SNR MAXIMIZATION HASHING FOR LEARNING COMPACT BINARY CODES

Honghai Yu †* *and Pierre Moulin*†*

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

ABSTRACT

In this paper, we propose a novel robust hashing algorithm based on signal-to-noise ratio (SNR) maximization to learn binary codes. We first motivate SNR maximization for robust hashing in a statistical model, under which maximizing SNR minimizes the robust hashing error probability. A globally optimal solution can be obtained by solving a generalized eigenvalue problem. The proposed algorithm is tested on both synthetic and real datasets, showing significant performance gain over existing hashing algorithms.

Index Terms— Robust hashing, SNR maximization, content identification, generalized eigenproblem

1. INTRODUCTION

Robust hashing, a.k.a. semantic hashing and fingerprinting, has received considerable attention from both academia and industry. For instance, robust hashing is used in the YouTube content ID system to detect registered audio and video uploads in real time. Shazam and SoundHound use robust hashing for music identification on mobile devices. Other applications include advertisement tracking, broadcast monitoring, copyright control, and law enforcement [1, 2, 3, 4]. In these applications, the content is encoded into compact binary hash codes (a fingerprint) which allows real-time search. The fingerprint must be robust to various content-preserving distortions, while being discriminative enough to distinguish perceptually different signals.

A popular family of hash functions, which assumes centered (mean-subtracted) inputs $\mathbf{x} \in \mathbb{R}^d$, linear projections $W \in \mathbb{R}^{d\times k}$, and binary quantization, is given by

$$h(\mathbf{x}, W) = \mathrm{sgn}(W^T\mathbf{x}), \tag{1}$$

where $\mathrm{sgn}(v) = 1$ if $v \geq 0$ and -1 otherwise. For a matrix or vector, $\mathrm{sgn}(\cdot)$ denotes the element-wise operation. Many robust hashing algorithms, such as in [5, 6, 7, 8, 9], fall in this category. Other families of hash functions such as kernelized hash functions [10], multilayer neural networks [11, 12], and boosting [13, 14] are more expensive to train and evaluate.

Traditionally, W was generated by randomly sampling a distribution that satisfies the locality-sensitive property [15, 16]. However, data-independent W can lead to inefficient codes, and thus require much longer codes (larger k) to work well. Recently, learning W from training datasets has been shown to yield better performance than data-independent W for the same code length [6, 8, 12, 9]. To learn W, the nondifferentiable and nonconvex sgn function of (1) is often approximated by either the identity function $h(\mathbf{x}; W) = W^T\mathbf{x}$ [6, 7, 8], which introduces large approximation error when the magnitude of $W^T\mathbf{x}$ is large, or the hyperbolic tangent function $h(\mathbf{x}; W) = \tanh(W^T\mathbf{x})$ [9], where the optimization may be trapped in a bad local optimum due to the nonconvexity of the tanh function.

In this paper, we analyze a statistical model for robust hashing and show that maximizing the signal-to-noise ratio (SNR) minimizes the robust hashing error probability. SNR has been used as the performance measure in many applications, such as lossy compression [17], matched filtering [18], relay functionality in memoryless relay networks [19], and beamforming in narrowband sensor arrays [20, 21]. However, to our knowledge, using SNR as the performance measure for robust hashing has never been considered in the literature.

Motivated by the analysis, we propose a SNR maximization hashing (SNR-MH) algorithm that iteratively finds uncorrelated projection directions that maximize the SNR. In doing so, we bypass the step of approximating the sgn function and finds the global optimal solution. Experimental results from both synthetic and real datasets demonstrate SNR-MH's superior performance in learning compact binary codes.

2. A MOTIVATIONAL MODEL

In this section, we consider a statistical model for robust hashing and motivate our signal-to-noise ratio maximization hashing (SNR-MH) by showing that a larger SNR leads to a smaller robust hashing error probability.

2.1. A Statistical Model for Robust Hashing

The statistical model consists of the following ingredients:

(A1) Assume $\mathbf{X}$ follows the distribution $P_{\mathbf{X}}$ with mean $\mathbf{0}$ and covariance matrix $C_X \in \mathbb{R}^{d\times d}$.

(A2) When the query item $\mathbf{Y}$ is a distorted version of $\mathbf{X}$, we assume the following noise model holds:

$$\mathbf{Y} = \mathbf{X} + \mathbf{Z}, \tag{2}$$

where $\mathbf{Z}$ is independent of $\mathbf{X}$ and follows the distribution $P_{\mathbf{Z}}$ with mean $\mathbf{0}$ and positive-definite covariance matrix C_Z.

(A3) When $\mathbf{X}$ and $\mathbf{Y}$ are not related, $\mathbf{Y}$ is independent of $\mathbf{X}$ and follows the distribution $P_{\mathbf{Y}}$.

(A4) Assume a projection matrix $W \in \mathbb{R}^{d\times k}, k \leq d$ such that W^TC_XW and W^TC_ZW are both diagonal [1], i.e., the transformed feature components $w_i^T\mathbf{X}$'s are uncorrelated and the transformed noise components $w_i^T\mathbf{Z}$'s are uncorrelated. Denote by $\{\sigma_i^2\}_{i=1}^d$ and $\{\lambda_i^2\}_{i=1}^d$ the diagonal entries of

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

[1] The existence of such W is guaranteed [22, Chapter 15].

$W^T C_X W$ and $W^T C_Z W$ respectively. For the i-th projection, we have $w_i^T \mathbf{Y} = w_i^T \mathbf{X} + w_i^T \mathbf{Z}$, which give rise to the i-th signal-to-noise ratio as

$$\mathrm{SNR}_i \triangleq \frac{\sigma_i^2}{\lambda_i^2}, \quad 1 \le i \le k. \tag{3}$$

(A5) To generate binary fingerprints, we use the component-wise sgn function:

$$\begin{aligned} \mathbf{F} &= \mathrm{sgn}(W^T \mathbf{X}) \in \{\pm 1\}^k \\ \mathbf{G} &= \mathrm{sgn}(W^T \mathbf{Y}) \in \{\pm 1\}^k, \end{aligned} \tag{4}$$

with $F_i = \mathrm{sgn}(w_i^T \mathbf{X})$ and $G_i = \mathrm{sgn}(w_i^T \mathbf{Y}), 1 \le i \le k$.

(A6) Upon seeing a $(\mathbf{x}, \mathbf{y})$ pair, a binary decision about whether $\mathbf{x}$ and $\mathbf{y}$ are *similar* or *dissimilar* must be made based on the fingerprints $\mathbf{f}$ and $\mathbf{g}$, where similar and dissimilar $(\mathbf{x}, \mathbf{y})$ pairs are defined as

$$\begin{aligned} &\text{Similar } (S): \mathbf{x} \text{ and } \mathbf{y} \text{ are related by (2);} \\ &\text{Dissimilar } (D): \mathbf{x} \text{ and } \mathbf{y} \text{ are independent.} \end{aligned}$$

(A7) We use the decision rule

$$d_H(\mathbf{f}, \mathbf{g}) \underset{D}{\overset{S}{\lesseqgtr}} \tau, \tag{5}$$

where $d_H(\mathbf{f}, \mathbf{g}) \triangleq \sum_{i=1}^k 1_{\{f_i \neq g_i\}}$ is the Hamming distance between $\mathbf{f}$ and $\mathbf{g}$ and $\tau \in \{0, 1, \ldots, k\}$ is a decision threshold. The decoder declares $(\mathbf{x}, \mathbf{y})$ similar when $d_H(\mathbf{f}, \mathbf{G}) \le \tau$ and dissimilar when $d_H(\mathbf{f}, \mathbf{G}) > \tau$.

2.2. Error Probability Analysis

Based on the above statistical model, we analyze the robust hashing error probabilities when $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, C_X)$ and $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, C_Z)$. For Gaussian random vectors, uncorrelatedness of $w_i^T \mathbf{X}$'s and $w_i^T \mathbf{Z}$'s implies independence. It then follows from (A4) that F_i's are independent and from (A2) that so are G_i's.

Denote by $P_{\mathbf{FG}}(\mathbf{f}, \mathbf{g}) = \prod_{i=1}^k P_{F_i G_i}(f_i, g_i)$ the joint distribution of $(\mathbf{F}, \mathbf{G})$ when $\mathbf{X}$ and $\mathbf{Y}$ are similar. Denote by $P_{\mathbf{F}} P_{\mathbf{G}}(\mathbf{f}, \mathbf{g}) = \prod_{i=1}^k P_{F_i} P_{G_i}(f_i, g_i)$ the product probability mass function (pmf) when $\mathbf{X}$ and $\mathbf{Y}$ are dissimilar. The performance of the hashing system is quantified using probability of miss:

$$P_M \triangleq P_{\mathbf{FG}}\{d_H(\mathbf{F}, \mathbf{G}) > \tau\}, \tag{6}$$

and probability of false alarm

$$P_F \triangleq P_{\mathbf{F}} P_{\mathbf{G}}\{d_H(\mathbf{F}, \mathbf{G}) \le \tau\}. \tag{7}$$

In the rest of this section, we prove the following proposition with the help of two lemmas:

Proposition 1. *For a fixed τ, P_M is a decreasing function of $\{\mathrm{SNR}_i\}_{i=1}^k$ and P_F is independent of $\{\mathrm{SNR}_i\}_{i=1}^k$.*

Proof. When $\mathbf{F} = \mathrm{sgn}(W^T \mathbf{X})$ and $\mathbf{G} = \mathrm{sgn}(W^T \mathbf{Y})$ are generated from dissimilar $\mathbf{X}$ and $\mathbf{Y}$, we have

$$P_{F_i} P_{G_i}\{F_i \neq G_i\} = \frac{1}{2}, \quad \forall 1 \le i \le k. \tag{8}$$

As the pairs $(F_i, G_i), 1 \le i \le k$ are independent, $d_H(\mathbf{F}, \mathbf{G})$ follows the binomial distribution with k trials and parameter $\frac{1}{2}$:

$$d_H(\mathbf{F}, \mathbf{G}) \sim \mathrm{Bi}(k, \frac{1}{2}). \tag{9}$$

Hence, P_F does not depend on $\{\mathrm{SNR}_i\}_{i=1}^k$.

When $\mathbf{F}$ and $\mathbf{G}$ are generated from similar $\mathbf{X}$ and $\mathbf{Y}$, define

$$p_i \triangleq P_{F_i G_i}\{F_i \neq G_i\}, \quad 1 \le i \le k. \tag{10}$$

As the pairs $(F_i, G_i), 1 \le i \le k$ are independent, the Hamming distance between $\mathbf{F}$ and $\mathbf{G}$ follows the Poisson binomial distribution (PBD) with parameter $\{p_1, \ldots, p_k\} \in [0, 1]^k$:

$$P_{\mathbf{FG}}\{d_H(\mathbf{F}, \mathbf{G}) = l\} = \sum_{A \in E_l} \prod_{i \in A} p_i \prod_{j \in A^c} (1 - p_j), \quad 0 \le l \le k, \tag{11}$$

where E_l is the set of all subsets of l integers that can be selected from $\{1, 2, \ldots, k\}$ and $A^c = \{1, 2, \ldots, k\} \backslash A$ is the complement of A.

Let $T_k^S = d_H(\mathbf{F}, \mathbf{G})$ when $\mathbf{F}$ and $\mathbf{G}$ are similar, so $T_k^S \sim \mathrm{PBD}(\{p_1, \ldots, p_k\})$. Then we have $P_M = Pr\{T_k^S > \tau\}$.

Lemma 1. *For a given decision threshold $\tau \in \{0, 1, \ldots, k-1\}$ and fixed $\{p_1, p_2, \ldots, p_{k-1}\}$, $Pr\{T_k^S > \tau\}$ is an increasing function of p_k.*

Proof. Let $T_{k-1}^S \sim PBD(\{p_1, \ldots, p_{k-1}\})$, for $l = 0, 1, \ldots, k$, we have

$$Pr\{T_k^S = l\} = p_k \times Pr\{T_{k-1}^S = l-1\} + (1-p_k) \times Pr\{T_{k-1}^S = l\}. \tag{12}$$

Since every PBD is unimodal, first increasing, then decreasing, and the mode is either unique or shared by two adjacent integers [23], let l^* be the unique mode (or the smaller of the two modes) of T_{k-1}^S. When $l \le l^*$, we have $Pr\{T_{k-1}^S = l-1\} < Pr\{T_{k-1}^S = l\}$, so $Pr\{T_k^S = l\}$ decreases with p_k. When $l > l^*$ (or $l > l^* + 1$ when there are two modes), we have $Pr\{T_{k-1}^S = l-1\} > Pr\{T_{k-1}^S = l\}$, so $Pr\{T_k^S = l\}$ increases with p_k.

Therefore, when $0 \le \tau \le l^*$, $Pr\{T_k^S > \tau\} = 1 - \sum_{l=0}^{\tau} Pr\{T_k^S = l\}$ is an increasing function of p_k. When $l^* + 1 \le \tau \le k-1$, $Pr\{T_k^S > \tau\} = \sum_{l=\tau+1}^{k} Pr\{T_k^S = l\}$ is also an increasing function of p_k. □

Lemma 2. *Under (A2) and (A4), p_i is a decreasing function of SNR_i for $i = 1, \ldots, k$.*

Proof. Denote by $\widetilde{X}_i = w_i^T \mathbf{X}$ and $\widetilde{Z}_i = w_i^T \mathbf{Z}$ the i-th transformed feature random variable and transformed noise random variable respectively. Then $\widetilde{X}_i \sim \mathcal{N}(0, \sigma_i^2)$ and $\widetilde{Z}_i \sim \mathcal{N}(0, \lambda_i^2)$. By (2) and (4), $F_i = \mathrm{sgn}(\widetilde{X}_i)$ and $G_i = \mathrm{sgn}(\widetilde{X}_i + \widetilde{Z}_i)$ are independent. It has been shown in [24, Equations 16 and 17] that

$$p_i = P_{F_i G_i}\{F_i \neq G_i\} = \frac{1}{\pi} \arctan\left(\frac{1}{\mathrm{SNR}_i}\right), \tag{13}$$

which is a decreasing function of SNR_i. □

Combining results from Lemma 1 and Lemma 2, we have shown that for a fixed τ, P_M is a decreasing function of $\{\mathrm{SNR}_i\}_{i=1}^k$. □

3. SNR MAXIMIZATION HASHING

Motivated by the analysis in Section 2, we propose a hashing algorithm based on SNR maximization, which bypasses the step of approximating the sgn function and finds the globally optimal solution.

Denote by $\mathbf{X} \in \mathbb{R}^d$ and $\mathbf{Z} \in \mathbb{R}^d$ the feature random vector and noise vector respectively (both $\mathbf{X}$ and $\mathbf{Z}$ have mean zero). The goal is to learn a $d \times k$ transformation matrix $W = [w_1, \dots, w_k]$ such that the transformed feature vector $W^T\mathbf{X} \in \mathbb{R}^k$ is uncorrelated and the SNR at each projection $\mathrm{SNR}_i = \mathrm{var}(w_i^T\mathbf{X})/\mathrm{var}(w_i^T\mathbf{Z})$ is maximized. Mathematically, for $i = 1, 2, \dots, k$, w_i is sequentially learnt via the following optimization:

$$\begin{aligned} w_i = \arg\max_{w} \quad & \frac{w^T C_X w}{w^T C_Z w} \\ \text{subject to} \quad & w^T C_X w_j = 0, \quad \forall j < i \\ & w^T C_Z w_j = 0, \quad \forall j < i \\ & w^T C_Z w = 1, \end{aligned} \tag{14}$$

where the last constraint is to normalize the transformed noise to unit power so the solution is unique. To ensure C_Z is invertible, a small constant may be added to the diagonal entries of C_Z, i.e., replacing C_Z with $C_Z + \alpha I$ where I denotes the identity matrix.

The optimization (14) is used in multiclass Fisher discriminant analysis (FDA) [25] to learn up to k linear projections when there are $k+1$ different classes. In multiclass FDA, C_X is the inter-class scatter matrix and C_Z is the intra-class scatter matrix. The solution of (14) is given by the k eigenvectors corresponding to the first k largest eigenvalues of the generalized eigenproblem [25]

$$C_X w = \gamma C_Z w, \tag{15}$$

where γ is the eigenvalue (also the SNR) associated with eigenvector w.

There are several ways to reduce (15) to a standard eigendecomposition problem [22]. One way is to form $C_Z^{-1}C_X$, but in general $C_Z^{-1}C_X$ is not symmetric so all the nice properties about diagonalizing symmetric matrices will be lost.

Another way to solve (15) is by using the Cholesky decomposition on C_Z. Let $C_Z = LL^T$ where L is a lower triangular matrix. Then (15) becomes

$$\left[L^{-1}C_X L^{-T}\right]\left[L^T w\right] = \gamma\left[L^T w\right], \tag{16}$$

which is a standard eigendecomposition problem. Note that this procedure is equivalent to applying a whitening transformation L^{-1} on the noise. After whitening, $L^{-1}\mathbf{Z}$ and $L^{-1}\mathbf{X}$ have covariance matrices $L^{-1}C_Z L^{-T} = I$ and $L^{-1}C_X L^{-T}$ respectively.

Connection to PCA Hashing: In PCA hashing [6, 8], W is given by the top k eigenvectors of C_X. This is equivalent to assuming C_Z is the identity matrix in (15). PCA hashing maximizes the transformed feature variance without considering the noise. The only case PCA hashing is optimal in the sense of SNR maximization is when the noise $\mathbf{Z}$ has uncorrelated components with equal variance.

Connection to Semi-Supervised Hashing (SSH): SSH [6] was formulated as minimizing empirical error on the labeled data while maximizing variance and independence of hash bits over the labeled and unlabeled data. After approximating the sgn function with the identity function, SSH maximizes the following objective function subject to the constraint $W^TW = I$:

$$\sum_{i=1}^{k}\left[w_i^T C_{XY} w_i - w_i^T C_{X\widehat{Y}} w_i + \beta w_i^T C_X w_i\right], \tag{17}$$

where C_{XY} and $C_{X\widehat{Y}}$ denote the cross covariance matrices between similar and dissimilar $\mathbf{X}$ and $\mathbf{Y}$ respectively and $\beta > 0$ is a weighting parameter chosen by cross-validation. The optimal projection matrix W then consists of the top eigenvectors of the matrix $C_{XY} - C_{X\widehat{Y}} + \beta C_X$.

Under the statistical model of Section 2, C_{XY} becomes C_X and $C_{X\widehat{Y}}$ is the zero matrix. As a result, the optimal projections of SSH are equivalent to those of PCA Hashing.

In the next section, we will compare empirical performance of SNR-MH and other hashing algorithms.

4. EXPERIMENTAL RESULTS

4.1. Results on Synthetic Data

We first run simulations on synthetic datasets where we compare performance between SNR-MH and PCAH under the statistical model in Section 2. We fix the feature dimension $d = 128$. The feature vector $\mathbf{X}$ consists of i.i.d. samples from $\mathcal{N}(\mathbf{0}, C_X)$. The covariance matrix $C_X = UD_XU^T$, where U is a random $d \times d$ orthogonal matrix and D_X is a $d \times d$ diagonal matrix with diagonal entries uniformly sampled from $(0.5, 1)$ and normalized so that their sum equals to $P = 128$ where P is the total signal power.

The noise vector $\mathbf{Z}$ consists of i.i.d. samples from $\mathcal{N}(\mathbf{0}, C_Z)$ where $C_Z = VD_ZV^T$ with V being a random orthogonal matrix and $D_Z = \mathrm{diag}\{d_{z1}, d_{z2}, \dots, d_{zd}\}$. Fixing the total noise power equal to P above, we consider three different scenarios depending on how $\{d_{z1}, d_{z2}, \dots, d_{zd}\}$ are generated:

1. **Uniform**: $d_{zi} = P/d, \quad \forall 1 \le i \le d$.
2. **Linear**: $d_{z_i} = a + (i-1)r, \quad \forall 1 \le i \le d$, where $a = 0.1$ and $\sum_{i=1}^{d} d_{zi} = P$.
3. **Exponential**: $d_{zi} = ar^{(i-1)}, \quad \forall 1 \le i \le d$, where $r = 1.05$ and $\sum_{i=1}^{d} d_{zi} = P$.

We generate 500,000 similar and dissimilar pairs for training and another 500,000 similar and dissimilar pairs for testing. Simulation results are shown in Fig. 1. The left column shows SNR_i for each projection w_i learnt by SNR maximization and PCA; the right column shows ROC curves for SNR-MH and PCAH at different code lengths. The rows corresponds to uniformly, linearly and exponentially generated $\{d_{z1}, d_{z2}, \dots, d_{zd}\}$ respectively.

Consider the left column first. As noted in Section 3, SNR-MH and PCA coincide in the uniform scenario. As we move to the linear (second row) and exponential (third row) scenarios where noise power is not evenly distributed, SNR_i increases from 1.32 to 9.98 and 81.83 in SNR maximization while remains largely unchanged in PCA. Moreover, more high SNR projections are learnt in the exponential case than in the linear case as exponential distribution creates more smaller d_{zi}'s. On the contrary, PCA performs similarly across the three scenarios and SNR_i is generally not a monotone function of the PCA projections as PCA seeks the variance-maximizing projections of the signal and ignores the noise structure.

Showing in the right column, SNR-MH and PCAH performs indistinguishably in the uniform scenario, whereas SNR-MH outperforms PCAH significantly in the linear and exponential scenarios, especially in the exponential scenario where the gain is in orders of magnitude due to the high SNR projections learnt by SNR maximization.

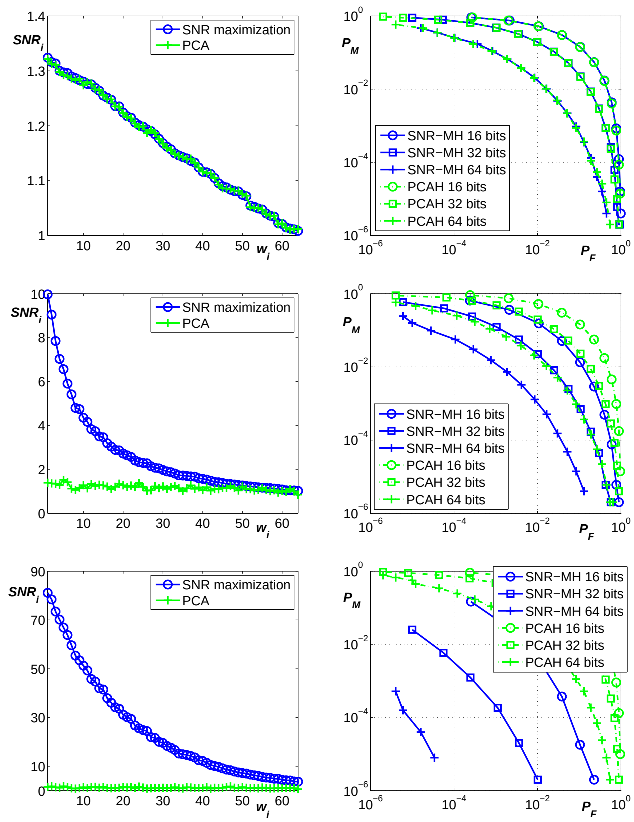
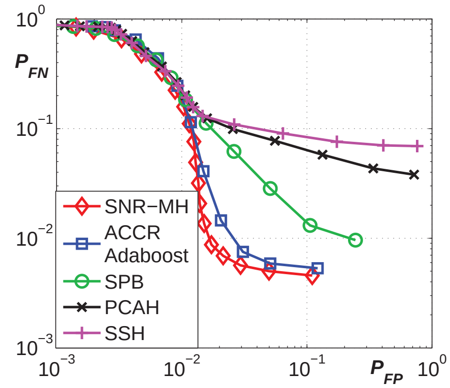
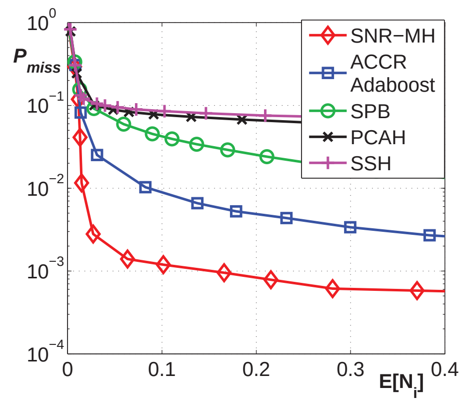


Fig. 1: Performance on the synthetic dataset. The left column shows SNR_i for each projection w_i learnt by SNR maximization or PCA; the right column shows ROC curves for SNR-MH and PCAH at different code lengths. Rows correspond to uniformly, linearly and exponentially generated $\{d_{z1}, d_{z2}, \ldots, d_{zd}\}$ respectively.



(a) Single-Output Decoder.



(b) List Decoder.

Fig. 2: Performance on audio content Identification. The query consists of 16 audio segments and 32 bits are extracted from each audio segment by each hashing algorithm.

4.2. Results on Audio Content Identification

Next, we test our proposed SNR-MH on an audio content identification (ID) system. The problem is to determine whether a given query **y** is related to some element of the database with M elements, and if so, identify which one. To this end, an algorithm must be designed, returning the decision $\psi(\mathbf{y}) \in \{0, 1, 2, \ldots, M\}$, where $\psi(\mathbf{y}) = 0$ indicates that **y** is unrelated to any of the database elements. This is a *single-output decoder*. Alternatively, a *variable-size list decoder* $\mathcal{L}(\mathbf{y}) \subseteq \{1, 2, \ldots, M\}$ might be used, returning 0, 1, 2 or more matches.

The audio dataset is a collection of 1,700 songs spanning a variety of music genres including classical, vocal, rock and pop. We randomly divide the 1,700 songs into training, validation, and testing subsets consisting of 100, 100, and 1,500 songs respectively. From the training songs, we generate 22,400 similar and 22,400 dissimilar feature pairs. The audio distortions considered are: 400 Hz to 4 kHz bandpass filtering, tunnel reverberation, boost bass, recording industry association of America (RIAA) equalization, down-sampling to 16 kHz, Attack-Decay-Sustain-Release (ADSR) envelop, and 64 kb/s WMA encoding. On top of the above distortions, each audio signal is encoded by 96 kb/s MP3 encoding and added a time delay of 92.9 ms.

We follow the same experimental setup as in [26] for audio fingerprinting. An audio signal is first normalized to mono with 11,025 Hz sampling rate, and then converted into overlapping segments by a window with size 371.52 ms and shift 185.76 ms. For every segment, an M-dimensional spectral subband centroid (SSC) vector is computed [27] from $M = 16$ critical subband linearly spaced in mel scale from 300 Hz to 5300 Hz. A SSC image, built from $N = 10$ consecutive SSC vectors, is the basic building block for fingerprint extraction. Given a 16×10 SSC image, we first convert it into a 160 dimensional vector and extract 32 bits from it by SNR-MH as the subfingerprint. For every shift of 185.76 ms, an SSC image is obtained from an audio segment of length 2.04 s. Every audio query is fixed to be 5 s long, corresponding to 16 SSC images.

To compare performance, we estimate probability of false positive (P_{FP}) and probability of false negative (P_{FN}) for the single-output decoder, and expected number of incorrect items on the list ($\mathbb{E}(N_i)$) and probability of miss (P_{miss}) for the list decoder [26]. Besides PCAH and SSH, we also compare with two boosting-based hashing algorithms, symmetric pairwise boosting (SPB) [13] and a regularized Adaboost (ACCR Adaboost) [14, 26], which have achieved excellent content ID performance on audio.

Fig. 2 shows the performance comparison on the audio content identification experiments. For both decoders, SNR-MH outperforms all other methods. For the list decoder, SNR-MH outperforms the next best by almost an order of magnitude .

5. REFERENCES

[1] R. Venkatesan, S.-M. Koon, M. H. Jakubowski, and P. Moulin, "Robust image hashing," in *IEEE International Conference on Image Processing: ICIP 2000*, Vancouver (BC), CA, September 2000.

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

[3] 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.

[4] 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.

[5] P. Indyk and R. Motwani, "Approximate nearest neighbors: Towards removing the curse of dimensionality," in *ACM STOC*, 1998, pp. 604–613.

[6] J. Wang, S. Kumar, and S.-F. Chang, "Semi-supervised hashing for scalable image retrieval," in *CVPR*, 2010, pp. 3424–3431.

[7] ——, "Sequential projection learning for hashing with compact codes," in *ICML*, 2010, pp. 1127–1134.

[8] Y. Gong and S. Lazebnik, "Iterative quantization: A procrustean approach to learning binary codes," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 2011, pp. 817–824.

[9] X. Li, G. Lin, C. Shen, A. van den Hengel, and A. R. Dick, "Learning hash functions using column generation." in *ICML*, ser. JMLR Proceedings. JMLR.org, 2013, pp. 142–150.

[10] B. Kulis and T. Darrell, "Learning to hash with binary reconstructive embeddings," in *in Proc. NIPS, 2009*, 2009, pp. 1042–1050.

[11] R. Salakhutdinov and G. Hinton, "Semantic hashing," *Int. J. Approx. Reasoning*, vol. 50, no. 7, pp. 969–978, Jul. 2009.

[12] M. Norouzi, D. J. Fleet, and R. Salakhutdinov, "Hamming distance metric learning," in *NIPS*, 2012, pp. 1070–1078.

[13] 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.

[14] H. Yu and P. Moulin, "Regularized Adaboost for content identification," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, Canada, 2013.

[15] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality-sensitive hashing scheme based on p-stable distributions," in *Proceedings of the twentieth annual symposium on Computational geometry*. ACM, 2004, pp. 253–262.

[16] M. Raginsky and S. Lazebnik, "Locality-sensitive binary codes from shift-invariant kernels," in *Advances in Neural Information Processing Systems (NIPS)*, Vancouver, Canada, December 2009, pp. 1509–1517.

[17] V. Bhaskaran and K. Konstantinides, *Image and Video Compression Standards: Algorithms and Architectures*, 2nd ed. Norwell, MA, USA: Kluwer Academic Publishers, 1997.

[18] J. R. Barry, D. G. Messerschmitt, and E. A. Lee, *Digital Communication: Third Edition*. Norwell, MA, USA: Kluwer Academic Publishers, 2003.

[19] K. S. Gomadam and S. A. Jafar, "Optimal relay functionality for snr maximization in memoryless relay networks," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 390–401, 2007.

[20] B. van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, pp. 4–24, April 1988.

[21] R. A. Monzingo, R. L. Haupt, and T. W. Miller, *Introduction to Adaptive Arrays*. Institution of Engineering and Technology, 2011.

[22] B. N. Parlett, *The Symmetric Eigenvalue Problem*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1998.

[23] S. M. Samuels, "On the number of successes in independent trials," *The Annals of Mathematical Statistics*, vol. 36, no. 4, pp. 1272–1278, 1965.

[24] F. Balado, N. Hurley, E. McCarthy, and G. Silvestre, "Performance analysis of robust audio hashing," *Information Forensics and Security, IEEE Transactions on*, vol. 2, no. 2, pp. 254–266, June 2007.

[25] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1988.

[26] H. Yu and P. Moulin, "Regularized adaboost learning for identification of time-varying content," *IEEE Transactions on Information Forensics and Security*, 2014, to appear.

[27] J. S. Seo, M. Jin, S. Lee, D. Jang, S. Lee, and C. D. Yoo, "Audio fingerprinting based on normalized spectral subband moments," *IEEE Signal Process. Lett.*, vol. 13, no. 4, pp. 209–212, 2006.