacks against watermarks: removal, presentation, protocol, and legal attacks. The last category, unlike the previous ones, refers to attempts made to prevent the perception of watermarking as a reliable proof of ownership by the law [13].

Watermarking application scenarios were described by Nikolaidis et al. [13] in terms of their main objectives, involved subjects, attacks that can be expected in a given situation, the type of watermarking scheme (zero-, single-, or multiple-bit), usage of private or public keys, and other properties of watermarks. Six general cases with subsequent various sub-scenarios were considered by the authors: copyright protection (without distribution network, with distribution network with and without a trusted third party), broadcast monitoring (piracy tracking and people metering), fingerprinting, authentication and integrity checking (either verified by the owner of the multimedia content or by its user), usage control, and information hiding (with a public, private, or hidden side-channel). Although some examples of usage in the context of a given media type (image, video) were mentioned, they seemed to intend to explain the scenarios better to the reader, not to narrow down the scope of the survey or to assign specific applications to particular multimedia. The described scenarios were so generic that they could be applied to any form of multimedia.

Cox and Miller [3] highlighted the importance of perceptual modeling; that is, placing a watermark in perceptually meaningful regions of the digital content. Their survey focused on applications related to copyright control problems, and the question of content authentication-related watermarking applications was intentionally not discussed. The authors explained the preferable properties of watermarks used for copyright control purposes and indicated that, in the case of content authentication purposes, the required characteristics of the watermark could differ. According to [3], a watermark used for copyright control needs to be hard to notice, robust to typical signal transformations, and resistant to tampering. Intentional malicious attacks and watermark modifications caused by common signal processing were distinguished in this paper and described according to two different properties: tamper resistance and robustness. Moreover, Cox and Miller [3] emphasized that completely imperceptible watermarks cannot be robust, as simple processing such as lossy compression based on the removal of elements invisible to humans could destroy the watermark; for this reason they chose to investigate the problem of perceptual modeling. The other essential features of watermarks indicated in [3] were a sufficient bit rate, the ability to modify or add multiple watermarks, and decoder scalability. The authors also presented a mathematical representation of the watermarking process for image content, taking into account two characteristics: whether the watermark is stored in perceptually important parts of the image and whether it is dependent on the original image (i.e., the linearity of the insertion).

The review of existing approaches performed by Cox and Miller [3] surveyed 22 publications and covered various types of media, although the majority of the examined papers focused on images (17 publications). The other mentioned approaches referred to watermarking of audio, text (2 papers in each case), video, and CDs (1 paper in each case). In general, the idea of each method was presented, and its various aspects, such as robustness and tamper resistance, decoding process, dependency on the original content, and usage of perceptually significant regions of the content, were discussed. The authors concluded that, for copyright protection purposes, it is essential to place the watermark in perceptually meaningful components of multimedia and to use a non-linear insertion process. As the watermark should also be difficult to notice by humans, the signal-to-noise ratio (where the signal is the watermark and the noise is the original content) should be much less than one.

The extensive survey presented by Menendez-Ortiz et al. [103] analyzed the robustness of invisible reversible watermarking schemes from various points of view. Unlike conventional watermarking, in which the multimedia content retrieved after the decoding process is very similar to the original one, but not exactly the same, reversible watermarking allows for reconstruction of the original content without any loss, which may be crucial in military or medical applications. However, for a long time, reversible schemes were fragile by default, such that the watermark and the content could have been decoded perfectly only if they were transmitted through a noise-free channel where no attacks could occur. The authors analyzed 134 different approaches, and categorized and compared them. Menendez-Ortiz et al. classified reversible watermarking approaches as fragile or robust, and the latter category was further subdivided into robust and semi-fragile schemes. The idea behind this classification is that fragile watermarks cannot deal with any attacks, while semi-fragile approaches can survive unintentional attacks, such as compression, and robust methods can even survive some intentional attacks. It is important that robust or semi-fragile reversible watermarking allows for retrieval of both the content and watermark only when there are no attacks. On the other hand, fragile reversible watermarking methods can focus on improving the payload capacity and imperceptibility.

The fragile methods described in [103] deal with image and audio content and belong to two major groups: error expansion-based and histogram shifting-based methods. The latter category contains only approaches designed for images, while the former consists of four types of approaches: amplitude expansion (proposed only for audio), difference expansion, interpolation error expansion, and prediction error expansion. All fragile techniques were briefly explained and compared, in terms of content type (image or audio), domain (spatial, frequency, or temporal), payload, and perceptual similarity between the original and watermarked content (PSNR for images and signal to noise ratio, SNR, and objective difference grade, ODG, for audio). Semi-fragile and robust methods were similarly described in terms of multimedia type, domain, and the attacks that they can survive. The analyzed semi-fragile schemes were proposed only for images and were resistant to only JPEG or JPEG2000 compression attacks. The majority of robust approaches presented in [103] were also designed for images, and there was only one paper proposing a solution for audio. All of them can survive various attacks, which were listed in detail.

In addition to reversible watermarking, Menendez-Ortiz et al. [103] included self-recovery watermarking and a combination of two different approaches –namely, reversible watermarking with watermark and signal robustness– in their survey. Self-recovery watermarking differs from reversible watermarking, in that it allows for reconstruction of the host signal (i.e., multimedia content) but, in this case, the payload does not contain useful watermarks, only information necessary to recover the original content (e.g., its compressed version). Reconstruction of a host signal can be approximate (covering the vast majority of solutions) or perfect. The survey compared the methods for self-recovery watermarking in terms of domain, host type, ability for tamper detection and correction, and attacks that the given method can resist. Approximate reconstruction solutions were the only category in the survey that contained methods suitable for video (6 papers), the most common content type was images (36 papers), and approaches for audio were rare (4 papers). The perfect reconstruction category contained much fewer techniques, as there were only four papers regarding images and one paper suited for audio, which was claimed to be the first solution proposed for this media type. All of these solutions allow for tamper detection, and only two approaches in the approximate reconstruction category did not allow for tamper correction.

As self-recovery watermarking can preserve the host signal and robust reversible watermarking can preserve the watermark in the presence of attacks, a combination of those two approaches allows for reconstruction of both. This field of study is very new, and only three such concrete methods were described in [103] (two for images and one for audio). The idea behind this approach is that fragile reversible watermarking is embedded into self-recovery watermarking with perfect reconstruction.

Although the survey by Menendez-Ortiz et al. [103] is very detailed, in most cases, it lacked information about the applications in which the given schemes can be used. The authors mentioned that watermarking was originally used for copyright protection purposes but, recently, it has also been used to enrich multimedia with their metadata. They explained the need for reversible watermarking techniques through the example of telemedicine, where the precision of the decoding watermarked content is essential for proper diagnosis. In the case of single approaches, their main applications were mentioned, including integrity verification, authentication (of content), and detection of tampering or tampered regions (although tamper detection and localization were rather described as properties, not as aims of methods), DICOM format of images, and MRI images. Fragile approaches were explained in the most detailed way in [103], and the idea of each method was described with a few sentences. Robust reversible techniques were rather described in terms of their properties, such as domains and attacks or disadvantages (e.g., the addition of noise or low capacity), but the algorithms themselves were presented rarely. In the case of self-recovery watermarking schemes, not all of the analyzed methods were described, with only the most crucial ones mentioned in the text, and the rest were compared in a table. They referred mainly to approximate reconstruction methods, many of which exist. Perfect reconstruction approaches were presented in a bit more detail. Reversible watermarking with the robustness of both watermark and signal was also described very briefly, some theoretical works on this topic were mentioned, and the general idea was explained.

Menendez-Ortiz et al. [103] provided many useful conclusions and future research directions. Their survey demonstrated that, among fragile methods, error expansion-based approaches are the most popular; however, histogram shifting-based techniques (proposed only for images) offer better capacities, provided that multi-level embedding is used. They indicated that it needs to be further investigated whether a better compromise between payload capacity and imperceptibility can be achieved through multi-level histogram shifting methods or multi-bit error expansion solutions. Robust reversible techniques were proposed mainly for images, and there is a gap to fill regarding semi-fragile and robust reversible algorithms for audio and video. Many schemes have been proposed for self-recovery watermarking with approximate reconstruction, covering all types of media; however, again, the vast majority of them are suited for images. Perfect reconstruction approaches are still very rare and, at this time, they can resist only a few types of attacks. Thus, the authors of [103] stated that research should be carried out to increase the scope of attacks to which such methods are robust. The novel combination of reversible and self-recovery watermarking also needs to be further explored, for example, in terms of robustness, transparency, complexity, and security [103]. The authors concluded that all types of robust approaches presented in their paper can resist only non-severe common types of attacks, such that there is a need to extend the types of attacks to which these methods are immune; however, other properties, such as imperceptibility, capacity and security, must also be considered.

Megías et al. [87] focused on fingerprinting schemes used in two types of decentralized distribution cases: peer-to-peer networks and broadcasting. They provided a general introduction to the watermarking topic, including descriptions of different applications, properties, and types of watermarks. They also pointed out the important features of fingerprinting, which is one of the applications of watermarking used to ensure copyright protection. According to them, piracy tracing, asymmetry, anonymity, collusion resistance, dispute resolution, non-repudiation, and unlinkability should be taken into account in every fingerprinting method [87]. Although those properties are usually achieved in traditional unicast distribution scenarios, where one merchant distributes the content to one buyer, such scenarios are not suitable with respect to current needs. Megías et al. investigated those properties in the case of more contemporary distribution methods. The different types of entities which appear in the fingerprinting process, technologies used in fingerprinting protocols, and attacks that have to be faced in case of fingerprinting were also explained. Moreover, for real-time distribution, fingerprinting schemes also need to be robust to lossy compression, as this mechanism is usually used to increase the efficiency of distribution.

Megías et al. [87] surveyed 14 papers describing fingerprinting methods in peer-to-peer distribution scenarios and 10 papers regarding fingerprinting for broadcasting. Each of those methods was explained, and the techniques used in each case were indicated. The methods were all critically reviewed, and their disadvantages were presented. Where appropriate, advantages or improvements in comparison to previous solutions from the literature were also described. However, in the vast majority of cases, the multimedia type that the given method is suited to was not clearly stated. In particular, the content type was only directly mentioned in two cases. The rest of the solutions seemed to be of general purpose. Moreover, Megías et al. focused mostly on the possibility of applying the given method in the streaming scenario, not in the distribution of separate whole files. Regarding possible attacks, the authors mostly examined the robustness of each approach against collusion. Copyright protection through piracy tracing, user privacy, asymmetry of the scheme, computational cost, and the efficiency of the approaches were also discussed. In addition to the deep review of fingerprinting schemes for peer-to-peer and broadcasting applications, Megías et al. [87] also analyzed two decentralized tracing protocols and a few supplementary techniques, which can be used to distinguish between guilty users who illegally redistributed multimedia content and innocent ones or to ensure better efficiency of the protocol.

The fingerprinting approaches presented in [87] were deeply discussed and compared. Although almost all of them offer piracy tracing, privacy (which is revocable when there is a need to indicate the personal data of the illegal content redistributor) and anonymity are not always provided. Robustness to collusion attacks is achieved by using collusion-resistant fingerprinting codes, but not all approaches are immune to this type of attack. Another threat considered by Megías et al. was the communication attack, which is also addressed in many solutions. Some peer-to-peer approaches can also be used (or adjusted) for real-time broadcasting. All of the methods analyzed in [87] were compared in a tabular form, in terms of applicability to peer-to-peer and broadcasting scenarios, collusion resistance, buyer frameproofness, and privacy.

The most promising solutions for fingerprinting in decentralized distribution scenarios were also indicated. As they belong to different domains and have different properties, Megías et al. [87] suggested that a combination of these different solutions could be used in order to achieve a more general approach, which is worth investigating in the future. Moreover, the biggest challenge identified in this survey was the proposal of fingerprinting methods that are suitable for broadcasting live events, which requires great robustness against various attacks, as well as buyer frameproofness, traceability, and revocable privacy [87].

The overview presented by Langelaar et al. [4] focused on the explanation of some watermarking algorithms and their possible improvements, rather than on some specific applications of watermarking or addressing particular types of attacks. Although the paper was dedicated to images and video, the vast majority of approaches presented there were suited for images, and only a few were designed to deal with videos; although some of the described methods can be applied to both images and videos. They also mentioned the question of audio watermarking, but none of the algorithms described were designed particularly for audio. Langelaar et al. [4] described typical watermarking applications, including copyright protection, fingerprinting, copy protection, broadcast monitoring, data authentication (defined as checking data authenticity, informing whether the content was modified, and providing localization of changes), indexing, medical safety (e.g., to include patient data in medical images), and data hiding. The authors also explained the most important features of watermarking schemes, including transparency, payload, robustness, security, and blindness, and indicated the relationships between them. Despite the fact that various applications and characteristics of watermarks were described in this survey, it did not focus on any specific application and, regarding the features of watermarks, only the methods used to increase the robustness of particular algorithms were discussed in a detail. The problem of attacks on watermarks was also not addressed directly; in the case of some algorithms, methods for avoiding the impact of lossy compression (especially JPEG), filtering, and geometric transformations of the content on the watermark were discussed, but this was rather framed in terms of robustness than in terms of resistance against particular types of attacks.

The authors mainly described watermarking algorithms in a very detailed way. They not only explained how the algorithms work, but also illustrated it with mathematical formulas used in algorithms, examples, and the results of experiments. The algorithms were classified into two major groups: correlation-based and noncorrelation-based. As the authors claimed, the first group of approaches is more commonly used, and, so, they were also more extensively described in the paper. Correlation-based techniques from both spatial and transform domains were presented. First, the basic correlation-based method in the spatial domain used in various algorithms –namely, the addition of pseudorandom noise to the pixels of an image– was explained in detail. Then, its various improvements, increasing the robustness and payload of the watermark, were described. Correlation-based algorithms from DFT, DCT, and DWT transform domains were also explained, as well as the spatial masking technique used to improve the robustness of watermarks. The paper also described methods to increase the robustness of watermarks against lossy compression, filtering, and geometrical transformations (e.g., shifting, scaling, cropping, and rotation). Although the majority of approaches were designed for images (some of them for images and video, but not particularly for video), two correlation-based algorithms for MPEG video watermarking which operate on DCT coefficients were also presented, and the error accumulation problem related to this type of method was explained.

The noncorrelation-based algorithms described by [4] belong to two main groups: those based on LSB modification and based on geometric relations. Six main approaches were explained: LSB modification, parity bit modification, DCT coefficient ordering, the differential energy watermark (DEW) algorithm, salient-point modification, and fractal-based watermarking [4]. Some of these approaches, especially parity bit modification and the DEW method, can be used not only for images but also for compressed video in MPEG format. The authors concluded their paper with a list of other approaches which had been identified as promising during the time of writing the paper (2000) and three main areas for future research. Research on watermarks dependent on the multimedia content (instead of general algorithms) to prevent copy attacks, watermarking of low bit-rate video, and the development of an international benchmarking system that could allow users to compare various watermarking schemes and choose the one that best suits their needs were indicated in [4] as the most important aspects requiring further work, at the time of writing. The authors did not provide any specific comparison of the presented algorithms, but the paper itself was written in an incremental way, where subsequent algorithms extend and improve upon the previous ones. The rationale for such improvements was explained.

Agarwal et al. [88] surveyed 49 watermarking algorithms. The vast majority of the analyzed schemes were suited for images (37 papers), including medical images (7 papers) and 2D barcodes (2 papers). The other described methods proposed watermarking approaches for 3D objects (3 papers), as well as video, audio, electroencephalography (EEG) data, datasets, relational data, social network data, identity cards, and data mining applications (1 paper in each case). The authors focused on robust and imperceptible solutions, but some fragile algorithms were also indicated in their work. The analyzed algorithms came from both spatial and transform domains.

In their general introduction to watermarking, Agarwal et al. [88] defined three types of watermark systems (blind, non-blind, and semi-blind), mentioned several crucial characteristics of watermarks (i.e., robustness, security, computational cost, data payload, tamper resistance, key restriction, fragility, embedding capacity, imperceptibility, and false positive rate), and defined eight types of applications of watermarking (i.e., copyright protection, broadcast monitoring, fingerprinting, medical applications, electronic voting systems, remote education, chip and hardware protection, and secure data on the cloud) and nine types of attacks on watermarks (i.e., active, passive, forgery, collusion, simple, ambiguity, cryptographic, removal, and geometric attacks), as well as the relationship between the characteristics and applications of watermarks. Various other types of watermark applications than those defined above were mentioned in different parts of the article (in the text, figure, and table).

Tabular comparison of the analyzed methods seemed to be the biggest advantage of that survey [88]. Watermarking schemes were compared in terms of the author's objectives, type of watermark system, techniques used in the given method, results of the algorithm, size of the cover object, and of the watermark; some remarks on each solution were also presented. Robustness, imperceptibility, security, embedding capacity, accuracy, efficiency, transparency, distortions, payload, computational complexity, BER, and preservation of image quality were also indicated for each analyzed watermarking scheme [88]. Finally, the articles using the given techniques were indicated, and a matrix presenting the distribution of watermark size with regard to image size was provided. Almost all of the surveyed algorithms were described in a few sentences, mainly in terms of the used techniques, features of the scheme, and comparison of the given algorithm with other existing approaches in the literature; the type of cover media was also sometimes indicated.

The authors concluded that the most important features of the watermarks identified in the surveyed papers were robustness, imperceptibility, security, and capacity [88]. Eight techniques used to increase the robustness of methods were identified. It turned out that majority of schemes use cover images and watermarks of variable size, and among those of fixed size, the most common image size was 512 × 512 pixels [88]. Agarwal et al. indicated only one potential area of future research; that is, comparison of the performance of watermarking solutions.

Although the survey written by Agarwal et al. [88] seems to provide a valuable contribution in the form of a tabular comparison of various methods at first sight, the general impression of the whole paper is much worse. The text of the paper turns out to be chaotic and inconsistent (e.g., different watermarking applications are in different places in the text and in the figure, some of them are defined, and some are not), and elements at different levels of detail are mixed (e.g., a collusion attack is a case of a removal attack but, in the paper, they are described as completely different categories). Some of the statements seem to be doubtful; for example, transparency and imperceptibility are mentioned as two different characteristics of watermarks and, moreover, sometimes the same analyzed paper is claimed to deal with imperceptibility but not with transparency, or vice versa; similarly, embedding capacity and payload are treated as two independent features of watermarking schemes, and in the case of some surveyed algorithms, one of them is claimed to be considered, while the other seems to not be covered; on the other hand, robustness and fragility are mentioned as two different characteristics of watermarks while, in fact, they both represent two opposite values of the same property. Some definitions do not explain anything, but describe some general facts. The conclusions after the performed survey were very short and straightforward. Not all abbreviations were defined and, so, it is sometimes impossible to understand what the authors meant. Last, but not least, there were many editing and language errors, typos, and mistakes (for instance, a lack of textual description for one of the analyzed papers). This all leads to the conclusion that the text of the paper is not only hard to read and understand but may also be of doubtful quality. As a result, it is unclear whether the survey results can be treated as a trustworthy source of information.

A better (but also not free from errors, omissions, duplications, and some other drawbacks) overview has been performed by Kumar et al. [5]. The authors presented classifications of watermarking based on three aspects: the type of multimedia, characteristics of the approach, and its applications [5]. The features of watermarking algorithms are mapped to particular applications. Limitations and challenges of watermarking are pointed out; however, although they are placed after the description of the surveyed articles, it seems that they are only re-written from various sources and do not result from the performed literature analysis. Performance measures are enumerated (according to their definitions provided by Kumar et al., these measures are for images; however, even in their survey, they are used for various types of multimedia) and attacks on watermarks are classified into four main categories (i.e., removal, geometric, protocol, and cryptographic attacks).

The authors of [5] analyzed 64 papers describing watermarking schemes for six types of cover objects: images (15 papers), video (14 papers), audio (12 papers), text (10 papers), graphics (4 papers), and databases (9 papers). The algorithms belonged to both spatial and transform domains. All analyzed methods were briefly described in terms of the utilized techniques, features of the method, its advantages and disadvantages, and the results of experiments performed by the authors of the given approach. For each type of cover media, they also enumerated which techniques, performance metrics, and applications had been identified as the most common, as well as the characteristics of the proposed schemes and attacks against which the algorithms for the given media type were tested.

However, the most interesting part of the paper [5] is the tables containing comparisons of examined approaches. Methods for each type of media were analyzed separately. Their purposes, techniques used in the algorithms, characteristics of the input, names of performance metrics calculated by the authors of the methods to verify their usefulness, attacks against which the algorithms were tested, and some additional remarks were provided for all schemes surveyed in [5]. Additional comparisons of watermarking schemes were provided in the case of images and video. The domain (spatial, transform or, in one case, dual), robustness, imperceptibility, and blindness of the algorithms were indicated, as well as the other features used to compare methods for a given media type (e.g., capacity, watermark type, and objectives in the case of images and video preprocessing, and message preprocessing in the case of video).

From these tables, it can be found that the majority of algorithms belong to the transform domain and offer robustness and imperceptibility. Unfortunately, the paper [5] did not draw any conclusions from the performed analysis. Watermarking schemes were described separately in the text and compared only in the tables, but there was no discussion of what had been placed in the tables and summarization of the obtained results was also missing. Although the paper contained some conclusions and indicated various future research directions for all types of analyzed cover media, these seemed to be a general discussion which was very weakly related to the particular results obtained during the survey. To be precise, Kumar et al. mentioned addressing the trade-off between various features of watermarking schemes, the preparation of benchmarking tools for reversible watermarking techniques, and the need for watermarking methods for 3D printing models and animation as general potential areas of future work. In the case of images, the authors indicated the necessity of real-time implementation of watermarking, blind watermark detectors, better perceptual models, dual watermarking, and improvement of watermarking security [5] as the main challenges. Similarly, shortening the operation time and meeting real-time requirements was pointed out as the area of further work for video watermarking while, in the case of audio, attacks should be better resisted. Moreover, it was stated that database watermarking requires improvements in terms of computational time, robustness, and the lack of distortions in the cover data, while the main problem with text watermarking approaches is that they are usually suited only for a given alphabet, but should be possible to use with any type of text [5].

The paper [1] is not a classical survey *per se*, but it is one of the most important and highly cited works the in information hiding field. It was published in 1996, and its content is not only limited to watermarking. Instead, it focused first on describing several techniques as possible methods for embedding data in a hidden manner in digital images, audio signals, and text. In this context, digital watermarking was mentioned as one of the potential applications of data-hiding techniques, especially for images, to enable proof of ownership and tracing the distribution of such content on the Internet. Two other applications were also mentioned: tamper-proofing and feature location. The information hiding techniques described were grouped into low and high bit-rate coding. In the former, the authors included Patchwork and Texture Block Coding while, in the latter, methods that replace the least significant luminance bit of image data with the embedded data were enclosed. Finally, the authors concluded that two of the presented solutions –namely, the Patchwork and Texture Block Coding techniques– are particularly promising for digital watermarking purposes.

In [2], Miller et al. presented a book chapter related to the characterization of digital watermarking principles and practices. In the second part of this work, a review of 29 existing watermarking schemes was also included. However, before that, the authors incorporated an informative introduction and explanations of the fundamentals of digital watermarking. For example, they listed several characteristics of watermarks (e.g., fidelity, robustness, fragility, tamper resistance) that might be desirable for various applications. They also discussed the specifics of six intentional and unintentional attacks which a watermark system may encounter. These include attacks on the content, statistical averaging, exploiting the presence of a watermark detector device, based on the presence of a watermark inserter, on the copy protection system, and collusion attacks. The authors also noted that, although a watermark can withstand several signal transformations that happen in commonly used signal processing operations, it is typically more challenging to obtain resistance to intentional tampering. As has already been mentioned, the last part of this work was a survey of several recent (at the time of writing) proposals for watermarking solutions and, for each of them, the authors presented their pros and cons. Of the 29 reviewed articles, the majority of them (26) discussed watermarking schemes for digital images, while 2 of them were for video and 1 considered audio signals. However, some kind of analytical overview or an overall conclusion drawn from this part of the chapter was omitted; instead, the authors included only a very brief summary of the content of their work.

In [89], Meerwald and Uhl presented an overview of wavelet-based watermarking schemes available around the year 2000. In more detail, the authors discussed how different existing watermarking methods are related to image compression. Then, they investigated the robustness of the chosen watermarking algorithms against image compression. An interesting aspect of this work is that the authors used many illustrative materials to show how different wavelet-based watermarking methods differ from each other.

As noted by the authors of [89], the wavelet transform has many benefits when compared, for example, with DCT, which makes it suitable for image compression and watermarking applications; in particular, it has good space-frequency localization for analyzing image features, allows for multi-resolution representation, and presents superior HVS modeling, linear complexity, and adaptivity.

The presented watermarking techniques were discussed in several categories: decomposition strategy, coefficient selection, human visual system modeling, embedding/extraction methods, and application. In terms of decomposition strategy, the authors concluded that the majority of solutions use three or four decomposition steps and apply well-known wavelet filters. As for the selection of the coefficients, they are typically modified in the watermarking process depending on the characteristics of the embedding technique and the application. Moreover, the masking properties of the human visual system are considered either implicitly or explicitly. Regarding the embedding process, the watermark is inserted into the selected coefficients by adding a pseudorandom noise-like spread-spectrum sequence or through a quantization-and-replace strategy. On the other hand, for watermark extraction, blind, semi-blind, and non-blind (or private) schemes can be utilized. Finally, from the perspective of the applications, wavelet-based watermarking has been proposed for copyright protection, image authentication and tamper detection, data hiding, as well as image labeling.

The paper of Xie et al. [90] reviewed several existing solutions for public key digital watermarking. The public key digital watermarking system differs from the symmetric or private watermarking schemes as, in the latter, the key used to embed the watermark is the same as that for its detection. On the other hand, in the public key algorithm, the detection of the watermark is possible with a public key that does not disclose too much information on the embedded watermark in order not to impair it. In this paper, the authors described six existing public key digital watermarking solutions, based on part-key, one-way signal processing, Legendre sequence, Eigenvector, QIM, and DCT coefficients, and custom watermark positioning. The authors concluded that the existing methods are not sufficiently robust against malicious attacks. Therefore, further research effort is needed in this area. It is also worth noting that, apart from the description of the existing schemes, the authors did not include any synthesis and did not point out any specific promising research directions.

Next, in [91], Khan et al. focused on surveying existing reversible watermarking schemes. The main characteristic of this type of solution is that it enables complete extraction of the watermark, along with full restoration of the cover. The authors stated that such methods are gaining interest due to their applications in military communication, healthcare, and law enforcement.

A review of the existing reversible watermarking solutions was performed by distinguishing four groups of schemes on which the watermarking technique is based: compression, histogram modification, quantization, and expansion. The authors briefly surveyed the first three groups (5, 16, and 5 papers, respectively) while their main focus was on expansion-based reversible watermarking techniques (29 papers), which were discussed in detail. The provided justification is that these solutions are capable of achieving high capacity and are computationally efficient. They were further classified into three subgroups: prediction error (22 papers), contrast mapping (3 papers), and interpolation error-based (4 papers). The performed review and comparison of existing reversible watermarking techniques were also presented as two tables, the first of which presented a comparison of different existing reversible watermarking techniques with respect to the type of watermark and whether it is a blind or semi-blind scheme. One of the conclusions is that these techniques do not need cover image-related information on the detection side. The second table provided a comparison of prediction error-based reversible watermarking methods, where expansion and predictor types were highlighted. All techniques mentioned there were fragile and blind. Moreover, in the case of prediction error expansion-based reversible watermarking, the embedding distortion caused to the image relies on the prediction error. For this reason, in order to decrease the extent of prediction error, various types of predictors and prediction contexts are utilized.

An added value of the paper [91] is that the authors also performed an experimental comparison of the selected reversible watermarking schemes with respect to their watermarking properties as well as computational time, which was included at the end of the paper. This analysis was performed on two datasets, one consisting of 300 and the other containing 1500 digital images. For the first dataset, imperceptibility (expressed as PSNR) was investigated at a constant capacity of 1 Kbits. Then, in the other experiment, the embedding capacity was kept at 0.25 bpp. For the second dataset, an embedding capacity of 10 Kbits was used, and PSNR and SSIM index were analyzed. Moreover, embedding and extraction time, which can be a significant challenge in real-life applications in large databases, were also inspected.

The paper concluded by listing some potential future research directions in this field, related to developing a tamper localization technique for reversible watermarking or benchmarking tools for the evaluation of reversible watermarking techniques.

In [92], Pal conducted a brief survey on selected watermarking methods for digital images and their various applications. To this aim, without clearly stating their criteria, the author arbitrarily chose 11 papers describing different watermarking techniques for QR (Quick Response) codes, authentication of medical images, e-commerce applications, and copyright protection and broadcast monitoring. Apart from briefly sketching the idea of each selected work, the author also included a table summarizing the digital watermarking algorithm used, its envisioned application, and its resulting performance. However, it must be noted that only ten papers were enclosed in the table, and the review was not described in a consistent manner.

Next, in [93], Kumari et al. focused on providing an overview of current digital watermarking methods for images, text, audio, and video content. The paper started by reviewing existing, popular, free, and paid watermarking tools that can be found online or installed locally. Then, apart from characterizing the various applications of existing techniques, the authors also described different classifications for digital watermarking algorithms as well as corresponding attacks. They also presented the steps comprising a typical watermarking process.

In the second part of the paper, the authors performed an analysis of the popularity of digital watermarking research, considering how many research works have been published in IEEE, Springer, and Elsevier. It was found that digital watermarking research has evolved very progressively throughout the years. Then, Kumari et al. surveyed 40 existing research works from the period 2013-2018 in a tabular form, characterizing what problem each method solves, the proposed approach, and what results were obtained. The article ended with several conclusions. First, publicly available watermarking tools are not sufficient to provide advanced security mechanisms for digital media content protection. Second, the majority of existing solutions have high cost and computational complexity, and they are prone to geometric attacks. Moreover, they highlighted that more different types of attacks should be considered for these schemes. Finally, according to the authors, with the increasing number of proposed watermarking techniques, there is a pressing need for an unbiased benchmarking technique to evaluate the effectiveness of various solutions from different perspectives, such as robustness, quality, or computational complexity.

Finally, the paper [18] by Byrnes et al. is a very recent survey that reviewed how deep learning methods have been used for data hiding purposes so far, especially for watermarking and steganography applications, based on the network architecture and noise injection methods. The research trend considering the utilization of deep-learning-based models for data hiding in either watermarking or steganography started in 2017, when the first papers began to investigate how convolutional neural networks (CNNs) can be used for data-hiding purposes. The focus has recently shifted towards the utilization of Generative Adversarial Networks (GANs). Information hiding using deep learning applies the encoder-decoder network structure to train models to achieve imperceptibility and robustness of the concealed data. This is an improvement, compared to the classical data hiding methods, as the resulting models can be retrained to resist various potential attacks and adjust to different scenarios. An essential advantage of applying deep learning techniques is that no expert knowledge is required when creating information hiding schemes. Moreover, as deep learning models are considered to be black boxes, this also provides improved security of the overall setup. It should also be noted that the paper [18] also included a very informative introduction to deep-learning-based data hiding architectures.

The deep-learning-based watermarking techniques surveyed in this paper were divided into those that are CNN- and GAN-based. Additionally, the former were classified into autoencoder-based models and adversarial training, while the latter were grouped into Wasserstein GAN and CycleGAN. The main conclusion of this part of the review is that, for state-of-the-art deep learning methods, the GAN framework is currently the most promising solution in terms of robustness and secrecy. The conducted analysis was also expressed in tabular form, where each proposed method was characterized according to the domain, embedding and extractor networks, host resolution adaptability, and control for influencing the trade-off between imperceptibility and robustness.

In the following part of the survey, the authors also devoted some space to the characterization of watermark removal techniques. They distinguished three categories: blind watermark removal (e.g., through compression and geometric distortions), key estimation, and tampering attacks. In the case of blind removal, deep learning methods can help to alleviate this threat through varied attack simulation strategies, although deep-learning-based watermark removal techniques also exist, which has lead to a kind of "arms race" between the defenders and attackers. The other types of attacks require some knowledge about the watermarking algorithm and, as already mentioned, deep-learning-based watermarking has a black-box nature; thus, its specifics cannot be readily uncovered by adversaries.

Moreover, it has been revealed that most existing works in this area focus on image-based data hiding. For this reason, this survey was concentrated on their comparison. As a result, the authors analyzed over 30 papers related to deep data-hiding methods. They also compared the chosen deep-learning-based watermarking models with respect to robustness (expressed using BER) in tabular form. For each method, they characterized its architecture, cover image dimensions, watermark bits, and BER. Then, the authors also systematically explained and compared various objective strategies –including the MSE, mean absolute error, cross entropy loss, mean-variance loss, adversarial loss, Kullback–Leibler (KL) divergence, cycle consistency loss, and Wasserstein loss– and evaluation metrics (robustness, quality, capacity), as well as various training datasets utilized for benchmarking existing deep data-hiding techniques.

Finally, potential future research directions related to deep learning data hiding techniques were outlined. This included, for example, expanding the applications of deep learning digital watermarking models, such as by using them for text watermarking in order to combat misinformation or mitigating watermark removal attacks. Additionally, the authors proposed to apply watermarking for the protection of machine learning models and launching backdoor attacks.

VIII. RESULTS OF THE META-SURVEY

In this section, we pinpoint the main conclusions derived from the analysis conducted on existing surveys. In more detail, we first present quantitative results, followed by a more general discussion.

A. QUANTITATIVE RESULTS

As previously mentioned, for this meta-survey, we analyzed 64 survey papers on multimedia watermarking. These papers were selected from 133 downloaded articles, and more than a half of the retrieved surveys had to be rejected due to their low quality. Moreover, even some of the selected research turned out to be of poor quality, as pointed out in their descriptions in previous sections.

A quantitative summary of the results of the conducted meta-survey (shown in Tables 2 and 3) is depicted in Table 4. This table presents the number of surveys in which the given type of application, attack or cover object is mentioned ($\surd$), purposely omitted ($\times$) or not mentioned at all (N/A). As the same paper can describe both robust and fragile/semi-fragile techniques (and, consequently, appear twice in Tables 2 and 3), in Table 4 we first present the results regarding the type of watermark (robust or fragile/semi-fragile), which are just sums of rows with the given value from Tables 2 and 3 and also include those overlapping papers, and then provide the unique results without distinction between the type of watermark; in these results, each article is included only once.

More than one-third of the analyzed surveys (23 out of 64) discussed both robust and fragile/semi-fragile approaches. In general, however, robust techniques were discussed more often than fragile and semi-fragile ones. The former category was described in 54 out of 64 papers (84%), while the latter was discussed in 33 out of 64 cases (52%).

Regarding surveys dealing with robust watermarking algorithms, it turns out that the copyright protection was the most explored application of these techniques: almost all of the studies (53 out of 54) indicated its usage. Copy control, fingerprinting and broadcast monitoring were other meaningful areas in which watermarks can be utilized; however, the results of our analysis indicated that they were mentioned in a half or less of cases (27, 25, and 21 out of 54 papers, respectively). The least-explored application of robust watermarking was device control (10 out of 54 papers).

In the case of applications related to fragile or semi-fragile algorithms, content authentication was described most often (27 out of 33 cases), followed by tamper detection, authentication, and tamper localization (mentioned in 23, 21, and 16 out of 33 papers, respectively). Content labeling seemed to be the most rarely analyzed application of fragile watermarking, appearing in only 6 of 33 papers.

All five applications of robust digital watermarking (i.e., copyright protection, copy control, device control, broadcast monitoring, and fingerprinting) were discussed in six surveys. Moreover, one of these studies, written by Begum and Shorif Uddin [14], also dealt with all of the applications of fragile watermarks (i.e., content authentication, tamper detection, authentication, tamper localization, and content labeling). As a result, this paper is the only one which described all applications of watermarking distinguished in our meta-survey. On the contrary, there was one paper in each category (robust or fragile/semi-fragile) which did not mention (deliberately or not) any of the applications from this category. First, Cox and Miller [3] purposely omitted content authentication and did not mention any other fragile applications, but focused on robust methods instead. Second, Pathak and Jain [73] did not present any applications of robust watermarks and did not deal with fragile or semi-fragile approaches, instead dealing with attacks.

Removal and geometric attacks seemed to be the most broadly described in the existing literature on robust watermarks. Both of them were presented in vast majority of papers on robust techniques (44 and 37 cases out of 54, respectively). Exploration of other types of attacks on watermarks in this branch of the literature was less extensive. While cryptographic and protocol attacks still appeared in more than 20% of cases (16 and 13 out of 54 papers, respectively), the copy-move attack was dealt with only in 15% of papers (8 out of 54), and collage and vector quantization attacks were discussed very rarely in the analyzed surveys: the former appeared in only 5 surveys, while the latter appeared in just 3 papers.

The frequency of description of particular types of attacks differed slightly in the papers on fragile and semi-fragile algorithms. Although removal and geometric attacks were still the most often discussed (23 and 20 cases out of 33, respectively), the occurrence of other types of attack differed. Protocol and copy-move attacks were dealt with in about one-third of this category of papers (13 and 11 out of 33 cases, respectively), while the last three types of attacks were still discussed in more than 25% of cases, with both cryptographic and collage attacks in 10 out of 33 papers and vector quantization in 9 out of 33 papers.

As a result, removal and geometric attacks were the most explored in the considered literature, regardless of the discussed watermark type (51 and 41 papers out of 64, respectively), while less attention was paid to cryptographic, protocol, and copy-move attacks (20, 18, and 14 out of 64 cases, respectively). The least-considered attacks were collage and vector quantization, discussed in 12 and 9 out of 64 papers, respectively.

The problem of attacks was not discussed at all –either in purpose or not– in eight papers describing robust approaches and in eight surveys on fragile watermarking. In total, 10 out of all 64 papers did not describe possible attacks on watermarks. Only one of them, by Bhattacharya et al. [68], which dealt with robust applications of watermarks, deliberately omitted some types of attacks. The rest of these papers simply did not mention any kind of attacks considered in our meta-survey. To the contrary, two studies mentioned all seven types of attacks which we have taken into account. Both of these papers dealt only with fragile/semi-fragile watermarks.

The vast majority of all analyzed surveys dealt with images (54 out of 64 cases), regardless of the type of watermark presented in the papers. In the case of articles presenting robust approaches, 44 out of 54 cases considered images while, in the fragile/semi-fragile category, this was 32 out of 33 papers. The second most explored type of cover object was video, but this was present in less than the half of the analyzed studies. This type of media was discussed in 29 out of all 64 papers and, considering the type of watermarks discussed in particular surveys, it was mentioned in 27 out of 54 papers on robust watermarks and in 17 out of 33 papers on fragile techniques. It seems that speech and audio are the least mature cover type in terms of digital watermarking, which were taken into account in only around one-third of all analyzed surveys (21 out of 64). Its exploration was a bit bigger when we distinguished between the analyzed types of watermark. In the studies on fragile algorithms, audio as a cover type was discussed in 12 out of 33 cases while, in the literature on robust watermarking, this type of media was mentioned in 20 out of 54 cases. Surveys typically focused on one or (at most) two types of cover media; however, in 13 out of 64 papers, all three types of the host signal (i.e., image, video, and audio) were discussed. Regarding the distinction between watermark type, 12 out of 54 papers on robust algorithms and 11 out of 33 papers on fragile techniques discussed all three types of cover media.

From Table 4, it can be seen that information about given types of applications and attacks was deliberately omitted (×) in very rare cases. More often, the given aspect of algorithms was not mentioned at all (N/A). To the contrary, the situation was opposite when considering the type of analyzed media. Here, in a vast majority of cases the given media types were purposely not discussed, as the authors focused on some specific type of cover. Completely missing information about given type of media was a very rare case.

As can be seen from the quantitative results, there are several areas which require further research efforts. First, watermarking of speech and audio content is much more rare than that of images. Although it is due to the fact that image watermarking techniques have been developed for a longer time –being the first type of watermarked multimedia– there is still a gap to be filled in this regard. Second, the discussion on some types of attacks on watermarks, particularly collage and vector quantization attacks, seems to be lacking in comparison with other types of watermarks. Last, but not least, several applications of watermarks were covered in the surveys many times less often than the others. Particularly, device control (in the case of robust algorithms) and content labeling (in terms of fragile watermarking) were rarely mentioned in the literature.

B. DISCUSSION

According to the meta-survey described in the previous sections, here we summarize the research trends of digital watermarking technologies for each digital media content area.

Throughout the meta-survey, the majority of the media content covered was image data; however, as summarized in the previous section, video, audio, text, and other data have also been investigated. Among image watermarking, the evaluation of relevant algorithms is mainly performed on small images, typically of size 256×256 or 512×512 pixels. For quality assessment, human visual and auditory sensitivity are considered when embedding watermarks; however, only the use of certain subjective metrics such as PSNR, SSIM, and Perceptual Evaluation of Audio Quality (PEAQ), as well as objective metrics such as ODG, have been mentioned for measuring the quality of watermarked content. Due to the difficulty of fairly comparing performance with respect to capacity, many surveys only listed the characteristics of each watermarking scheme.

Most surveys have attempted to classify watermarking techniques in terms of their embedding/extraction domain, applications, and performance improvements. The typical domains are mainly the spatial or frequency domains (transformed by DFT, DCT, DWT, and so on). As a tendency of the transformed domain, DWT is sometimes selected due to its affinity to human perceptual characteristics such as those of the HVS and HAS. This can be regarded as the extraction of feature elements from a given multimedia content, and the choice strongly depends on the requirements of the particular attack or application. While many surveys have enumerated advantages and disadvantages, there has been no clear description of the suitability of the domains in terms of maximizing robustness, imperceptibility, and capacity. Due to the several factors that must be considered in each application, it seems difficult to find an optimal domain and approach for watermarking techniques.

In the analyzed literature, watermarking techniques consisted of three main operations: extracting the host signal from given multimedia content, inserting a message into the host signal, and detecting/extracting the message. Through this meta-survey, it was found that robust techniques have mainly been investigated for the first operation, in order to improve the tolerance for attacks such as lossy compression, filtering operations, and geometric modifications. On the other hand, direct modification of spatial signals has been explored in the context of fragile watermarking, which can be used to authenticate multimedia content. The first operation can be regarded as the feature extraction of multimedia content, where the extracted features have specific meaningful properties with respect to robustness, imperceptibility, and capacity. However, there is no universal feature set that satisfies all the requirements of watermarking algorithms.

TABLE 4. Summary of quantitative results of the meta-survey presented in Tables 2 and 3. (Abbreviations used are the same as in Tables 2 and 3).

Type of watermark	Application					Attacks							Cover object		
Robust (54 papers in total, including 23 describing both types of watermarks)	**CP**	**CC**	**DC**	**BM**	**FP**	**RM**	**GM**	**CR**	**PR**	**CM**	**CG**	**VQ**	**Image**	**Speech / Audio**	**Video**
√	53	27	10	21	25	44	37	16	13	8	5	3	44	20	27
×	0	5	5	5	5	1	4	10	11	8	7	8	10	32	27
N/A	1	22	39	28	24	9	13	28	30	38	42	43	0	2	0
Fragile / Semi-fragile (33 papers in total, including 23 describing both types of watermarks)	**CA**	**TD**	**AU**	**TL**	**CL**	**RM**	**GM**	**CR**	**PR**	**CM**	**CG**	**VQ**	**Image**	**Speech / Audio**	**Video**
√	27	23	21	16	6	23	20	10	13	11	10	9	32	12	17
×	1	0	0	0	1	0	2	1	1	1	0	1	1	18	15
N/A	5	10	12	17	26	10	11	22	19	21	23	23	0	3	1
Both (64 unique papers)	not applicable					**RM**	**GM**	**CR**	**PR**	**CM**	**CG**	**VQ**	**Image**	**Speech / Audio**	**Video**
√						51	41	20	18	14	12	9	54	21	29
×						1	5	10	11	8	7	8	10	40	34
N/A						12	18	34	35	42	45	47	0	3	1

Therefore, researchers have investigated techniques that can cover multiple requirements by combining some additional operations at the embedding and detection/extraction stages.

In the case of robust watermarking, tolerance against removal and geometric attacks is essential, and most surveys have taken these attacks into account, as described in Tables 2, 3, and 4. On the other hand, cryptographic and protocol attacks are considered when intentional modification is expected in the application. The fingerprinting scheme is one of the applications of robust watermarking, which is assisted by the cryptographic protocol to ensure the integrity during content distribution and to identify illegal users from pirated copies.

With advances in machine learning techniques, several techniques such as NN, SVM, PCA, and DL have been employed to improve the performance of watermarking in terms of robustness and imperceptibility. Once the threat conditions are clearly defined, the corresponding robust feature space can be determined by using machine learning algorithms. If all conditions are explicitly given in advance, the distortion caused by embedding the watermark signal can also be minimized. One of the difficulties in watermarking techniques is that the conditions are not always fixed and the determination of threshold values is generally ambiguous. For example, the visual and auditory quality of multimedia content is very sensitive to user preferences.

Depending on the application, the computational cost becomes important to realize real-time applications, although this term is not detailed in Tables 2 and 3. In the case of video watermarking, it is general to execute watermarking algorithms on encoded video formats such as MPEG-2, MPEG-4, H.264/AVC, H.265, and HEVC, due to the large file size. The adaptation of new video formats is one of major issues in video watermarking.

Interesting application trends were also observed from the meta-survey on fragile watermarking. The applications for medical and military images are the main targets for reversible (lossless) methods, as they are considered sensitive to even small changes. On the other hand, self-recovery schemes may compromise the quality of reconstructed content to make it robust against minor changes caused by unintentional signal processing operations, such as lossy compression. In the case of video streams, it seems difficult to apply reversible methods, as they assume the raw format of the content. Hence, only self-recovery methods were discussed in relevant surveys.

IX. DIGITAL WATERMARKING TECHNIQUES FOR FAKE NEWS DETECTION

In the last few years, the accelerated adoption of social media platforms has facilitated information sharing between users, who can create and share information much more easily than in the past. However, this facility has been exploited to create and disseminate hoaxes and fake news. Disinformation is mainly distributed through social media platforms and is not restricted to textual data, with manipulated digital contents (audio, images, and video) increasingly being used to distribute fake news.

This problem is continuously increasing. In fact, in the 2016 U.S. election, it was estimated that social media platforms accounted for more than 41.8% of the fake news data traffic regarding the election [110], which is far more effective than traditional communication channels (e.g., radio, television, or the printed media). Recently, the SARS-CoV-2 pandemic has also been the focus of many disinformation campaigns concerning the virus, its severity, infection, treatment, and vaccination, resulting in deaths across the world.

The use of digital watermarking techniques to help in fake news detection and other security issues in Online Social Networks has been recently proposed in [111], [112], [113], [114], [115], [116], and [117]. A complete architecture for the detection of fake news distributed as multimedia content was described in [15]. The following paragraphs present an overview of these recent works.

A. EXISTING WORKS

Reference [111] dealt with several issues related to security in Online Social Networks, including disinformation. This work also mentioned (in its Section 4.1) several applications of digital watermarking in the context of Online Social Network security, with different objectives such as proof of ownership, protecting user privacy, and tracking activities such as re-posting or modification of the contents owned by other users. However, this reference did not consider the application of watermarking to fake news or disinformation detection but, rather, focused on user privacy and preventing the theft of private user data. In any case, this reference overviewed some works that proposed a dual robust and (semi)fragile watermarking approach for some applications, as we also propose for fake news detection in this section.

In [112], a so-called "steganographic" method based on neural networks (stegoNN) was proposed to detect tampering (photomontage) manipulations in pre-embedded images. The proposed approach uses a neural network to create a series of attributes that can be used later on to detect tampered images. Once the images are "signed" with the proposed approach, there is no need for any external data or the original image to detect modifications. The embedded information can also be used to localize the particular areas of the images that are modified.

Next, in [113], a proof-of-concept of a deepfake video news detection and prevention system using watermarking and blockchain was presented. In the proposed approach, the Digimarc's robust audio and video watermarking schemes are used to embed the watermark both in the audio and video tracks of news clips prior to their distribution. Blockchain technology is used to store the video and its metadata for future forensic analyses. The watermarks can be detected from online social network portals, nodes, and back-ends. A two-stage detection process is carried out: first, the Digimarc watermark reader is applied to the video frames; second, the information stored in the blockchain is retrieved and applied. The proposed scheme provides an informal security analysis of the deepfake detection scheme and presents simulation results using a face-swap algorithm as a proof of concept.

Reference [114] is a Bachelor's Thesis that presented an LSB-based digital watermarking scheme for fact-checking and fake news detection. The proposed approach embeds a descriptive watermark into a 2D image. The watermark includes detailed information about the image, such as its owner, its contents, the date on which it was taken, and the location. Photographers and news agencies should embed a descriptive watermark in each photo they own using this scheme, and social network owners should include the scheme as part of their content-sharing procedure. Whenever an image is uploaded by a user, the watermark is extracted. If the caption entered by the user does not match the extracted watermark, the image will not be posted. Although the proposed solution is too fragile to be used in practice, it points out some interesting ideas to be used in this context.

A watermarking-based image tagging method to be used in deepfake provenance tracking was introduced in [116]. More precisely, a deep-learning-based approach, called FakeTagger, was proposed, consisting of a simple encoder and decoder design to embed a message in a facial image. Experimental results showed that the embedded message can be recovered with high confidence (over 95%), even after several GAN-based deepfake transformations. The embedded message can be used to encode the identity of the facial images and can contribute to deepfake detection and provenance. The purpose of the proposed method is to prevent personal photos from being deepfaked.

Then, in [117], a proof-of-concept deepfake detection system was presented. The proposed method aimed to detect fake news video clips generated using voice impersonation, using digital watermarks that are embedded in the audio track of a video using a hybrid speech watermarking technique. A standalone software application can perform the detection of robust and fragile watermarks. The paper also presented different simulations, performed to evaluate the embedded watermark's robustness against common signal processing and video integrity attacks, which can be regarded as one of the first few attempts to use digital watermarking for fake content detection.

In [15], the architecture of a fake news detection system, which is being developed within the ongoing Detection of fake newS on SocIal MedIa pLAtfoRms (DISSIMILAR) [115] project, was presented and described. The proposed architecture was designed for the protection of digital media content (i.e., images, video, and speech) and, to fulfill its goals, it combines digital watermarking, signal processing, and machine learning techniques. As partial solutions for the detection of deepfakes and fake content have not been proven successful enough to date, the combination of different technologies is expected to provide a more robust and consistent approach to the challenge of detecting and tracing disinformation on Online Social Networks.

Finally, a recent preprint [118] has discussed the possibility of using a deep-learning-based semi-fragile watermarking scheme for the detection of deepfakes. Instead of detecting fake media directly through machine learning classification, the proposed approach embeds a semi-fragile watermark into the media (e.g., an image, as used in their experiments) using a deep learning algorithm, such that its authenticity can be proved later on. The proposed framework is fragile against facial manipulations and tampering attacks, but robust to different benign image processing operations, such as compression, scaling, saturation, contrast adjustment, and so on. In this way, images that are distributed over the Internet can be verified unless deepfake modifications are applied to them.

B. DIGITAL WATERMARKING APPLICATIONS FOR FAKE NEWS DETECTION

In [15], the authors discussed three different areas in which existing digital watermarking solutions can contribute to the detection of fake news: (1) to validate the provenance of legitimate news (i.e., to discard that some news is fake), (2) to authenticate legitimate news (i.e., if the news is not identified as legitimate, they shall be further investigated), and (3) once fake news has been identified, to trace the source and classify it as a suspected malicious fake news creator/distributor. Therefore, we consider three new application areas for digital watermarking: Legitimate News Provenance (LNP), Legitimate News Authentication (LNA), and Fake News Tracking (FNT).

A more detailed description of these three new application areas is provided below.

- Legitimate News Provenance (LNP): Among the efforts to combat the proliferation of fake news, LNP represents a pivotal application area for digital watermarking technologies. LNP involves the integration of digital watermarks into news content to validate its origin. By embedding unique identifiers within legitimate digital contents, news agencies and media can provide an irrefutable trail of provenance, allowing consumers to verify the source and trace the history of the news they encounter. This not only serves to discard suspicions of fake news, but also bolsters trust in the veracity of news sources. The watermarking technique can include metadata such as timestamps, author information, and publication details, enabling consumers to confidently differentiate between credible journalism and potentially misleading or fabricated content.
- Legitimate News Authentication (LNA): LNA harnesses the power of digital watermarking to address the critical issue of verifying the authenticity of news contents in the modern information landscape. In an era where the dissemination of misinformation and deepfakes is prevalent, LNA offers a robust solution. By embedding watermarks into news content, publishers can certify the legitimacy of their reports. When readers encounter a digitally watermarked news piece, they can be assured that it has undergone a rigorous authentication process and no malicious manipulation has been applied. If some content lacks such authentication, or it is found to be modified in some areas, it may signal the need for further scrutiny or investigation. LNA not only safeguards against the spread of false information, but also empowers individuals to make informed choices about the credibility of the news they consume.
- Fake News Tracking (FNT): The proliferation of fake news is a pressing concern in the modern digital landscape, necessitating innovative approaches for identifying and mitigating its impact. FNT represents a proactive application of digital watermarking technology in the battle against misinformation. FNT involves the insertion of traceable markers or forensic watermarks into news contents that have been identified as fake or malicious. These watermarks serve as digital breadcrumbs, enabling investigators and fact-checkers to trace the source of the misinformation back to its origins. By categorizing and classifying suspected fake news creators and distributors, FNT not only aids in the identification of malicious actors but also supports legal and regulatory efforts to curb the dissemination of false information. FNT, in conjunction with other measures, offers a multifaceted approach for tackling the root causes of the fake news epidemic.

Indeed, digital watermarking is a very promising tool that can help in fake news detection, in combination with other solutions, such as multimedia forensics (to detect malicious manipulations of contents) and machine learning. The main advantage of digital watermarking is that watermark extraction can be performed very efficiently, even for advanced watermarking algorithms. In this way, it can be relatively easy to determine if some contents are legitimate when a watermark of the producer is detected. If legitimate news producers embed authentication (LNP and LNA) watermarks in the contents they distribute over social networks and websites, the task of fake news detection can be significantly simplified.

LNP watermarks need to be robust at least against the transformations that occur in the communication channel of social media platforms and, more specifically, to cropping, scaling, re-sampling, and (re)compression attacks. Additionally, some filters may be applied to the platforms to highlight some details of the contents. In that case, such filters should also be included in the collection of robustness attacks. Regarding LNA watermarks, they would also be embedded at the origin by the media producer, with the aim of detecting manipulations. Typically, LNA watermarks should be semi-fragile, allowing for some alterations (e.g., re-compression), but would be erased if stronger modifications were applied.

Finally, traceability is another relevant challenge in the management of fake news. Once some content is identified as fake, determining the source of the manipulation is the desired property, as such a source could be labeled as unreliable and suffer from reputation loss in the future. Traceability and data provenance are possible applications of digital watermarking using techniques similar to those used in transaction tracking (or digital fingerprinting). Social media platforms would have to embed a user-specific watermark within the content published by a user. This could also be recorded in a database or a blockchain, as proposed in [117]. Extracting the user-specific watermark may suffice to determine the source of particular fake content.

The properties required for the three new applications of digital watermarking for the detection of fake news are detailed in Table 5.

1) Typically, the contents on social media platforms are not characterized by high quality, as they are often

TABLE 5. Properties required for the new watermarking applications: Legitimate News Provenance (LNP), Legitimate News Authentication (LNA), and Fake News Tracking (FNT).

Application	Imperceptibility	Real time (low cost)	Blindness	Robustness	Attacks
LNP	Medium/High	Yes	Yes	Robust	Re-compression, some filters, cropping, scaling, and others
LNA	Medium/High	Yes	Yes	Semi-fragile	Re-compression and channel-specific attacks
FNT	Medium/High	Yes	Yes	Robust	Re-compression, some filters, cropping, scaling, and others

compressed to minimize their size. For this reason, in general, high imperceptibility will not be a hard requirement of the watermarking systems used for LNP, LNA, and FNT. This is an advantage compared to other areas of application, such as Digital Rights Management (DRM) or streaming solutions, where high imperceptibility is always required. In particular, the imperceptibility requirements only become stronger when the platform allows high-quality content.

2) Regarding the real-time extraction or detection of watermarks, all three applications would require a quick analysis to give feedback to users. Hence, a real-time response is always required in this framework.
3) As blind extraction/detection is concerned, we have a similar situation. In general, there would be no original content to use for extracting the watermarks. Similarly, semi-blind methods that require knowledge about the embedded watermark, in order to check whether it is embedded or not in some content, should also be avoided, as they may require performing several tests to check specific content, and the feedback to the user could be too slow in such a case.
4) Finally, LNP and FNT applications must be robust against different attacks, including re-compression, (some) filtering, cropping, scaling, and other signal processing operations. On the other hand, the LNA watermarks should only be preserved in the case of re-compression or certain channel-specific attacks (maybe re-scaling, if the platform carries out such kinds of transformations); however, any other attack should make the authentication watermark undetectable and, hence, reveal a possibly malicious modification or forgery.

X. FUTURE RESEARCH DIRECTIONS

The necessity of content protection in digital media has led to massive growth in the field of digital watermarking, where researchers are motivated to devise innovative solutions for this purpose. At present, cyberspace is hectic and having control over it is rarely possible. In accordance with the IFPI (International Federation of the Phonographic Industry), 95% of music is illegally downloaded. Moreover, Digital Life America has published a survey showing the same results for movies. Hence, there is strong advocacy to adopt new rules and create systems to protect intellectual property, which is why new watermarking techniques need to be developed.

Particular care must be taken to ensure the survival of the embedded watermark against such attacks to attain the required functionalities in the target application. For medical images, regarding applications such as tele-radiography, in which the medical images may be transferred through communication networks, the watermarked images should be robust against signal noises of the transmission channel. With the development of smart cities, some services such as smart health systems have also been noticed, in which the medical information of patients can be transmitted to the hospitals digitally. Hence, another direction of digital watermarking is the protection of digital patient profiles from tampering.

Furthermore, considering the nature of digital watermarking, in which hidden data are transferred with the main media, these techniques do not require additional memory or space to carry out the watermark bits. As a result, they are good candidates for Internet of Things (IoT) applications, in which lightweight design is prioritized for implementation. Thus, another future venue to pursue in the watermarking field is to apply watermarking in lightweight IoT applications for privacy protection, control of information leakage, and data tampering prevention, such as in smart home environments.

Another different potential research direction is the utilization of digital watermarking for fake news detection in multimedia content [15]. This is the most recent application of such techniques, which definitely requires further systematic investigations and research efforts. Clearly, combining watermarking for digital content marking and the tracing of sources with existing technologies such as AI and multimedia forensics to create a more complete solution is a promising research direction.

Next, one of the current rising trends of research is related to the application of deep learning techniques to improve the characteristics of digital watermarking techniques [18]. Moreover, it should be noted that similar methods may be used to boost the efficacy of attacks against embedded watermarks. Therefore, analyzing the resistance of digital watermarking techniques against deep-learning-empowered attacks is currently a promising research direction in this area.

Regarding the watermarking of multimedia content, the possibility of using this type of technique in real time – especially for video– seems to require further attention [5]. This refers not only to traditional video files, but also to the streaming of live events, which must fulfill more strict requirements on robustness and other characteristics of the watermark [87]. Megías et al. [87] have indicated that the combination of different existing approaches for fingerprinting in decentralized distribution scenarios is worth exploring to improve their copyright protection.

Another interesting area is reversible watermarking. Research on this topic can be twofold. First, there is a need for assessment of such kinds of approaches, but there is a lack of benchmarks that could be used for this purpose [5]. Second, the development of reversible watermarking methods with various types of robustness (of the watermark or the host signal) against a wider range of attacks is still an open issue [103]. According to Menendez-Ortiz et al. [103], robust reversible schemes (which guarantee restoration of the watermark) for audio and video, as well as self-recovery schemes allowing for perfect reconstruction of the signal and the combination of the latter with fragile reversible watermarking (which allows one to obtain both the original signal and watermark) for various types of multimedia, require further research efforts.

Watermark embedding, which can be categorized into correlation- and quantization-based methods, has introduced several extended operations over the past two decades. These operations are also strongly related to the detection and extraction operations. Unfortunately, few papers in this meta-survey mentioned the differences between detection and extraction. To prevent false positives, we must first check the presence of a hidden message in a given content, which is detection. Especially in the application of fingerprinting, we must avoid catching innocents, even if some guilty users may not be identified. In binary classification, the equal error rate is one of the key metrics for performance evaluation; however, in the case of fingerprinting, false positives should be noted even if false negatives are ignored.

After the presence of a hidden message is confirmed, the message should be restored. To improve robustness, ECC should be used. It is worth noting that the amount of watermark information is not equal to the energy of the watermark signal. For example, suppose that a 100-bit message is encoded in an ECC codeword. Even if the bit-length of the codeword is greater than 100, the amount of information is still 100 bits. Furthermore, if a logo image is inserted as a watermark, the presence of the logo image can be detected by the NCC, but the amount of information is not equal to the bit-length of the logo image. Depending on the applications, it is necessary to take this into account and use the terms detection and extraction differently.

From a different point of view, fragile watermarking techniques are intended to investigate tamper detection and correction. The self-recovery capability is one of the attractive properties in the variants of fragile techniques, which can be extended to protection against illegal manipulations (e.g., DeepFakes) if official multimedia content are pre-processed before being revealed to the public. One of the recent movements of content management is to record the all editing history of multimedia content, called the content credential,[1] and a content credential initiative has been launched.[2] The use of fragile or semi-fragile watermarks will contribute to the realization of such a new framework, but this surely requires further attention.

XI. CONCLUSION

In this study, we performed a meta-survey of existing surveys related to different aspects of digital watermarking in multimedia content between 1996 and 2022. In total, we retrieved and examined more than 130 surveys covering various parts of watermarking research, applications, and attacks, of which, based on their content and quality, we selected the 64 articles that are described in this work.

The results of the conducted meta-survey clearly indicate that the majority of approaches were focused on image data. Although the same research covered both of the topics of robust and fragile watermarking in many cases, robust approaches were discussed much more often than fragile ones. Copyright protection and content authentication turned out to be most-explored applications of robust and fragile watermarking, respectively. Regarding attacks on watermarks, the meta-survey showed that removal and geometric attacks were discussed the most often, while collage and vector quantization attacks require further attention.

Moreover, we analyzed how watermarking methods can help in thwarting issues related to fake news, which is the most recent application of such techniques. To this end, we presented some early approaches addressing this problem and defined the properties required for new watermarking applications, namely, legitimate news provenance, legitimate news authentication, and fake news tracing.

Finally, we indicated some promising future research directions, including new applications such as applying digital watermarking for protection in IoT environments or limiting the spread of fake news. Moreover, applying deep learning techniques in order to improve the characteristics of digital watermarking techniques as well as to improving resistance against deep-learning-empowered potential attacks, are presently dynamic research trends in this area. Other interesting research avenues involve investigating how to conveniently watermark live streams in real time or the development of new reversible techniques and novel approaches for fingerprinting.

REFERENCES

[1] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Syst. J.*, vol. 35, no. 3.4, pp. 313–336, 1996.

[2] M. Miller, I. Cox, and J.-P. Linnartz, "A review of watermarking principles and practices," in *Digital Signal Processing in Multimedia Systems*, K. K. Parhi and T. Nishitani, Eds. Marcell Dekker Inc., 1999, pp. 461–485.

[3] I. J. Cox and M. L. Miller, "A review of watermarking and the importance of perceptual modeling," *Proc. SPIE*, vol. 3016, pp. 92–99, Jun. 1997. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/3016/0000/Review-of-watermarking-and-the-importance-of-perceptual-modeling/10.1117/12.274502.short?SSO=1

[1] https://helpx.adobe.com/photoshop/using/content-credentials.html

[2] https://contentauthenticity.org/

[4] G. C. Langelaar, I. Setyawan, and R. L. Lagendijk, "Watermarking digital image and video data. A state-of-the-art overview," *IEEE Signal Process. Mag.*, vol. 17, no. 5, pp. 20–46, Sep. 2000.

[5] S. Kumar, B. K. Singh, and M. Yadav, "A recent survey on multimedia and database watermarking," *Multimedia Tools Appl.*, vol. 79, nos. 27–28, pp. 20149–20197, Jul. 2020.

[6] X. Niu, C. Shao, and X. Wang, "A survey of digital vector map watermarking," *Int. J. Innov. Comput., Inf. Control*, vol. 2, no. 6, pp. 1301–1316, Dec. 2006.

[7] J.-U. Hou, D. Kim, W.-H. Ahn, and H.-K. Lee, "Copyright protections of digital content in the age of 3D printer: Emerging issues and survey," *IEEE Access*, vol. 6, pp. 44082–44093, 2018.

[8] A. Iacovazzi and Y. Elovici, "Network flow watermarking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 512–530, 1st Quart., 2017.

[9] R. Halder, S. Pal, and A. Cortesi, "Watermarking techniques for relational databases: Survey, classification and comparison," *J. Universal Comput. Sci.*, vol. 16, no. 21, pp. 3164–3190, Dec. 2010.

[10] A. Soltani Panah, R. Van Schyndel, T. Sellis, and E. Bertino, "On the properties of non-media digital watermarking: A review of state of the art techniques," *IEEE Access*, vol. 4, pp. 2670–2704, 2016.

[11] F. Boenisch, "A survey on model watermarking neural networks," 2020, *arXiv:2009.12153*.

[12] Y. Li, H. Wang, and M. Barni, "A survey of deep neural network watermarking techniques," *Neurocomputing*, vol. 461, pp. 171–193, Oct. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S092523122101095X

[13] A. Nikolaidis, S. Tsekeridou, A. Tefas, and V. Solachidis, "A survey on watermarking application scenarios and related attacks," in *Proc. Int. Conf. Image Process.*, 2001, pp. 991–994.

[14] M. Begum and M. S. Uddin, "Digital image watermarking techniques: A review," *Information*, vol. 11, no. 2, p. 110, Feb. 2020. [Online]. Available: https://www.mdpi.com/2078-2489/11/2/110

[15] D. Megías, M. Kuribayashi, A. Rosales, K. Cabaj, and W. Mazurczyk, "Architecture of a fake news detection system combining digital watermarking, signal processing, and machine learning," *J. Wireless Mobile Netw., Ubiquitous Comput., Dependable Appl.*, vol. 13, no. 1, pp. 33–55, Mar. 2022.

[16] P. Korus and N. Memon, "Content authentication for neural imaging pipelines: End-to-end optimization of photo provenance in complex distribution channels," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8613–8621.

[17] P. Kadian, S. M. Arora, and N. Arora, "Robust digital watermarking techniques for copyright protection of digital data: A survey," *Wireless Pers. Commun.*, vol. 118, no. 4, pp. 3225–3249, Jun. 2021.

[18] Z. Wang, O. Byrnes, H. Wang, R. Sun, C. Ma, H. Chen, Q. Wu, and M. Xue, "Data hiding with deep learning: A survey unifying digital watermarking and steganography," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 6, pp. 2985–2999, Dec. 2023.

[19] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*, 2nd ed. San Francisco, CA, USA: Morgan Kaufmann, 2008.

[20] S. Rohith and K. H. Bhat, "A simple robust digital image watermarking against salt and pepper noise using repetition codes," *Int. J. Signal Image Process.*, vol. 3, no. 1, pp. 47–54, 2012.

[21] S. P. Mohanty, K. R. Ramakrishnan, and M. Kankanhalli, "A dual watermarking technique for images," in *Proc. 7th ACM Int. Conf. Multimedia*, Oct. 1999, pp. 49–51.

[22] Y. Tang, K. Li, C. Wang, S. Bian, and Q. Huang, "A two-stage robust reversible watermarking using polar harmonic transform for high robustness and capacity," *Inf. Sci.*, vol. 654, Jan. 2024, Art. no. 119786.

[23] B. Wang, Y. Wu, and G. Wang, "Adaptor: Improving the robustness and imperceptibility of watermarking by the adaptive strength factor," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 11, pp. 6260–6272, Nov. 2023.

[24] M. A. Akhaee and F. Marvasti, "A survey on digital data hiding schemes: Principals, algorithms, and applications," *ISeCure*, vol. 5, no. 1, pp. 5–36, 2013.

[25] E. Hussein and M. A. Belal, "Digital watermarking techniques, applications and attacks applied to digital media: A survey," *Threshold*, vol. 5, p. 6, Jan. 2012.

[26] S. Malshe, H. Gupta, and M. K. Baghel, "Digital image watermarking in robust feature region set," *Int. J. Comput. Appl.*, vol. 55, no. 16, pp. 41–47, Oct. 2012.

[27] Y. Liu, Y. Wang, and X. Zhu, "Novel robust multiple watermarking against regional attacks of digital images," *Multimedia Tools Appl.*, vol. 74, no. 13, pp. 4765–4787, Jul. 2015.

[28] A. Nikolaidis, "Local distortion resistant image watermarking relying on salient feature extraction," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, pp. 1–17, Dec. 2012.

[29] A. Cheddad, J. Condell, K. Curran, and P. Mc Kevitt, "Digital image steganography: Survey and analysis of current methods," *Signal Process.*, vol. 90, no. 3, pp. 727–752, Mar. 2010.

[30] R. Singh, M. Saraswat, A. Ashok, H. Mittal, A. Tripathi, A. C. Pandey, and R. Pal, "From classical to soft computing based watermarking techniques: A comprehensive review," *Future Gener. Comput. Syst.*, vol. 141, pp. 738–754, Apr. 2023.

[31] X. Zhang, Q. Su, Y. Sun, and S. Chen, "A robust and high-efficiency blind watermarking method for color images in the spatial domain," *Multimedia Tools Appl.*, vol. 82, no. 18, pp. 27217–27243, Jul. 2023.

[32] S. Gaur and V. Barthwal, "An extensive analysis of digital image watermarking techniques," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 1, pp. 121–145, 2024.

[33] S. Sharma, J. J. Zou, G. Fang, P. Shukla, and W. Cai, "A review of image watermarking for identity protection and verification," *Multimedia Tools Appl.*, pp. 1–63, Sep. 2023.

[34] C. Song, S. Sudirman, M. Merabti, and D. Llewellyn-Jones, "Analysis of digital image watermark attacks," in *Proc. 7th IEEE Consum. Commun. Netw. Conf.*, Jan. 2010, pp. 1–5.

[35] L. Kumar Saini and V. Shrivastava, "Analysis of attacks on hybrid DWT-DCT algorithm for digital image watermarking with MATLAB," 2014, *arXiv:1407.4738*.

[36] T. K. Araghi, A. A. Manaf, and S. K. Araghi, "A secure blind discrete wavelet transform based watermarking scheme using two-level singular value decomposition," *Exp. Syst. Appl.*, vol. 112, pp. 208–228, Dec. 2018.

[37] S. Sudha and K. Rahini, "Prevention of watermarking attacks using cryptography method," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 3, no. 2, pp. 5050–5053, 2014.

[38] S. Sadeghi, S. Dadkhah, H. A. Jalab, G. Mazzola, and D. Uliyan, "State of the art in passive digital image forgery detection: Copy-move image forgery," *Pattern Anal. Appl.*, vol. 21, no. 2, pp. 291–306, May 2018.

[39] S. Sadeghi, H. A. Jalab, and S. Dadkhah, "Efficient copy-move forgery detection for digital images," *Int. J. Comput. Inf. Eng.*, vol. 6, no. 11, pp. 1339–1342, 2012.

[40] C. Wang, H. Zhang, and X. Zhou, "Review on self-embedding fragile watermarking for image authentication and self-recovery," *J. Inf. Process. Syst.*, vol. 14, no. 2, pp. 510–522, 2018.

[41] C.-C. Lin, T.-S. Nguyen, and C.-C. Chang, "LRW-CRDB: Lossless robust watermarking scheme for categorical relational databases," *Symmetry*, vol. 13, no. 11, p. 2191, Nov. 2021.

[42] F. Di Martino and S. Sessa, "Fragile watermarking tamper detection with images compressed by fuzzy transform," *Inf. Sci.*, vol. 195, pp. 62–90, Jul. 2012.

[43] M. J. Page, D. Moher, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan, and R. Chou, "PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews," *BMJ*, vol. 372, Mar. 2021. [Online]. Available: https://www.bmj.com/content/372/bmj.n160

[44] J. A. López-Sánchez, J. C. Patiño-Vanegas, A. Valencia-Arias, and J. Valencia, "Use and adoption of ICTs oriented to university student learning: Systematic review using PRISMA methodology," *Cogent Educ.*, vol. 10, no. 2, Dec. 2023, Art. no. 2288490, doi: 10.1080/2331186x.2023.2288490.

[45] M. Farrús, "Automatic speech recognition in L2 learning: A review based on prisma methodology," *Languages*, vol. 8, no. 4, p. 242, 2023. [Online]. Available: https://www.mdpi.com/2226-471X/8/4/242

[46] M. Sokouti and B. Sokouti, "A PRISMA-compliant systematic review and analysis on color image encryption using DNA properties," *Comput. Sci. Rev.*, vol. 29, pp. 14–20, Aug. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574013717300485

[47] B. Peixoto, R. Pinto, M. Melo, L. Cabral, and M. Bessa, "Immersive virtual reality for foreign language education: A PRISMA systematic review," *IEEE Access*, vol. 9, pp. 48952–48962, 2021.

[48] *Information Technology—Syntactic Metalanguage—Extended BNF*, Standard ISO/IEC 14977, Int. Org. Standardization, 1996.

[49] V. M. Potdar, S. Han, and E. Chang, "A survey of digital image watermarking techniques," in *Proc. 3rd IEEE Int. Conf. Ind. Informat.*, Aug. 2005, pp. 709–716.

[50] D. Zheng, Y. Liu, J. Zhao, and A. E. Saddik, "A survey of RST invariant image watermarking algorithms," *ACM Comput. Surv.*, vol. 39, no. 2, p. 5, Jul. 2007, doi: 10.1145/1242471.1242473.

[51] M. Prasad.R and S. Koliwad, "A comprehensive survey of contemporary researches in watermarking for copyright protection of digital images," *Int. J. Comput. Sci. Netw. Secur.*, vol. 9, no. 4, pp. 91–107, 2009.

[52] P. Singh and R. S. Chadha, "A survey of digital watermarking techniques, applications and attacks," *Int. J. Eng. Innov. Technol. (IJEIT)*, vol. 2, no. 9, pp. 165–175, 2013.

[53] H. Tao, L. Chongmin, J. Mohamad Zain, and A. N. Abdalla, "Robust image watermarking theories and techniques: A review," *J. Appl. Res. Technol.*, vol. 12, no. 1, pp. 122–138, Feb. 2014.

[54] S. M. Mousavi, A. Naghsh, and S. A. R. Abu-Bakar, "Watermarking techniques used in medical images: A survey," *J. Digit. Imag.*, vol. 27, no. 6, pp. 714–729, Dec. 2014.

[55] T. K. Araghi, A. A. Manaf, and M. Araghi, "Taxonomy and performance evaluation of feature based extraction techniques in digital image watermarking," *Int. J. Image Process. Techn.*, vol. 3, no. 1, pp. 20–23, 2016.

[56] C. Kumar, A. K. Singh, and P. Kumar, "A recent survey on image watermarking techniques and its application in e-governance," *Multimedia Tools Appl.*, vol. 77, no. 3, pp. 3597–3622, Feb. 2018.

[57] T. K. Araghi, "Digital image watermarking and performance analysis of histogram modification based methods," in *Proc. Sci. Inf. Conf.*, 2018, pp. 631–637.

[58] A. Mohanarathinam, "Digital watermarking techniques for image security: A review," *J. Ambient Intell. Humanized Comput.*, vol. 11, no. 8, pp. 3221–3229, 2020.

[59] O. Evsutin, A. Melman, and R. Meshcheryakov, "Digital steganography and watermarking for digital images: A review of current research directions," *IEEE Access*, vol. 8, pp. 166589–166611, 2020.

[60] K. J. Giri, S. M. K. Quadri, R. Bashir, and J. I. Bhat, "DWT based color image watermarking: A review," *Multimedia Tools Appl.*, vol. 79, nos. 43–44, pp. 32881–32895, Nov. 2020.

[61] A. Ray and S. Roy, "Recent trends in image watermarking techniques for copyright protection: A survey," *Int. J. Multimedia Inf. Retr.*, vol. 9, no. 4, pp. 249–270, Dec. 2020.

[62] D. K. Mahto and A. K. Singh, "A survey of color image watermarking: State-of-the-art and research directions," *Comput. Electr. Eng.*, vol. 93, Jul. 2021, Art. no. 107255. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045790621002408

[63] T. K. Araghi, A. A. Alarood, and S. K. Araghi, "Analysis and evaluation of template based methods against geometric attacks: A survey," in *Innovative Systems for Intelligent Health Informatics*, F. Saeed, F. Mohammed, and A. Al-Nahari, Eds. Cham, Switzerland: Springer, 2021, pp. 807–814.

[64] W. H. Alshoura, Z. Zainol, J. S. Teh, M. Alawida, and A. Alabdulatif, "Hybrid SVD-based image watermarking schemes: A review," *IEEE Access*, vol. 9, pp. 32931–32968, 2021.

[65] O. P. Singh, A. K. Singh, G. Srivastava, and N. Kumar, "Image watermarking using soft computing techniques: A comprehensive survey," *Multimedia Tools Appl.*, vol. 80, no. 20, pp. 30367–30398, Aug. 2021.

[66] S. Wadhera, D. Kamra, A. Rajpal, A. Jain, and V. Jain, "A comprehensive review on digital image watermarking," in *Proc. Workshop Comput. Netw. Commun.*, vol. 2889, G. Ganesan, R. Saha, and G. K. Sharma, Eds. 2021, pp. 126–143.

[67] O. Evsutin and K. Dzhanashia, "Watermarking schemes for digital images: Robustness overview," *Signal Process., Image Commun.*, vol. 100, Jan. 2022, Art. no. 116523. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0923596521002551

[68] S. Bhattacharya, T. Chattopadhyay, and A. Pal, "A survey on different video watermarking techniques and comparative analysis with reference to H.264/AVC," in *Proc. IEEE Int. Symp. Consum. Electron.*, Jun. 2006, pp. 1–6.

[69] T. Jayamalar and V. Radha, "Survey on digital video watermarking techniques and attacks on watermarks," *Int. J. Eng. Sci. Technol.*, vol. 2, no. 12, pp. 6963–6967, 2010.

[70] X. Chang, W. Wang, J. Zhao, and L. Zhang, "A survey of digital video watermarking," in *Proc. 7th Int. Conf. Natural Comput.*, vol. 1, 2011, pp. 61–65.

[71] N. A. Shelke and P. N. Chatur, "A survey on various digital video watermarking schemes," *Int. J. Comput. Sci. Eng. Technol.*, vol. 4, no. 12, pp. 1447–1454, 2013.

[72] P. a. S. Sethuraman and R. Srinivasan, "Survey of digital video watermarking techniques and its applications," *Eng. Sci.*, vol. 1, no. 1, pp. 22–27, 2016.

[73] D. Pathak and A. Jain, "A survey on video watermarking features and techniques," *Int. J. Sci. Eng. Res.*, vol. 7, no. 6, pp. 390–394, 2016.

[74] M. Asikuzzaman and M. R. Pickering, "An overview of digital video watermarking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 9, pp. 2131–2153, Sep. 2018.

[75] X. Yu, C. Wang, and X. Zhou, "A survey on robust video watermarking algorithms for copyright protection," *Appl. Sci.*, vol. 8, no. 10, p. 1891, Oct. 2018. [Online]. Available: https://www.mdpi.com/2076-3417/8/10/1891

[76] R. Artru, A. Gouaillard, and T. Ebrahimi, "Digital watermarking of video streams: Review of the state-of-the-art," 2019, *arXiv:1908.02039*.

[77] T. N. Hummadia and N. F. Hassanb, "Survey of recent video watermarking techniques," *Eng. Technol. J.*, vol. 39, no. 1B, pp. 165–174, Mar. 2021.

[78] M. A. T. Alsalami and M. M. Al-Akaidi, "Digital audio watermarking: Survey," in *Proc. 17th Eur. Simul. Multiconf.*, 2003.

[79] S. P. Singh Chauhan and S. A. M. Rizvi, "A survey: Digital audio watermarking techniques and applications," in *Proc. 4th Int. Conf. Comput. Commun. Technol. (ICCCT)*, Sep. 2013, pp. 185–192.

[80] M. A. Nematollahi and S. A. R. Al-Haddad, "An overview of digital speech watermarking," *Int. J. Speech Technol.*, vol. 16, no. 4, pp. 471–488, Dec. 2013.

[81] J. Bajpai and A. Kaur, "A literature survey—Various audio watermarking techniques and their challenges," in *Proc. 6th Int. Conf.-Cloud Syst. Big Data Eng.*, 2016, pp. 451–457.

[82] R. Nayyar and R. Singh, "A review on improved audio watermarking methods," *Int. Res. J. Eng. Technol. (IRJET)*, vol. 3, no. 7, pp. 1832–1838, 2016.

[83] G. Hua, J. Huang, Y. Q. Shi, J. Goh, and V. L. L. Thing, "Twenty years of digital audio watermarking—A comprehensive review," *Signal Process.*, vol. 128, pp. 222–242, Nov. 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165168416300263

[84] R. Jain, M. C. Trivedi, and S. Tiwari, "Digital audio watermarking: A survey," in *Advances in Computer and Computational Sciences*, S. K. Bhatia, K. K. Mishra, S. Tiwari, and V. K. Singh, Eds. Singapore: Springer, 2018, pp. 433–443.

[85] N. Lalitha, C. S. Rao, and P. J. Sree, "A rewiev of digital audio watermarking schemes," *J. Crit. Rev.*, vol. 7, no. 7, pp. 870–880, 2020.

[86] S.-J. Lee and S.-H. Jung, "A survey of watermarking techniques applied to multimedia," in *Proc. IEEE Int. Symp. Ind. Electron.*, vol. 1, Jun. 2001, pp. 272–277.

[87] D. Megías, M. Kuribayashi, and A. Qureshi, "Survey on decentralized fingerprinting solutions: Copyright protection through piracy tracing," *Computers*, vol. 9, no. 2, p. 26, Apr. 2020. https://www.mdpi.com/2073-431X/9/2/26

[88] N. Agarwal, A. K. Singh, and P. K. Singh, "Survey of robust and imperceptible watermarking," *Multimedia Tools Appl.*, vol. 78, no. 7, pp. 8603–8633, Apr. 2019.

[89] P. Meerwald and A. Uhl, "Survey of wavelet-domain watermarking algorithms," *Proc. SPIE*, vol. 4314, pp. 505–516, Aug. 2001, doi: 10.1117/12.435434.

[90] R. Xie, K. Wu, J. Du, and C. Li, "Survey of public key digital watermarking systems," in *Proc. 8th ACIS Int. Conf. Software Eng., Artif. Intell., Netw., Parallel/Distrib. Comput.*, vol. 2, 2007, pp. 439–443.

[91] A. Khan, A. Siddiqa, S. Munib, and S. A. Malik, "A recent survey of reversible watermarking techniques," *Inf. Sci.*, vol. 279, pp. 251–272, Sep. 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025514004150

[92] M. M. Pal, "A survey on digital watermarking and its application," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 1, pp. 153–156, 2016.

[93] R. R. Kumari, V. Vijaya, and K. R. Naidu, "Existing trends of digital watermarking and its significant impact on multimedia streaming: A survey," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 2, pp. 126–139, 2019.

[94] C. Rey and J.-L. Dugelay, "A survey of watermarking algorithms for image authentication," *EURASIP J. Adv. Signal Process.*, vol. 2002, no. 6, pp. 613–621, Jun. 2002.
[95] X.-L. Liu, C.-C. Lin, C.-C. Chang, and S.-M. Yuan, "A survey of fragile watermarking-based image authentication techniques," *J. Inf. Hiding Multimedia Signal Process.*, vol. 7, no. 6, pp. 1282–1292, 2016.
[96] X. Yu, C. Wang, and X. Zhou, "Review on semi-fragile watermarking algorithms for content authentication of digital images," *Future Internet*, vol. 9, no. 4, p. 56, Sep. 2017. [Online]. Available: https://www.mdpi.com/1999-5903/9/4/56
[97] K. Sreenivas and V. Kamkshi Prasad, "Fragile watermarking schemes for image authentication: A survey," *Int. J. Mach. Learn. Cybern.*, vol. 9, no. 7, pp. 1193–1218, Jul. 2018.
[98] C.-F. Lee, J.-J. Shen, and F.-W. Hsu, "A survey of semi-fragile watermarking authentication," in *Recent Advances in Intelligent Information Hiding and Multimedia Signal Processing*, J.-S. Pan, A. Ito, P.-W. Tsai, and L. C. Jain, Eds. Cham, Switzerland: Springer, 2019, pp. 264–271.
[99] N. R. N. Raj and R. Shreelekshmi, "A survey on fragile watermarking based image authentication schemes," *Multimedia Tools Appl.*, vol. 80, no. 13, pp. 19307–19333, May 2021.
[100] A. Anand and A. K. Singh, "Watermarking techniques for medical data authentication: A survey," *Multimedia Tools Appl.*, vol. 80, no. 20, pp. 30165–30197, Aug. 2021.
[101] R. Ghafoor, D. Saleem, S. S. Jamal, M. Ishtiaq, S. Ejaz, A. Jamal Malik, and M. F. Khan, "Survey on reversible watermarking techniques of echocardiography," *Secur. Commun. Netw.*, vol. 2021, pp. 1–19, Mar. 2021.
[102] L. Rakhmawati, W. Wirawan, and S. Suwadi, "A recent survey of self-embedding fragile watermarking scheme for image authentication with recovery capability," *EURASIP J. Image Video Process.*, vol. 2019, no. 1, pp. 1–22, Dec. 2019.
[103] A. Menendez-Ortiz, C. Feregrino-Uribe, R. Hasimoto-Beltran, and J. J. Garcia-Hernandez, "A survey on reversible watermarking for multimedia content: A robustness overview," *IEEE Access*, vol. 7, pp. 132662–132681, 2019.
[104] P. F. Alcantarilla, A. Bartoli, and A. J. Davison, "Kaze features," in *Computer Vision—ECCV*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Berlin, Germany: Springer, 2012, pp. 214–227.
[105] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[106] D. Kirovski and H. S. Malvar, "Spread-spectrum watermarking of audio signals," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 1020–1033, Apr. 2003.
[107] D. Gruhl, A. Lu, and W. Bender, "Echo hiding," in *Information Hiding*, R. Anderson, Ed. Berlin, Germany: Springer, 1996, pp. 295–315.
[108] X. Wang, W. Qi, and P. Niu, "A new adaptive digital audio watermarking based on support vector regression," *IEEE Trans. Audio, Speech Language Process.*, vol. 15, no. 8, pp. 2270–2277, Nov. 2007.
[109] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1423–1443, May 2001.
[110] H. Allcott and M. Gentzkow, "Social media and fake news in the 2016 election," *J. Econ. Perspect.*, vol. 31, no. 2, pp. 211–236, May 2017. [Online]. Available: https://www.aeaweb.org/articles?id=10.1257/jep.31.2.211
[111] S. Rathore, P. K. Sharma, V. Loia, Y.-S. Jeong, and J. H. Park, "Social network security: Issues, challenges, threats, and solutions," *Inf. Sci.*, vol. 421, pp. 43–69, Dec. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0020025517309106
[112] R. Jarusek, E. Volna, and M. Kotyrba, "Photomontage detection using steganography technique based on a neural network," *Neural Netw.*, vol. 116, pp. 150–165, Aug. 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608019300966
[113] A. Alattar, R. Sharma, and J. Scriven, "A system for mitigating the problem of deepfake news videos using watermarking," *Electron. Imag.*, vol. 32, no. 4, pp. 117–1, Jan. 2020. [Online]. Available: https://www.ingentaconnect.com/content/ist/ei/2020/00002020/00000004/art00011
[114] N. Alaa, "Watermarking images for fact-checking and fake news inquiry," B.Sc thesis, Fac. Manag. Technol., Bus. Inform. Dept., German Univ. Cairo, New Cairo City, Egypt, 2021. [Online]. Available: https://www.researchgate.net/publication/352772064_Watermarking_Images_for_Fact-Checking_and_Fake_News_Inquiry
[115] D. Megías, M. Kuribayashi, A. Rosales, and W. Mazurczyk, "DISSIMILAR: Towards fake news detection using information hiding, signal processing and machine learning," in *Proc. 16th Int. Conf. Availability, Rel. Secur.* New York, NY, USA: Association for Computing Machinery, Aug. 2021, pp. 1–9, doi: 10.1145/3465481.3470088.
[116] R. Wang, F. Juefei-Xu, M. Luo, Y. Liu, and L. Wang, *FakeTagger: Robust Safeguards against DeepFake Dissemination via Provenance Tracking*. New York, NY, USA: Association for Computing Machinery, 2021, pp. 3546–3555, doi: 10.1145/3474085.3475518.
[117] A. Qureshi, D. Megías, and M. Kuribayashi, "Detecting deepfake videos using digital watermarking," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Dec. 2021, pp. 1786–1793.
[118] P. Neekhara, S. Hussain, X. Zhang, K. Huang, J. McAuley, and F. Koushanfar, "FaceSigns: Semi-fragile neural watermarks for media authentication and countering deepfakes," 2022, *arXiv:2204.01960*.

AGNIESZKA MALANOWSKA received the B.Sc. and M.Sc. degrees in computer science from Warsaw University of Technology, Poland, in 2017 and 2019, respectively. Since 2019, she has been a Research and Teaching Assistant with the Institute of Computer Science, Division of Software and Computer Architecture, Warsaw University of Technology. Her main research interests include software engineering, particularly UML modeling, effort estimation, and software testing. As a Teacher, she conducts classes in object-oriented programming, software engineering, and compiling techniques in Polish and English. She is the coauthor of six articles.

WOJCIECH MAZURCZYK (Senior Member, IEEE) received the B.Sc., M.Sc., Ph.D. (Hons.), and D.Sc. (Habilitation) degrees in telecommunications from Warsaw University of Technology (WUT), Warsaw, Poland, in 2003, 2004, 2009, and 2014, respectively. He is currently an University Professor with the Institute of Computer Science, WUT. He is also a Researcher with the Parallelism and VLSI Group, Faculty of Mathematics and Computer Science, FernUniversität, Germany. His research interests include bioinspired cybersecurity and networking, information hiding, and network security. He is involved in the technical program committee of many international conferences and is a reviewer of major international magazines and journals. Since 2016, he has been the Editor-in-Chief of an open access *Journal of Cyber Security and Mobility*. From 2018 to 2020, he was a Mobile Communications and Networks Series Editor and a Technical Associate Editor of *IEEE Communications Magazine*, from 2013 to 2018. Between 2018 and 2021, he was also served as an Associate Editor for IEEE Transactions on Information Forensics and Security. Has been indicated three times as one of the Top 2% of Scientists of their main subfield discipline in the Stanford List of Top Scientists, in 2019, 2020, 2021, and 2022.

TANYA KOOHPAYEH ARAGHI received the Ph.D. degree (Hons.) from Universiti Teknologi Malaysia (UTM), in 2017. She has recorded three copyrights for intellectual property in the country of Malaysia under protection of UTM and MYIPO (Malaysia), from January 2017 to September 2017. She was selected for two consecutive years as the best Researcher with Iranian Social Security Organization and got the Best Researcher Award from the Department of Employment and Social Affairs, in January 2018, and again awarded the Best Researcher Prize from the Iranian Department of Co-Operation and Social Welfare, in January 2019. Currently, she is a Postdoctoral Researcher with Universitat Oberta de Catalunya. Since 2021, she has been participating in several national and international projects in Spain. Her research interests include watermarking, data hiding, multimedia security, and wireless network security. She received the Best Post Graduate Award for the Ph.D. degree.

DAVID MEGÍAS (Member, IEEE) received the Ph.D. degree in computer science from Universitat Autònoma de Barcelona (UAB), in July 2000. He is currently a Full Professor and a Principal Investigator of the KISON Research Group, Internet Interdisciplinary Institute (IN3), Universitat Oberta de Catalunya (UOC). Since October 2001, he has been with UOC with a permanent position (currently as a Professor). At UOC, he has held several academic positions, until he was appointed as the Director of IN3, in April 2015. His current teaching is mostly related to computer networks, information security (watermarking and steganography), and research techniques and methodologies in the field of network and information technologies. He has published more than 130 research papers in numerous international journals and conferences, 40 of them in journals indexed in JCR, and has participated in several national joint research projects both as a contributor and as a principal investigator. He has supervised four doctoral theses and is a member of the editorial board and programme committees of several journals and conferences in the area of security and privacy. His current research interests include information security and privacy, the security and privacy in multimedia content distribution (mainly in the watermarking and fingerprinting topics), steganography and steganalysis, and privacy concerns in different applications of decentralized networks.

MINORU KURIBAYASHI (Senior Member, IEEE) received the B.E., M.E., and D.E. degrees from Kobe University, Japan, in 1999, 2001, and 2004, respectively. He was a Research Associate and an Assistant Professor with Kobe University, from 2002 to 2007 and from 2007 to 2015, respectively. He was an Associate Professor with the Graduate School of Natural Science and Technology, Okayama University, from 2015 to 2023. Since 2023, he has been a Professor with the Center for Data-Driven Science and Artificial Intelligence, Tohoku University. His research interests include multimedia security, digital watermarking, cryptography, and coding theory. He has published more than 150 research papers in numerous international journals and conferences. He has been serving as an Associate Editor for IEEE Signal Processing Letters, since 2022, and *Journal of Information Security and Applications*, from 2014 to 2021. He is the Chair of APSIPA TC of Multimedia Security and Forensics and a TC Member of IEEE SPS Information Forensics and Security. He received the Young Professionals Award from IEEE Kansai Section, in 2014, and the Best Paper Award from IWDW, in 2015 and 2019.

• • •