REGULARIZED ADABOOST FOR CONTENT IDENTIFICATION

Honghai Yu [†*] *and Pierre Moulin*[†*]

[†]ECE Dept., University of Illinois at Urbana-Champaign, USA
[*]Advanced Digital Sciences Center, Singapore

ABSTRACT

This paper proposes a regularized Adaboost learning algorithm to extract binary fingerprints by filtering and quantizing perceptually significant features. The proposed algorithm extends the recent symmetric pairwise boosting (SPB) algorithm by taking feature sequence correlation into account. Information and learning theoretic analysis is given. Significant performance gains over SPB are demonstrated for both audio and video fingerprinting.

***Index Terms*—** Content identification, fingerprinting, learning theory, mutual information

1. INTRODUCTION

The problem of fingerprint-based content identification (ID) has received considerable attention from both academia and industry. YouTube uses content ID to detect copyrighted audio and video uploads in real time. Shazam and SoundHound use it for music identification on mobile devices. Other applications include advertisement tracking, broadcast monitoring, copyright control, and law enforcement [1, 2, 3]. The fingerprinting algorithm encodes signal into a short fingerprint which allows for real-time search. The fingerprint must be robust to various content-preserving distortions, while being discriminative enough to distinguish perceptually different signals.

Many fingerprinting algorithms have been proposed based on various signal features [1, 2, 3, 4]. A theoretical framework for fingerprint-based content ID systems was presented in [5], and derived a fundamental relation between database size and query length under some statistical assumptions. Decoding of correlated fingerprints is studied in [6, 7] and the related problem of physical object identification is studied in [8]. Fingerprinting algorithms that employ a variation of Adaboost to select filters and quantizers, such as Asymmetric Pairwise Boosting (APB) [9] and Symmetric Pairwise Boosting (SPB) [10, 11], have demonstrated excellent content ID performance.

This paper presents two contributions. First, we provide an information theoretic analysis of SPB and show that each iteration of SPB maximizes a lower bound on the mutual information between matching fingerprint pairs. Second, we propose a regularized Adaboost algorithm, which tackles SPB's implicit assumption that signal frames are statistically independent which does not hold in practice because of frame overlapping. The proposed algorithm is tested on both audio and video content ID systems, demonstrating significantly better performance than SPB.

Notation: we follow the convention that uppercase letters represent random variables while lowercase letters represent particular realizations of these random variables. A vector is denoted by an underscore (e.g., $\underline{f}$) and a temporal sequence by a boldface letter (e.g., $\mathbf{f}$).

2. STATEMENT OF THE CONTENT ID PROBLEM

Following [5], a *content database* is defined as a collection of M elements, $\mathbf{x}(m) \in \mathcal{X}^N, m = 1, 2, \ldots, M$, each of which is a sequence of N *frames* $\{x_1(m), x_2(m), \ldots, x_N(m)\}$. A frame could be a short video segment, a short sequence of image blocks, or a short audio segment. Frames may be overlapping spatially, temporally, or both, to prevent misalignment during matching [1].

The problem is to determine whether a given *query* consisting of $L < N$ frames, $\mathbf{y} \in \mathcal{X}^L$, is related to some element of the database, and if so, identify which one. To this end, an algorithm $\psi(\cdot)$ must be designed, returning the decision $\psi(\mathbf{y}) \in \{0, 1, 2, \ldots, M\}$, where $\psi(\mathbf{y}) = 0$ indicates that $\mathbf{y}$ is unrelated to any of the database elements.

3. STRUCTURED CONTENT ID CODES

In this paper, we restrict our attention to the following fairly general class of fingerprint-based content ID codes. The codes of [1, 10, 11] among others, fall in this category.

Definition 1 *[12] A (M, N, L) structured content ID code for a size-M database populated with $\mathcal{X}^N$-valued content items, and granularity L, is a pair consisting of a mapping $\phi : \mathcal{X} \to \mathcal{F}$ generating an encoding function $\Phi : \mathcal{X}^N \to \mathcal{F}^N$ that returns a fingerprint $\boldsymbol{f} = \Phi(\boldsymbol{x})$ with components $\underline{f_i} = \phi(x_i)$ for each $1 \le i \le N$, and a decoding function $\psi : \mathcal{F}^L \to \{0, 1, \ldots, M\}$ returning $\hat{m} = \psi(\Phi(\mathbf{y}))$.*

This work was supported by NSF grant CCF 12-19145 and by AUIP fellowship from the A*STAR Graduate Academy, Singapore.

Hence the mapping ϕ is applied independently to each frame. It might be convenient to impose additional structure on the code. For instance, the mapping $\phi : \mathcal{X} \to \mathcal{F}$ in [10, 11] is obtained by applying a set of J optimized filters to each frame and quantizing each of the J real-valued filter outputs to four levels. Hence $\mathcal{F}$ takes the form $\mathcal{A}^J$ with $\mathcal{A} = \{a, b, c, d\}$. In this case we view the fingerprint as an array $\mathbf{f} = \{f_{ij}, 1 \le i \le N, 1 \le j \le J\}$ and the query fingerprint as an array $\mathbf{g} = \{g_{ij}, 1 \le i \le L, 1 \le j \le J\}$ where i denotes time and j filter index. We also use the notation $\underline{f} = \{f_j, 1 \le j \le J\}$ for the subfingerprint associated with a given frame. We also write ϕ in vector form as $\phi = \{\phi_j, 1 \le j \le J\}$. The length of the binary subfingerprint $\underline{f}$ is $J \log_2 |\mathcal{A}|$.

The decoding function ψ, in most content ID systems, measures distance between fingerprints [10, 11, 5, 7]. If the distance is less than a predefined decision threshold τ, the fingerprint is declared as a match for the query. This is a *variable-size list decoder*: the number of matches could be 0, 1, 2 or more. Alternatively a single-output decoder might be used, returning only the index of the closest match. In this paper, Hamming distance metric with a list decoder is used to make a fair comparison with the SPB algorithm of [10, 11].

4. SPB FOR FILTER AND QUANTIZER SELECTION

Primitive signal processing features such as filters and quantizers had been heuristically chosen until learning-based methods, such as Adaboost, were proposed [9, 10, 11]. Adaboost-chosen filters and quantizers outperform the heuristically chosen ones [9, 10, 11].

The *symmetric pairwise boosting* (SPB) algorithm [10, 11] operates as follows. A training set $\mathcal{T} \triangleq \{(x_t, y_t, z_t) \in \mathcal{X}^2 \times \{\pm 1\}, t \in \mathcal{T}\}$ is comprised of a subset $\mathcal{T}_+$ of $|\mathcal{T}|/2$ *matching pairs* and a subset $\mathcal{T}_-$ of $|\mathcal{T}|/2$ *nonmatching pairs*, where a pair $(x_t, y_t) \in \mathcal{X}^2$ is said to be matching if the second signal is a distorted version of the first, and nonmatching if the two signals are independent. The binary variable (label) z_t is equal to 1 (resp. -1) if (x_t, y_t) is matching (resp. nonmatching). Define a set of J weak classifiers $h_j : \mathcal{X}^2 \to \{\pm 1\}, 1 \le j \le J$, as

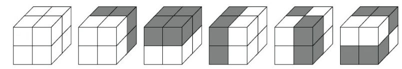
$$h_j(x, y) = \begin{cases} +1 & \text{if } \phi_j(x) = \phi_j(y) \\ -1 & \text{otherwise} \end{cases} \qquad (1)$$

where ϕ_j is parameterized by a filter $\lambda_j : \mathcal{X} \to \mathbb{R}$ and a quantizer $Q_j : \mathbb{R} \to \mathcal{A}$,

$$\phi_j(x) = Q_j(\lambda_j(x)). \qquad (2)$$

Denote by $\mathcal{H}$ the class of feasible classifiers (indexed by the choice of filters and quantizers).

A popular family of filters is the Haar-like filters used in [9, 10, 11] which are easy to compute and rich enough to describe perceptually significant visual features. The filter outputs for the 3-D Haar-like filters in [11] are the average difference between values in light and dark regions shown in Fig. 1.



(a) (b) (c) (d) (e) (f)

Fig. 1. 3-D Haar-like filters [11]: (a) spatio-temporal average, (b) temporal difference, (c,d) spatial difference, and (e,f) spatio-temporal difference.

Input: training set $\mathcal{T} \triangleq \{(x_t, y_t, z_t) \in \mathcal{X}^2 \times \{\pm 1\}, t \in \mathcal{T}\}$
Initialization: define equal weights $w_t^{(1)} = 1/|\mathcal{T}|, \forall t \in \mathcal{T}$
Do for $j = 1, \ldots, J$

1. Choose the classifier h_j that minimizes the weighted error over $h \in \mathcal{H}$
$$e_j = \sum_{t \in \mathcal{T}} w_t^{(j)} \mathbb{1}\{h(x_t, y_t) \neq z_t\}. \qquad (3)$$
2. Compute $\alpha_j = \frac{1}{2} \log \frac{1 - e_j}{e_j}$.
3. Update the weights
$$w_t^{(j+1)} = w_t^{(j)} \exp\{-\alpha_j z_t h_j(x_t, y_t)\}.$$
4. Normalize the weights so that $\sum_{t \in \mathcal{T}} w_t^{(j+1)} = 1$.

Output: J pairs of filter and quantizer $\{(\lambda_j, Q_j)\}_{j=1}^J$ parameterizing the chosen J classifiers $\{h_j\}_{j=1}^J$.

Table 1. Adaboost for filter and quantizer selection.

To reduce the computational complexity of the training, a limited number of candidate quantizers are evaluated.

The SPB algorithm is an adaptation of the well-known Adaboost classification algorithm given in Table 1. Upon completion of the algorithm, Adaboost would output the *boosted classifier* $h_{\mathrm{B}}(x, y) \triangleq \mathrm{sgn}\left[\sum_{1 \le j \le J} \alpha_j h_j(x, y)\right]$. However [10, 11] do not use the boosted classifier, only the filter λ_j and quantizer Q_j associated with each h_j are used to produce the fingerprints.

5. CONTENT ID CAPACITY

A content ID system, like any other communication system, is subject to a fundamental capacity limit that upper bounds the rate at which information can be decoded with arbitrarily low probability of error. For an iid signal process $\mathbf{X}$, memoryless degradation channel, and fixed structured content ID code (Def. 1) with mapping $\phi : \mathcal{X} \to \mathcal{F}$, the content ID capacity is given by $C = I(\underline{F}; \underline{G})$ [5]. If ϕ is a code design parameter, then $C = \max_\phi I(\underline{F}; \underline{G})$. If the signal $\mathbf{X}$ is an ergodic stationary process and the degradation channel from $\mathbf{X}$ to $\mathbf{Y}$ is stationary ergodic, one may conjecture that $C = \max_\phi I(\mathbf{F}; \mathbf{G})$ where $I(\mathbf{F}; \mathbf{G})$ is the mutual information rate between the random process $\mathbf{F}$ and $\mathbf{G}$. The mutual information rate is a nondecreasing function of the number of filters J, which is fixed here.

6. INFORMATION THEORETIC ANALYSIS OF SPB

In this section, we show that, at each iteration $1 \le j \le J$, SPB maximizes a lower bound on the mutual information

$I(F_j; G_j) = H(F_j) - H(F_j|G_j)$ associated with the joint probability distribution $P_{F_j G_j}$ induced by the choice (2) of ϕ_j. Indeed it follows from (3) that SPB selects the weak classifier that minimizes the weighted error:

$$h_j = \arg\min_{h\in\mathcal{H}} \Bigg[\sum_{t\in\mathcal{T}_+} w_t^{(j)} \mathbb{1}\{h(x_t, y_t) = -1\} + \sum_{t\in\mathcal{T}_-} w_t^{(j)} \mathbb{1}\{h(x_t, y_t) = 1\} \Bigg], \quad (4)$$

where the two error terms are the empirical weighted *false-negative* and *false-positive* error probabilities, respectively. For a given classifier $h \in \mathcal{H}$, the empirical version of the false-negative error probability for matching fingerprints, $P_{e,j} = P_{F_j G_j}(F_j \neq G_j)$, is given by

$$\widehat{P}_{e,j} = \widehat{\Pr}(F_j \neq G_j|\mathcal{T}_+, h) = \sum_{t\in\mathcal{T}_+} w_t^{(j)} \mathbb{1}\{h(x_t, y_t) = -1\},$$

and the empirical version of the false-positive error probability, $P_{F_j} P_{G_j}(F_j = G_j)$, is

$$\widehat{\Pr}(F_j = G_j|\mathcal{T}_-, h) = \sum_{t\in\mathcal{T}_-} w_t^{(j)} \mathbb{1}\{h(x_t, y_t) = 1\}.$$

First, we derive a link between $\widehat{\Pr}(F_j \neq G_j|\mathcal{T}_+, h)$ and $H(F_j|G_j)$. Fano's inequality gives

$$H(F_j|G_j) \leq h_2(P_{e,j}) + P_{e,j} \log(|\mathcal{A}| - 1), \quad (5)$$

where $P_{e,j} = P_{F_j G_j}(F_j \neq G_j)$, $\mathcal{A}$ is the alphabet for scalar fingerprint introduced in Section 3, and $h_2(P_{e,j})$ is the binary entropy function [13]. We have observed that inequality (5) is tight. In Fig. 2a, we show the empirical equivocation $\widehat{H}(F_j|G_j)$ and Fano's upper bound $h_2(\widehat{P}_{e,j}) + \widehat{P}_{e,j} \log(|\mathcal{A}| - 1)$ evaluated from 16,000 matching pairs and 16 classifiers. Thus, minimizing $\widehat{\Pr}(F_j \neq G_j|\mathcal{T}_+, h)$ is equivalent to minimizing a tight upper bound on $H(F_j|G_j)$.

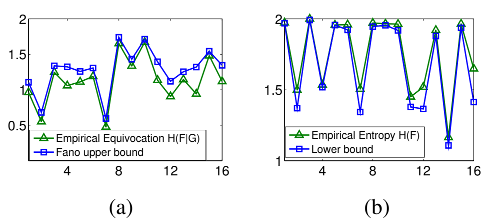


Fig. 2. (a) $\widehat{H}(F_j|G_j)$ and $H(\widehat{P}_{e,j}) + \widehat{P}_{e,j} \log(|\mathcal{A}| - 1)$. (b) $\widehat{H}(F_j)$ and $-\log \widehat{\Pr}(F_j = G_j)$. The x-coordinate is the classifier index j.

Next, we derive a link between $\widehat{\Pr}(F_j = G_j|\mathcal{T}_-, h)$ and $H(F_j)$. When F_j and G_j are generated from nonmatching pairs, we model them by a product distribution with identical marginals. From Lemma 2.10.1 in [13], we have $P_{F_j} P_{G_j}(F_j = G_j) \geq 2^{-H(F_j)}$, for two iid random variables F_j and G_j. Then $H(F_j)$ is lower bounded by

$$H(F_j) \geq -\log P_{F_j} P_{G_j}(F_j = G_j). \quad (6)$$

Again, we can verify the tightness of (6) from nonmatching pairs, see Fig. 2b. Thus, minimizing $\widehat{\Pr}(F_j = G_j|\mathcal{T}_-, h)$ is equivalent to maximizing a tight lower bound on $H(F_j)$.

From the above argument, we conclude that each iteration $1 \leq j \leq J$ of SPB simultaneously minimizes an upper bound on $H(F_j|G_j)$ and maximizes a lower bound on $H(F_j)$, thus maximizes a lower bound on $I(F_j; G_j) = H(F_j) - H(F_j|G_j)$.

However, frames are temporally overlapped to overcome misalignment during matching, which leads to temporally correlated fingerprints $\mathbf{F}_j = \{F_{1j}, F_{2j}, \ldots, F_{Lj}\}$ for each chosen classifier h_j. In a memoryless channel where each G_{ij} only depends on $\mathbf{F}_j$ only via F_{ij}, we have $I(\mathbf{F}_j; \mathbf{G}_j) \leq \sum_{i=1}^{L} I(F_{ij}; G_{ij})$ [13]. Equality holds when $\{F_{1j}, F_{2j}, \ldots, F_{Lj}\}$ are independent. Conversely, $I(\mathbf{F}_j; \mathbf{G}_j) \ll \sum_{i=1}^{L} I(F_{ij}; G_{ij})$ when $\{F_{1j}, F_{2j}, \ldots F_{Lj}\}$ are highly correlated. Thus we can increase the mutual information by decorrelating temporal fingerprints. Many frame-wise distortions can be modeled as memoryless channels, including resizing, cropping and rotation.

In the next section, we show that the classifiers' ability to decorrelate frames differs dramatically across different types of filters. In order to increase mutual information by decorrelating temporal fingerprints, we propose to use a regularizer to effectively eliminate those filters that generate highly correlated fingerprints from the candidate pool $\mathcal{H}$. Experiments demonstrate the effectiveness of this regularizer.

7. REGULARIZED ADABOOST

For a given filter λ (such as those from Fig. 1), the response $\lambda(\mathbf{X}) = \{\lambda(X_i), 1 \leq i \leq L\}$ is a L-dimensional random vector. Denote by $\rho(s,t) \in [-1, +1]$ the correlation coefficient between two random variables $\lambda(X_s)$ and $\lambda(X_t)$. Define the average correlation coefficient of $\lambda(\mathbf{X})$ as

$$\overline{\rho}(\lambda) \triangleq \frac{1}{L^2 - L} \sum_{s\neq t} |\rho(s,t)|. \quad (7)$$

The functional $\overline{\rho}(\lambda)$ captures the filter's ability to decorrelate overlapping frames and can be easily estimated from the training dataset. In Fig. 3, we show the estimated $\overline{\rho}(\lambda)$ for the family of Haar-like filters of Fig. 1 applied to video data. Within the family, type (b), (e) and (f) filters can decorrelate overlapping frames extremely well, type (a) and (c) filters produce almost perfectly correlated responses, and type (d) filters produce moderately correlated responses.

In our new regularized Adaboost algorithm, filters with large average correlation coefficient are penalized with the new objective function

$$e_j^{\text{REG}} = \sum_{t\in\mathcal{T}} w_t^{(j)} \mathbb{1}\{h(x_t, y_t) \neq z_t\} + \gamma\overline{\rho}(h), \quad (8)$$

where γ is the tuning parameter which can be chosen by cross-validation and $\overline{\rho}(h) = \overline{\rho}(\lambda)$ indicates the weak classifier h is parameterized by the filter λ. The regularized Adaboost is obtained by simply replacing (3) in Table 1 with (8) .

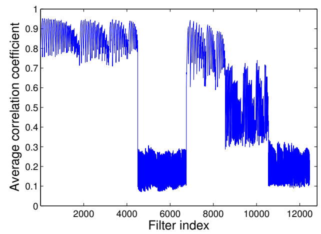


Fig. 3. Average correlation coefficient $\overline{\rho}(f)$ for the family of Haar-like filters on video frames.

8. CONVERGENCE OF THE REGULARIZED ADABOOST ALGORITHM

The Discrete Adaboost algorithm in Table 1 (with predictor variable (X, Y) and binary response variable Z) admits a known interpretation as an iterative procedure for fitting an additive logistic regression model [14, 15] $\delta(x,y) = \sum_{1\le j\le J} \alpha_j h_j(x,y)$ under the exponential loss function. For regularized Adaboost, the loss function is now the *regularized exponential loss*

$$L(z,\delta(x,y)) \triangleq \exp\{-z\delta(x,y)\} + \gamma \sum_{1\le j<J} 2\sinh(\alpha_j)\overline{\rho}(h_j).$$

At iteration j, we solve

$$(\alpha_j, h_j) = \arg\min_{\alpha\in\mathbb{R}, h\in\mathcal{H}} \left[\sum_{t\in\mathcal{T}} w_t^{(j)} \exp\{-\alpha z_t h(x_t,y_t)\} + 2\sinh(\alpha)\gamma\overline{\rho}(h)\right], \quad (9)$$

where $w_t^{(j)} = \exp\{-z_t\delta_{j-1}(x_t,y_t)\}$ and

$$\delta_i(x,y) \triangleq \sum_{1\le j\le i} \alpha_j h_j(x,y).$$

Using the fact that $h(x,y) \in \{-1,1\}$ and $\sum_{t\in\mathcal{T}} w_t^{(j)} = 1$, the objective function of (9) can be rewritten as

$$2\sinh(\alpha)\left[\sum_{t\in\mathcal{T}} w_t^{(j)} \mathbb{1}\{h(x_t,y_t)\neq z_t\} + \gamma\overline{\rho}(h)\right] + e^{-\alpha}. \quad (10)$$

The minimum over $h\in\mathcal{H}$ is given by

$$h_j = \arg\min_{h\in\mathcal{H}} \sum_{t\in\mathcal{T}} w_t^{(j)} \mathbb{1}\{h(x_t,y_t)\neq z_t\} + \gamma\overline{\rho}(h). \quad (11)$$

Plugging h_j into (10) and solving for α, we obtain

$$\alpha_j = \frac{1}{2}\log\frac{1-e_j^{\text{REG}}}{e_j^{\text{REG}}}, \quad (12)$$

where e_j^{REG} is given by (8). Equation (11) and (12) are equivalent to Step 1 and 2 of the regularized Adaboost algorithm.

Notice that the above derivation does not depend on the specific form of $\overline{\rho}(h)$. As long as the regularizer is a functional of h, it can be plugged into the regularized Adaboost algorithm, which makes this approach fairly general.

9. EXPERIMENTAL RESULTS

We follow the experimental setups in [10] and [11] with $J = 8$ and $L = 4$ for audio fingerprinting and $J = 16$, $L = 41$ for video fingerprinting. A set of matching and nonmatching pairs is randomly selected from the training dataset, where matching pairs consist equal amount of each distortion. The audio distortions considered are 64 kbps lossy compression using WMA encoding, addition of a 20% echo, and band-pass filtering in the 0.4-4 kHz range. The video distortions considered are 50% cropping, vertical mirroring, frame rotation at 15 degrees and frame shifting downward and left by 100 pixels each. We consider geometric distortions only as they represent the most challenging video distortions to detect, and SPB works nearly perfectly for simple distortions, such as lossy compression, resizing, and frame rate change [11] (so does regularized Adaboost). For the weight of the regularizer γ, values between 0.1 and 0.3 worked well in our experiments. The results shown are obtained using $\gamma = 0.2$.

To quantify ID performance, we plot the receiver operating characteristics (ROC) curves for different distortions. Each point on the curve represents a false negative rate and false positive rate pair corresponding to a decoding threshold τ. As shown in Fig. 4 and Fig. 5, irrespective of the signal type used, regularized Adaboost outperforms SPB for all the considered distortions.

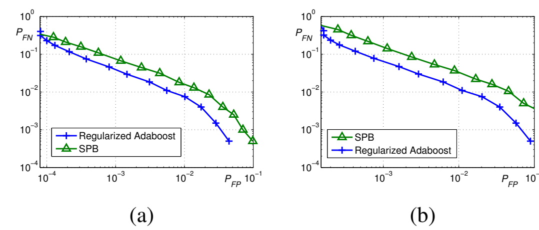


Fig. 4. ROC curves for audio distortions: (a) 20% echo; (b) WMA compression + bandpass filtering.

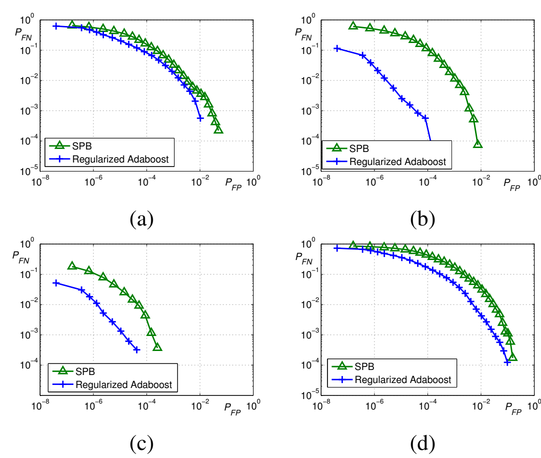


Fig. 5. ROC curves for video distortions: (a) cropping; (b) vertical mirroring; (c) rotation; (d) shift.

10. ACKNOWLEDGEMENT

The authors would like to thank Rohit Naini for providing experimental results on audio.

11. REFERENCES

[1] J. Haitsma and T. Kalker, "A highly robust audio fingerprinting system," in *Proceedings of the International Symposium on Music Information Retrieval*, Paris, France, 2002.

[2] S. Baluja and M. Covell, "Audio fingerprinting: Combining computer vision data stream processing," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Honolulu, Hawaii, April 2007.

[3] J. Lu, "Video fingerprinting for copy identification: from research to industry applications," in *SPIE Electric Imaging Symposium on Media Forensics and Security I*, San Jose, CA, USA, 2009.

[4] P. Cano, E. Batlle, T. Kalker, and J. Haitsma, "A review of audio fingerprinting," *VLSI Signal Processing*, vol. 41, no. 3, pp. 271–284, 2005.

[5] P. Moulin, "Statistical modeling and analysis of content identification," in *Information Theory and Applications Workshop*, San Diego, CA, February 2010.

[6] A. L. Varna and M. Wu, "Modeling and analysis of correlated binary fingerprints for content identification," *IEEE Transactions on Information Forensics and Security*, vol. 6, pp. 1146–1159, 2011.

[7] R. Naini and P. Moulin, "Model-based decoding metrics for content identification," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Kyoto, Japan, March 2012, pp. 1829 –1832.

[8] T. Holotyak, S. Voloshynovskiy, O. J. Koval, and F. Beekhof, "Fast physical object identification based on unclonable features and soft fingerprinting," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, 2011, pp. 1713–1716.

[9] Y. Ke, D. Hoiem, and R. Sukthankar, "Computer vision for music identification," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Washington, DC, USA, 2005, pp. 597–604.

[10] D. Jang, C. D. Yoo, S. Lee, S. Kim, and T. Kalker, "Pairwise boosted audio fingerprint," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 4, pp. 995–1004, Dec. 2009.

[11] S. Lee, C. D. Yoo, and T. Kalker, "Robust video fingerprinting based on symmetric pairwise boosting," *IEEE Trans. Circ. and Sys. for Video Technol.*, vol. 19, no. 9, pp. 1379–1388, Sept. 2009.

[12] R. Naini and P. Moulin, "Real adaboost for content identification," in *IEEE Statistical Signal Processing Workshop*, Ann Arbor, MI, Aug 2012.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*, Wiley-Interscience, July 2006.

[14] T. Hastie, R. Tibshirani, and J.H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, New York: Springer-Verlag, 2001.

[15] J. H. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," *Annals of Statistics*, vol. 28, pp. 400–407, 1998.