Digital Watermarking and Perceptual Hashing of Audio Signals with Focus on their Evaluation

Andreas Lang, Jana Dittmann, Christian Krätzer
Research Group Multimedia and Security,
Department of Computer Science,
Otto-von-Guericke-University of Magdeburg, Germany

Abstract. Digital watermarking is a growing research area to mark digital content by embedding information into the content itself. Perceptual hashing is used to identify a specific content or to identify integrity violations up to a specific threshold. The evaluation of watermarking algorithms provides a fair and automated analysis of specific watermarking schemes for selected application fields. In this paper, we present a theoretical framework, the design and formalization of an evaluation profile especially for perceptual hashing algorithms. Based on this profile, on one hand, the transparency of the digital watermarking scheme or on the other hand the robustness of the perceptual hashing function can be evaluated. The introduction of practical tests and their setup show the benefit of the evaluation profile.

1 Motivation and Introduction

Mostly, the development of digital watermarks is directly connected to their evaluation. In literature different strategies are introduced to provide a widely applicable and comparable evaluation of watermarking algorithms [2–4]. The existing evaluation techniques differ in their strategy and procedure. For example, a simple evaluation procedure for robustness can be provided by using single attacks with following detection or retrieval of the embedded information. If the watermarking schemes should be evaluated in the context of application scenarios, profiles are more realistic and are of higher use for the evaluation. With profiles, the evaluation is easier, abstracted and useable for the developer as well as for end users with no or only few limited inside knowledge on watermarking techniques. Therefore, application profiles have model typical application scenarios. Furthermore, the comparability of given watermarking schemes can be analyzed for a specific application scenario or a subset of application fields. In this paper, the application profile Perceptual Hashing $P_{E/A}$ is introduced, designed and formalized especially to evaluate perceptual hashing algorithms. Thereby, the influence of the embedding function of a watermarking

---

1 Note, we use the same notation as introduced in [3]. Thereby “E/A” means, that the embedding and/or attacking function is addressed, whereby “E” addresses only the embedding and “A” only the attacking function.
scheme to the perceptual hash as well as the changed perceptual hash as result of the embedded watermark within an application scenario are analyzed and discussed.

The bases for all evaluation techniques and evaluation profiles are standardized measurements of one or more specific properties or characteristics. Seven basic profiles are introduced in [4], defining the single characteristics of a digital watermarking scheme. These elementary properties are *transparency*, *capacity*, *robustness*, *complexity*, *invertible*, *verification* and *security*. Derived from the introduction of the single properties in [5] their formalization and measurement is defined and presented. Furthermore test results are presented, which compare different watermarking schemes with each other.

A perceptual hash is also known as content fingerprinting or content based identification [1, 7]. Considering the example of audio content, the principle of such a perceptual hash is, that acoustically relevant characteristics of a piece of audio material are identified and the computed features are stored in a database. When an unknown piece of audio signal is presented, then the same features are computed and matched against those stored in the database. The authors in [8] split the audio signal into $n$ frames. The frequency domain representation of each frame is divided in the range 300Hz to 3000Hz into 32 frequency bands $m$ to compute a 32 bit content hash value $H(n,m)$. In this paper, we do not consider framing of the audio signal for simplification reasons. In our application scenarios and the watermark evaluation, the perceptual hash $H(S)$ is computed from the complete audio signal $S$. Many applications for perceptual hashes exist, some of them are summarized in [7]. The following itemization summarizes some of the most common of these application scenarios:

- **The Identification** is used, to identify an unknown audio signal, for example a piece of music and to return media data like the author and title. This perceptual hash is also known as secure perceptual hash. In this process, the perceptual hash is computed from the known audio signal $H(S)$ and stored in the database. If an unknown or modified audio signal $S'$ is presented, then $H(S')$ is computed and compared with the hash values stored in the database to identify $S'$. Exact match or threshold based comparison identifies a specific or known audio signal.

- **The Integrity Verification** is an application, where the perceptual hashes are used to detect an alteration of the audio signal. If the original audio signal $S$ is modified ($S'$) by a malicious or non-malicious function and if as a result $H(S) \neq H(S')$ is computed, then this alteration is detectable by using a fragile perceptual hashing function. If the computation of the perceptual hash is chosen with the knowledge of possible audio modifications, the perceptual hash could be designed to allow certain modifications and could be fragile for others.

- **As Watermark Support**, perceptual hashes can be used, to compute the required secret key for the embedding and detection/retrieval function of the
fragile audio watermark from the audio signal itself [10]. If an alteration of
the marked audio signal occur, then the perceptual hash differs, the detec-
tion/retrieval fails and the alteration or tampering is detected.

– In the application field of Monitoring, which is a more specific application
of the identification approach introduced above, the perceptual hash is used
to identify transmitted content via radio or TV. Thereby, the content dis-
tributer has an easy way to know, when by whom content was broadcasted.
For the content owner it is important, because he needs to know, if the
distributer has the right to broadcast or not or whether commercials are
broadcasted or not.

For the different application scenarios, different perceptual hashing algorithms
with different required parameter sets exists. In [9] are three different perceptual
hashing algorithms are compared with each other. The authors compute $H^*(S)$,
whereby $S$ is the original audio signal and $H^*$ is one of the perceptual hash
functions. Then, the audio signal $S$ is modified by 11 different kinds of attacks
and from the resulting modified audio signals, the perceptual hashes are com-
puted $H^*(S')$. A comparison of $H^*(S)$ with $H^*(S')$ identifies the robustness of
$H^*$ against a specific attack.

The evaluation presented in [9] is focused on the perceptual hashing algorithms
and their robustness against selected attacks. The authors use a non-standardized
evaluation strategy which makes it difficult for others to compare their own test
results with those presented in [9]. To bridge this gap, we introduce profile based
evaluation based on established evaluation strategies and we focus in this paper
on the profile based robustness and fragility evaluation of perceptual hashes.

This paper is organized as follows: Section 2 introduces the application profile.
Thereby, in subsection 2.1, the transparency measure for the embedding function
is defined and formalized. In subsection 2.2 the attacking function as part of the
digital watermarking scheme is in focus and its transparency measure defined.
Based on both definitions, in subsection 2.3 the transparency evaluation of a
digital watermarking scheme in context of perceptual hashing as well as the
robustness evaluation of a perceptual hashing function with a digital watermark
scheme are introduced. The section 3 summarizes our approach and its impacts
and draws conclusion for future work.

2 Application Profile: Design and Formalization

In this section, the application profile $P_{E/A-PerceptualHash}$ for the evaluation of
the embedding and attacking function of a digital watermarking scheme and the
evaluation of a perceptual hash function with a digital watermarking scheme are
designed and introduced.

The usage of a digital watermarking scheme and its general processes can be
simplified as follows. An unmarked (mostly original) signal ($S$) is the source
signal, where the watermark \( w \) is embedded by using an embedding function \( E \). The result of it is the marked signal \( S_E \). It can be defined, that this process is done in a secure environment. The following step could be for example the distribution of \( S_E \) over the Internet or storage of it to provide authenticity or integrity checks. These processes can been seen as an insecure part, where attacks \( (A_{i,j} \in A) \) occur on \( S_E \). After distribution of \( S_E \), the signal is defined as \( S_{EA} \) because potential attacks could destroy or modify the watermark. A detecting function \( D \) tries to detect the watermark \( w \) or a retrieval function \( R \) tries to retrieve the embedded message \( m' \). The detection/retrieval can be done in a secure or insecure environment, depending on the used application of the watermarking algorithm.

The complete introduced scenario is also called life cycle of a watermark, because it begins with embedding and ends with detection/retrieval. The following Figure 1 introduces this life cycle and shows, where the secure and insecure parts are expected.

![Fig. 1: General Embedding-, Attacking- and Detecting Functions](image)

The usage of a watermarking algorithm and the classification from the audio signal point of view in the above summarized embedding \( E \), attacking \( A \) and detection \( D \), retrieval \( R \) functions is the motivation for an evaluation with the watermarking algorithm perspective. Therefore, the three main functions\(^2\) \( (E, A \) and \( D,R) \) are used to provide a classification of evaluation processes. Note, that the functions, shown in Figure 1 require parameters to setup and control the function internals.

In general, the evaluation within the application field of perceptual hashes can be considered from two completely different points of view. On one hand, one could focus on the evaluation of a digital watermarking scheme, whereby the perceptual hashing function is used to measure specific properties of the watermarking

\(^2\) There are three main functions, because the detection and retrieval function \( D \) and \( R \) are seen as one function
algorithm. On the other hand, properties of the perceptual hashing function can be evaluated with a digital watermarking scheme and its embedding and/or attacking function. Derived from both points of view and the motivation of enhancing the application evaluation profiles, the application profile definition is introduced and formalized.

When evaluating a digital watermarking scheme with focus on an application scenario using perceptual hashes, the following questions arise:

1. How can an evaluation of given watermarking schemes be done and performed providing comparability between the different schemes?

2. How can an assignment of the best fitting candidates for a given application scenario be achieved?

3. Does the embedding function of a digital watermarking scheme with the given embedding parameter set \( p_E \in \mathcal{P}_E \) change the perceptual hash?

4. Does the attacking function with the given attacking parameter set \( p_A \in \mathcal{P}_A \) change the perceptual hash?

Note, that, derived from these questions, different evaluation methods for the different functions are required. Based on the profile based evaluation technique introduced in [2–4], we firstly focus on the embedding function of a digital watermarking scheme and secondly on the attacking function by evaluating them within the application field of perceptual hashing. In the literature [5, 6] is the definition and formalization of digital watermarking schemes and their embedding-, attacking as well as detection/retrieval function introduced. We use the same notation here.

### 2.1 Evaluation of the Embedding Function with a Perceptual Hash

The embedding of a message \( m \) into the cover signal \( S \in \mathcal{S} \) by using an embedding function \( E \) of a digital watermarking scheme, results in signal modifications within the marked signal \( S_E \). This can be the reason, that a perceptual hash, computed from the cover signal \( S \) changes after embedding the digital watermark and the changes occurred by embedding needs to be evaluated. Additionally, the selected embedding parameter set \( p_E \in \mathcal{P}_E \) has an effect on the marked signal \( S_E \) and therefore on its perceptual hash.

In this case, the evaluation of the embedding function \( E \) is done within the application field of perceptual hashes. The goal thereby is to measure the impact of the embedding function on the perceptual hash. The classification of this evaluation scenario is associated to the embedding profiles already introduced in [3] and defined as follows:

\[
P_{E-\text{Perceptual Hash}} = (S_{\text{IN}} \parallel S_{\text{OUT}} \parallel \text{param})
\]

\[
\text{param} = (\text{alg} \parallel \text{alg-opt} \parallel \text{hashalg} \parallel \text{hashalg-opt})
\]
The parameter $S_{IN}$ defines the input audio signal ($S \in S$) and the parameter $S_{OUT}$ defines the marked output audio signal $S_E$. The parameter “alg” defines the watermarking scheme $\Omega$, which is evaluated with its parameters defined with “alg-opt”. Note, that in this case “alg-opt”=$p_E \in P_E$. The used perceptual hashing function is defined with “hashalg” and its required parameters with “hashalg-opt”.

The internals of this application profile are the measurement of the transparency $T$ between $S$ and $S_E$ in the closed interval $[0,1]$ where 0 provides the worst case ($S$ and $S_E$ are so different that $S_E$ cannot be recognized as a modified version of $S$) and 1 is the best case (an observer does not perceive any significant difference between $S_E$ and $S$).

$$T(S, S_E) \rightarrow [0,1] \quad (3)$$

The perceptual hash can be seen as transparency measure of $E$, whereby the relative embedding transparency of a particular audio signal $S$, watermarking scheme $\Omega^*$ and perceptual hashing function $H^*$ are defined as:

$$tra_{E,rel}(\Omega^*, H^*, S) = T(S, S_E), \quad (4)$$

where $S_E$ is obtained by applying the embedding function. However, this definition of transparency is related to a particular object $S$. It is usually better to provide some absolute value of transparency which is not related to a particular object $S$. A definition of “absolute” transparency is related to a family $S$ of objects to be marked and we could apply any of the following definitions:

- Average transparency:
  $$tra_{E,ave}(\Omega^*, H^*) = \frac{1}{|S|} \sum_{S \in S} tra_{E,rel}(\Omega^*, H^*, S). \quad (5)$$

- Maximum transparency:
  $$tra_{E,max}(\Omega^*, H^*) = \max_{S \in S} \{tra_{E,rel}(\Omega^*, H^*, S)\}. \quad (6)$$

- Minimum transparency:
  $$tra_{E,min}(\Omega^*, H^*) = \min_{S \in S} \{tra_{E,rel}(\Omega^*, H^*, S)\}. \quad (7)$$

The goal of the embedding profile to evaluate the effect of the transparency of the embedding function of a given digital watermarking scheme $\Omega^*$ on the perceptual hash is visualized in Figure 2. After selecting the cover signal $S \in S$ and the embedding parameter set $p_E \in P_E$, the general principle is as follows:

1. Compute the perceptual hash value from $S$ and store it in the data base $(H(S))$. 
2. Embed the message $m$ into $S$ with the embedding function $E$ and its required parameters $p_E$. The result is the marked signal $S_E$.

3. Compute the perceptual hash value from $S_E$ ($H(S_E)$).

4. Compare $H(S)$ with $H(S_E)$ and if $H(S) = H(S_E)$, then $E$ and the selected parameters $p_E$ do not effect the perceptual hash. Otherwise, if $H(S) \neq H(S_E)$, then $E$ and/or $p_E$ result in audio signal modifications, that change the perceptual hash.

$$\text{tra}_{E_{\text{rel}}} = \begin{cases} 0, & H(S) \neq H(S_E) \\ 1, & H(S) = H(S_E) \end{cases}$$ \hspace{1cm} (8)

**Fig. 2:** Embedding Profile $P_{E-\text{PerceptualHash}}$ for the Evaluation of the Embedding Function of a digital Watermarking Scheme

### 2.2 Evaluation of the Attacking Function with a Perceptual Hash

The evaluation of the attacking function $A$ of a digital watermarking scheme is alike the evaluation of the embedding function introduced in the previous subsection. The differences are, that instead of $E$, the attacking function $A$ modifies the audio signal and produces the attacked audio signal $S_A$. Derived from this modification, the attacking transparency of $A$ is measured to identify the impact on the perceptual hash. The formal definition of this attacking profile is given as follows:

$$P_{A-\text{PerceptualHash}} = (S_N || S_{OUT} || \text{param})$$ \hspace{1cm} (9)

$$\text{param} = (\text{alg} || \text{alg-opt} || \text{hashalg} || \text{hashalg-opt})$$ \hspace{1cm} (10)

The parameter $S_N$ defines the input audio signal and the parameter $S_{OUT}$ defines the marked output audio signal. The parameter “alg” defines the attacking
function $A \in A$, which is evaluated with its parameters $p_A \in P_A$ defined with “alg-opt”. The used perceptual hashing function is defined with “hashalg” and its required parameters with “hashalg-opt”.

The relative, average, minimum and maximum attacking transparency measurement is similar to the embedding transparency and defined as follows:

- Relative transparency:
  \[
  \text{tra}_A^{rel}(\Omega^*, H^*, S') = T(S', S_A),
  \]  
  \[\text{(11)}\]

- Average transparency:
  \[
  \text{tra}_A^{ave}(\Omega^*, H^*) = \frac{1}{|S|} \sum_{S \in S} \text{tra}_A^{rel}(\Omega^*, H^*, S').
  \]  
  \[\text{(12)}\]

- Maximum transparency:
  \[
  \text{tra}_A^{max}(\Omega^*, H^*) = \max_{S' \in S} \{\text{tra}_A^{rel}(\Omega^*, H^*, S')\}.
  \]  
  \[\text{(13)}\]

- Minimum transparency:
  \[
  \text{tra}_A^{min}(\Omega^*, H^*) = \min_{S' \in S} \{\text{tra}_A^{rel}(\Omega^*, H^*, S')\}.
  \]  
  \[\text{(14)}\]

The usage of this attacking profile to evaluate the transparency of the attacking function with the perceptual hash is visualized in Figure 3. After selecting the cover signal $S' \in S$ and the attacking parameter set $p_A \in P_A$, the general principle is as follows:

1. Compute the perceptual hash value from $S'$ and store it in the data base ($H(S')$).

2. Attack $S'$ with the selected attack and its parameters $p_A$. The result is the attacked signal $S_A$.

3. Compute the perceptual hash value from $S_A$ ($H(S_A)$).

4. Compare $H(S')$ with $H(S_A)$ and if $H(S') = H(S_A)$, then $A$ and the selected parameters $p_A$ do not effect the perceptual hash. Otherwise, if $H(S') \neq H(S_A)$, then $A$ and/or $p_A$ result in such audio signal modifications, that the perceptual hash changes.

\[
\text{tra}_A^{rel} = \begin{cases} 
0, & H(S') \neq H(S_A) \\
1, & H(S') = H(S_A) 
\end{cases}
\]  
\[\text{(15)}\]
If this scenario is slightly modified, then it can be used to evaluate the robustness of a given perceptual hashing algorithm like done in [9] against specific selected attacks or specific attack parameters. Thereby, different attacks are selected and the perceptual hashes \( H(S') \) and \( H(S_A) \) are computed and compared. If these hash values are equal, then the perceptual hashing function is robust against the specific attack \( A \). Otherwise, if the attack parameters \( p_A \) are slightly changed, then the threshold of secure perceptual hash and fragile perceptual hash can be identified.

2.3 Evaluation of a Digital Watermarking Scheme with Perceptual Hashes and Vice Versa

The evaluation of a digital watermarking algorithm in the application field of a perceptual hashing function can be done with the embedding and attack evaluation steps introduced in the subsections above. Derived from them, two general evaluation goals exist.

- The first evaluation goal can be the evaluation of a digital watermarking scheme within the application field of perceptual hashing. Thereby, the computed perceptual hash is used as transparency measurement of the embedding and/or attacking function.

- The second test goal can be the robustness evaluation of the perceptual hashing function by embedding a digital watermark or attacking the audio signal with different attacks, which represent malicious or non-malicious signal modifications.

For both different evaluation goals, the following Figure 4 visualizes the evaluation scenario. In general, the embedding function \( E \) embeds the digital wa-
termark with the given parameter set \( p_E \in P_E \) into the audio signals \( S \in \mathcal{S} \). For the input and output audio signal, the corresponding perceptual hashes \( h^1 = H^1(S) \) and \( h^2 = H^2(S_E) \) are computed and stored in the database. After, for example, distributing the marked signal \( S_E \), different signal modifications occur, simulated with the attack function \( A \) and its required parameter set \( p_A \in P_A \). From the marked, attacked signal \( S_{EA} \) is also a perceptual hash computed \( h^3 = H^3(S_{EA}) \) and stored in the database. The comparison of the perceptual hash values stored within the database is used to evaluate the transparency of the embedding and/or attacking function of the digital watermarking scheme, or the robustness of the perceptual hashing function. The computed results are the evaluation test results and can be used for a recommendation of the watermarking scheme within the application of the selected perceptual hashing function with the used parameter and test sets. Finally, detection/retrieval functions \( D \) and \( R \) tries to detect and/or retrieve the watermark to verify the successfully embedding of \( m \) into the given signal \( S \in \mathcal{S} \).

Fig. 4: Example of the Application Profile “Perceptual Hash” \( P_{A−\text{Perceptual\_Hash}} \)

Depending on the used application scenario with perceptual hashes and its derived application goals, a digital watermarking scheme can be evaluated to get a recommendation or not. For the evaluation itself, the application profile \( P_{E/A−\text{Perceptual\_Hash}} \) is used and defined as follows:

\[
P_{A−\text{Perceptual\_Hash}}(S_{IN} || S_{OUT} || \text{param}) \tag{16}
\]

\[
\text{param} = (\text{alg} || \text{alg-opt} || \text{hashalg} || \text{hashalg-opt} || \text{att} || \text{att-opt}) \tag{17}
\]

The parameter \( S_{IN} \) is the input audio signal \( (S \in \mathcal{S}) \) and the parameter \( S_{OUT} \) defines the resulting output audio signal \( S_{EA} \). The parameter “alg” defines the
watermarking scheme, which is evaluated with its parameters defined with “alg-opt”. The used perceptual hashing function is defined with “hashalg” and its required parameters with “hashalg-opt”. To compute the hash values $h_1$, $h_2$ and $h_3$, always the same hash function “hashalg” is used. The attack, to modify the marked signal is defined within “att” and its required parameter set in “att-opt”.

Depending on the required application goals, the evaluation steps are introduced in the following listing. Firstly, we focus on the transparency evaluation of the embedding and attacking function of the digital watermarking scheme and secondly on the robustness of the perceptual hashing function.

- If the transparency of the embedding function $E$ and/or attack function $A$ is evaluated, then the general principle can be seen separately as introduced in the previous subsections. The embedding transparency $tra_E$ and/or attacking transparency $tra_A$ as defined above are computed.

- If the robustness of the perceptual hashing function is evaluated, then the signal modification occurred by embedding a digital watermark or attacking a marked signal are seen as “robustness attack” against the perceptual hash. Thereby, the “attack” goal is to change the signal with, or without changing the perceptual hash depending on the application scenario. With the predefined transparency measurements $tra_E$ and $tra_A$ the robustness measurement is seen as follows:

  The perceptual hash is robust against a specific embedded watermark or attack, if the transparency measurement is equal to 1.0 ($tra_E = 1.0$ or $tra_A = 1.0$ independent of the relative or average transparency). Otherwise, the perceptual hash is not robust against the embedded watermark with the selected embedding parameter set or the selected attack with its parameter set.

3 Summary

The evaluation of digital watermark provides detailed information about the properties or useable application scenarios. The evaluation itself is classified into different classes of profiles, which provide better comparability and usage. In this paper, we introduced a new application profile focused on perceptual hashes, which can be used twofold. Either, as new transparency measure for the embedding or attacking function of a digital watermarking scheme, or as robustness measure for the perceptual hashing function. For evaluation the embedding transparency, the perceptual hash is computed between the original and marked signal. Based on existing differences between the perceptual hashes over a large test set, the average, minimum and maximum transparency is defined. For the robustness measurement of the perceptual hash, the introduced transparency measure is used.
Acknowledgements

The work about the profile definition described in this paper has been supported in part by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT. The information in this document is provided as is, and no guarantee or warranty is given or implied that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

References

1. Özer, H. and Sankur, B. and Memon, N., Robust audio hashing for audio identification, 12th European Signal Processing Conference (EUSIPCO), September, 2004
9. Özer, H. and Sankur, B. and Memon, N., Robust audio hashing for audio identification, 12th European Signal Processing Conference (EUSIPCO), September, 2004