Watermarking Based on Balanced Incomplete Block Designs

Zhi-Fang Yang
Department of Computer Science and Information Engineering, National Taipei University
151, University Rd., San Shia, Taipei, 237 Taiwan, R. O. C.
zfyang@mail.ntpu.edu.tw

ABSTRACT

Watermarking is a digital rights protection approach capable of associating information about media within the digital content during the whole lifetime of the contents. Representing messages by watermarks is the first essential stage for watermarking. In previous works, watermarks are generally designed by directly encoding messages into vectors. When more than one bit of information should be embedded, it is time consuming to identify a correct watermark in a large set of possible ones. In this work, a mathematical structure called balanced incomplete block designs (BIBD) is introduced to achieve watermark design. The purposes are to generate non-orthogonal vectors based on a smaller set of orthogonal ones and to reduce search time at the stage of watermark detection. Experimental results on real images are given to show the feasibility of the method, and discussions on future works are made.

1: INTRODUCTIONS

Digital multimedia distribution has become an important feature of the modern world due to the advent of modern communication infrastructures and technologies. Because of the ease of access and manipulation of digital media, the concept of digital rights protection has been introduced [1]. Among a variety of possible technologies for digital rights protection, watermarking is one potential academic field to provide such approaches to some of the various problems involved, including broadcast monitoring, owner identification, transaction tracking, content authentication, etc.[2] Capability of embedding information in digital contents within the lifetime is one of the many advantages of watermarking solutions [3].

Several types of watermarking models have been proposed to establish ways of thinking about watermarking systems, for instance, communication-based models and geometric models [3]. Regardless of which models are adopted, mapping messages into watermarks is the first essential step. According to watermark detection techniques, two types of messages can be found: one is for the verification of the existence of a certain kind of information, namely, one bit of information, and the other is for the identification of a specific message in a set of messages. In practice, most applications require multi-messages rather than only one bit of information [3].

For multi-message watermarking, the whole problem is actually a problem of mapping messages into message vectors at the sender site and then mapping those vectors back into messages at the receiver site [3]. For correlation-based watermark detection, finding the transmitted watermark message is generally done by computing correlations between the received message vector and each one in the message set, and then by searching the highest detection value [5].

In previous works, orthogonal message vectors are generally preferred for good code separation [3]. However, massive storage and long codes are usually required for practical problems. Furthermore, computation of correlations grows along with the size of the message set. Besides, the conflict is worse between the requirement of robustness and the need of imperceptibility.

In this work, to tackle the above problems, messages are designed based on a mathematical structure called balanced incomplete block designs (BIBD) [6-8]. Mathematical properties of BIBD are utilized to generate more non-orthogonal message vectors based on fewer orthogonal basis vectors. The number of correlation computation for watermark identification can be reduced to not more than an analytic formula defined with parameters in the BIBD. Watermarking based on BIBD can also be found in Trappe et al. [5], where anti-collusion fingerprinting is proposed by giving up some resiliency to uniquely identify K or fewer colluders with K a constant defined in the block designs.

The rest of this paper is organized as follows. Section 2 gives the proposed approach of BIBD-based watermark design and the corresponding detection scheme. Section 3 offers experimental results on a watermarking approach based on the proposed BIBD-based watermarks. Conclusions are finally made in Section 4.

2: WATERMARK DESIGN BASED ON BALANCED INCOMPLETE BLOCK DESIGNS

First, a general watermarking model is introduced, followed by more detailed figures of the watermark

embedder and extractor with BIBD coding and decoding functions, respectively, and then the proposed approach of BIBD coding and decoding is given.

2.1 OVERVIEW

Fig. 1 shows a general model for watermarking systems. A watermarking system embeds a message into contents using an embedder, sends the watermarked contents through a channel, and extracts the message using an extractor. The contents may be used at both the encoder and the decoder.

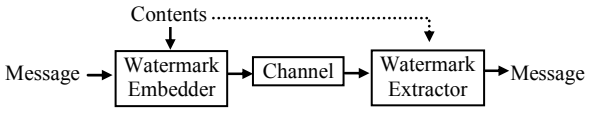

Figure 1: A general watermarking model.

Fig. 2 depicts possible functions performed by a watermark embedder. The system generates a BIBD code for an input message, modulates the sequence of symbols into a pattern compatible with the contents in which the message will be embedded, modifies the pattern according to properties of the contents, and finally embeds it into the contents. This work focuses on the part of BIBD coding.

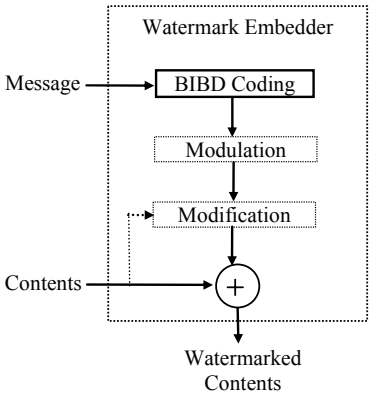

Figure 2: Watermark embedder with BIBD coding.

Fig. 3 depicts possible functions performed by a watermark extractor. The system extracts the embedded pattern possibly by subtracting the original contents or by some other schemes, recovers the modifications based on properties of the contents, demodulates the pattern into a sequence of symbols, and finally identifies which message it is in the message set.

2.2 BIBD CODING

The theory of BIBDs belongs to the field of combinatorial mathematics and has applications in codes, computer science, cryptography, etc. [6]. The definition of the BIBD is as follows [8].

Definition 1: A BIBD, a (v, b, r, k, λ)-configuration, is a pair (**V**, **B**) where **V** is a v-set and **B** is a collection of b k-subsets of **V** (blocks). Every 2-subset of **V** appears in exactly λ blocks, and each element of **V** .appears in exactly r blocks.

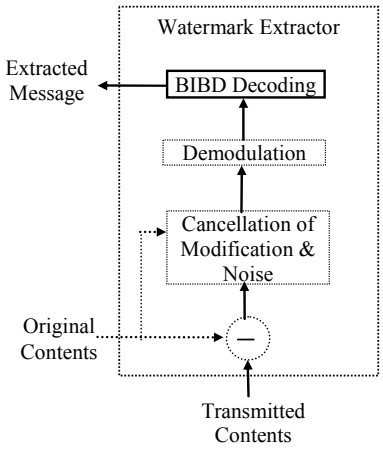

Figure 3: Watermark extractor with BIBD decoding.

A BIBD has at least many blocks as it has points [6-8]. This property is known as Fisher's inequality, which is the most basic necessary condition for the existence of a BIBD [8]:

Theorem 1 (Fisher's inequality): A BIBD (v, b, r, k, λ)-configuration exists only if $b \geqq v$ [7].

In this work, Fisher's inequality is used to guarantee that a set of non-orthogonal message vectors can be built based on a set of fewer orthogonal basis vectors. In order to achieve the above purpose, the incidence matrix of a BIBD is utilized, which is defined as follows [6].

Definition 2: The incidence matrix of a BIBD (v, b, r, k, λ)-configuration is a $v \times b$ matrix $\mathbf{A} = (a_{ij})$ with $a_{ij} = 1$ if the ith element of **V** appears in the jth block of **B**, and $a_{ij} = 0$ otherwise.

Thus, watermarks can be designed as follows.

Definition 3 (proposed watermark design): For a BIBD (v, b, r, k, λ)-configuration with the incidence matrix **A,** the corresponding set of message vectors, denoted as $\mathbf{W} = \{ \mathbf{w}_1, \ldots, \mathbf{w}j, \ldots, \mathbf{w}_b\}$, is defined to be the set of b $1 \times v$ vectors with $\mathbf{w}j$ the jth column vector of **A**. And the corresponding orthogonal basis set, denoted as $\mathbf{O} = \{\mathbf{o}_1, \ldots, \mathbf{o}_i, \ldots, \mathbf{o}_v\}$, is the v $1 \times v$ vectors with the ith element of $\mathbf{o}_i$ equals 0 and 1 for the rest of the vector. The proof of the orthogonal property of the basis vectors is straightforward.

The following is an example of a set of BIBD codes for watermarks. If a BIBD (16, 20, 5, 4, 1)-configuration [6] was chosen, the corresponding incidence matrix **A** is as shown in Eq. (1). Then, there are sixteen 1×16 orthogonal basis vectors and twenty 1×16 message vectors.

$$\mathbf{A} = \begin{pmatrix}
1&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
1&0&0&0&0&1&1&1&1&0&0&0&0&0&0&0&0&0&0&0\\
1&0&0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0&0&0\\
1&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1&0&0\\
0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0&0\\
0&1&0&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0\\
0&1&0&0&0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1\\
0&0&1&0&0&1&0&0&0&0&1&0&0&0&0&0&1&0&0&1\\
0&0&1&0&0&0&0&1&0&0&0&0&1&0&1&0&0&1&0&0\\
0&0&1&0&0&0&0&0&1&0&0&1&0&1&0&0&0&0&1&0\\
0&0&0&1&0&1&0&0&0&0&0&0&1&0&0&1&0&0&1&0\\
0&0&0&1&0&0&1&0&0&0&0&1&0&0&0&0&1&1&0&0\\
0&0&0&1&0&0&0&0&1&1&0&0&0&0&1&0&0&0&0&1\\
0&0&0&0&1&0&1&0&0&0&0&0&1&1&0&0&0&0&0&1\\
0&0&0&0&1&0&0&1&0&1&0&0&0&0&0&1&0&1&0&0\\
0&0&0&0&1&0&0&0&1&0&1&0&0&0&0&1&0&1&0&0
\end{pmatrix} \quad (1)$$

2.3 BIBD DECODING

Detection of messages should be conservative because the transmitted messages may be distorted during transmission. Thus, only BIBD codes with not-used basis vectors are not taken into detections. That is, after the transmitted BIBD message vector, denoted as **m**, is extracted at the receiver site, correlations between **m** and each vector in the set of orthogonal basis vectors **O** are performed to find which basis vectors are not involved. Only the result of zero is treated as not-used. That is, the set of not-used basis vectors, $\mathbf{O}'$, is actually the following set.

$$\mathbf{O}' = \mathbf{O} - \{\mathbf{o}_k\}, \quad (2)$$

where $\mathbf{o}_k \in \mathbf{O}$ and the kth element of **m** is 0.

The set of candidate message vectors, $\mathbf{W}'$, is found by discarding the message vectors containing any basis vectors in $\mathbf{O}'$:

$$\mathbf{W}' = \mathbf{W} - \{\mathbf{w}_k\}, \quad (3)$$

where $\mathbf{w}_k \in \mathbf{W}$ and $\mathbf{w}_k \cdot \mathbf{o}_k' \neq 0$ for some $\mathbf{o}_k' \in \mathbf{O}'$ and $\cdot$ the correlation operator.

Finally, the priorities of the candidate message vectors can be derived by computing correlations between the detected vector **m** and each vector in the set $\mathbf{W}'$. Higher priorities are given to candidate message vectors with larger correlation values.

Theorem 2: If p zero elements are found in a detected message vector **m**, the number of correlation computations is not more than $b - pr + 0.5 \times \lambda p(p-1)$ for a BIBD (v, b, r, k, λ)-configuration.

Proof: According to the proposed detection strategy, the number of not-used basis vectors in $\mathbf{O}'$ is p since p zero elements are found in a detected message vector **m**. Then the number of vectors in the set of candidate message vectors, $\mathbf{W}'$, is the number of all blocks, b, minus the number of blocks containing the p not-used basis vectors. According to the definition of BIBD, each basis vector appears in exactly r blocks. For n not-used basis vectors, the number of blocks containing the p not-used basis vectors is the value of pr minus possible repetition counting, which is not more than $0.5 \times \lambda p(p-1)$. Then the the number of correlation computations is not more than $b - pr + 0.5 \times \lambda p(p-1)$.

3: EXPERIMENTAL RESULTS

The experiments were performed on real images. A watermarking approach is designed based on the proposed BIBD watermark design. The BIBD (16, 20, 5, 4, 1)-configuration used in Section 2.2 and the corresponding incidence matrix **A** shown in Eq. (1) is utilized.

The watermark embedder and extractor are basically based on the quantization methods proposed in [9-10]. The media contents are transformed into a wavelet domain, and the watermark pattern is embedded bit by bit in the randomly chosen detail coefficients. However, the watermark pattern bits are not hidden in the highest resolution, and the quantization parameter is a constant. Note that, in order to increase robustness of the watermarks, each detail coefficient to be hidden is adjusted to the center of the bin of the quantization before embedding.

Thus, for each watermark bit $f(i)$, a wavelet coefficient, $w_{k,l,m,n}$, is randomly selected, where $k = h$, v, and d denote "horizontal', "vertical", and "diagonal" detail coefficients, respectively; $l = 1, 2, \ldots, L$ specifies a resolution level, and (m, n) represents a spatial location. Each wavelet coefficient can be chosen only once. The embedding rules are as follows:

$w_{k,l,m,n}$:=

$$\begin{cases} w'_{k,l,m,n} + \varepsilon & \text{if } w_{k,l,m,n} \le 0 \text{ and } Q(w_{k,l,m,n}) \neq f(i) \\ w'_{k,l,m,n} - \varepsilon & \text{if } w_{k,l,m,n} \ge 0 \text{ and } Q(w_{k,l,m,n}) \neq f(i) \\ w'_{k,l,m,n} & \text{if } Q(w_{k,l,m,n}) = f(i) \end{cases}, \quad (4)$$

where $w'_{k,l,m,n}$ is defined by the following function

$$w'_{k,l,m,n} = \frac{\varepsilon}{2}\left(2\left\lfloor \frac{w_{k,l,m,n}}{\varepsilon} \right\rfloor + 1\right), \quad (5)$$

and := the assignment operator, ε the quantization parameter, and $Q(.)$ a quantization function defined as follows.

$$Q(w_{k,l,m,n}) = w_{k,l,m,n} \bmod 2. \quad (6)$$

After distribution or redistribution, the watermark is extracted rather simple by the coefficient selection key. The distributed contents are first transformed into the predefined wavelet domain. Then the coefficient selection key is used to locate the wavelet coefficient positions in which the watermark bits are embedded, and the watermark bits can be extracted directly by applying Eq. (6).

Figure 4(a) shows the original 512×512 image — Lena. The Haar wavelet transform is used and resolution level is three. A quantization parameter, 2^3 is used. The selected watermark, $\mathbf{w}_1$, is embedded 1024 times in the

corresponding image copy, and the PSNR is around 44. Fig. 4(b) shows the watermarked image.

(a) (b)

Figure 4: Example of watermarked image: (a) original; (b) fingerprinted with PSNR = 44.2.

No noise is added in order to show the correctness of the proposed approach. Thus, at the receiver site, the set of not-used basis vectors, $\mathbf{O}'$, is found to be $\{\mathbf{o}_5, \mathbf{o}_6, \mathbf{o}_7, \mathbf{o}_8, \mathbf{o}_9, \mathbf{o}_{10}, \mathbf{o}_{11}, \mathbf{o}_{12}, \mathbf{o}_{13}, \mathbf{o}_{14}, \mathbf{o}_{15}, \mathbf{o}_{16}\}$, and the candidate watermark message vector is exactly the correct answer $\mathbf{f}_1$.

4 CONCLUSIONS

Digital rights protection is more and more important due to the advanced technologies of digital multimedia distribution, and watermarking can provide solutions to some of the problems involved in it. This study develops a novel watermark design approach based on a mathematical structure called balanced incomplete block designs. The mathematical properties of BIBDs are used. A list of candidate watermarks with priorities can be provided. A theorem is proposed to show how the computation time of correlations can be reduced, and the experimental results show that the correct embedded watermark message vector can be found. Future research may be aimed at developing more robust embedding methods, solving problems caused by possible rounding and truncation involved in the transmission process.

References

[1] R. H. Koenen, J. Lacy, M. Mackay, and S. Mitchell, "The long march to interoperable digital rights management," *Proceedings of the IEEE*, 92(6):883–897, June 2004.

[2] S. Katzenbeisser and F. A. P. Petitcolas, *Information Hiding Techniques for Steganography and Digital Watermarking*. Boston: Art House, 2000.

[3] I. J. Cox, M. L. Miller, J. A. Bloom, *Digital Watermarking*. Boston: Morgan Kaufmann, 2001.

[4] W. Trappe, M. Wu, Z. J. Wang, and K. J. R. Liu, "Anti-collusion fingerprinting for multimedia", *IEEE Trans. Signal Processing*, *51(4):1069–1087,April 2003.*

[5] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital data," *IEEE Trans. Inform. Theory*, 44:1897–1905, Sep. 1998.

[6] C. J. Colbourn and J. H. Dinitz, *The CRC Handbook of Combinatorial Designs*. Boca Raton: CRC, 1996.

[7] Ryser, H. J, *Combinatorial Mathematics.* Buffalo, NY: Math. Assoc. Amer., 1963.

[8] Dinitz, J. H. and Stinson, D. R, "A Brief Introduction to Design Theory," Ch. 1 in *Contemporary Design Theory: A Collection of Surveys*, (Ed. J. H. Dinitz and D. R. Stinson). New York: Wiley, pp. 1-12, 1992.

[9] D. Kundur and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication," *Proceedings of the IEEE*, 87(7):1167–1180, July 1999.

[10] Z. F. Yang and W. H. Tsai, "Watermark approach to embedded signature-based authentication by channel statistics," *Optical Engineering*, 42(4):1157–1165, April 2003.