MODELING CONTENT FINGERPRINTS USING MARKOV RANDOM FIELDS

Avinash L. Varna and Min Wu
Department of Electrical and Computer Engineering
Institute for Advanced Computer Studies
University of Maryland, College Park

ABSTRACT

Content fingerprints are widely employed for identifying multimedia in various applications. A "fingerprint" of a video or audio is a short signature that captures unique characteristics of the signal and can be used to perform robust identification. Several fingerprinting techniques have been proposed in the literature and are often evaluated using benchmark databases. To complement these experimental evaluations, this paper develops a theoretical model for content fingerprints and evaluates the identification accuracy. Fingerprints and the noise are modeled as Markov Random Fields and the optimal decision rule for matching is derived. An algorithm to compute the probability of correct detection and the false alarm rate by estimating the density of states is described. Numerical results are provided for a model of a block based binary fingerprinting scheme and the influence of the fingerprint correlation and the noise on the detection accuracy is studied.

Index Terms— Content fingerprints, content identification, Markov Random Fields, Wang-Landau density of state estimation.

1. INTRODUCTION

The Internet is emerging as a new and powerful medium for multimedia distribution and consumption. These new distribution channels have also raised several challenges in multimedia management and rights enforcement. Popular copyrighted videos are often reposted on user generated content (UGC) websites, such as Youtube, without authorization. To ensure that the content is being used in accordance with the content owner's guidelines, UGC websites should be able to correctly identify posted videos. Content fingerprinting is emerging as a promising technology for multimedia identification, wherein, for each video or audio, a short "fingerprint" is computed that captures robust and unique characteristics of the signal. Given a video/audio that needs to be identified, a fingerprint is computed and compared with a database of known fingerprints. If a match is found, then the content has been identified, else it is deemed as unknown.

Content identification also has several applications in multimedia management. Many multimedia databases are often not fully annotated, and manual annotation is impractical. Fingerprints can be used to automatically identify the database content and annotate them. Fingerprints have also been deployed in products that identify audio from short clips recorded by mobile phones.

Several multimedia fingerprinting techniques have been proposed in the literature and have been reviewed in [1]. These techniques have mostly been evaluated through experiments on benchmark databases of limited sizes. In practical systems, the reference database typically contains millions of videos, and it is difficult to predict the performance of fingerprinting schemes on these large databases from moderate-scale experiments. There is a strong need for theoretical analysis that can complement experimental evaluations to provide understanding of the scalability and performance of fingerprinting systems and guide the design of better schemes. Using aspects from decision theory and game theory, our previous work [2, 3] has provided guidelines for designing fingerprints and choosing parameters, such as the length, to achieve a desired performance. The prior work assumed that the fingerprint components were independent and identically distributed (i.i.d.), but many practical schemes generate fingerprints with correlated components. In this paper, we develop a model using Markov Random Fields (MRFs) to analyze the performance of such fingerprinting schemes whose components are correlated.

Section 2 provides a brief overview of MRFs and Section 3 describes a model for content fingerprints using MRFs. The matching accuracy using fingerprints is also examined in Section 3, and an algorithm is developed to compute the probabilities of detection and false alarm. Section 4 provides numerical results and Section 5 summarizes the main results and contributions of this paper.

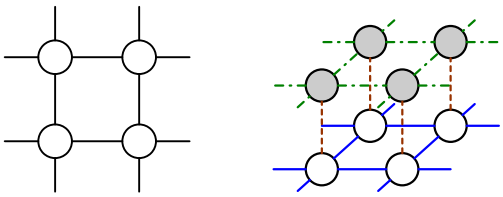
2. MARKOV RANDOM FIELDS

Markov Random Fields (MRFs) are a generalization of Markov chains in which time indices are replaced by space indices [4]. MRFs are undirected graphical models and represent conditional independence relations among random variables. In this section, we briefly review key concepts related to MRFs.

An MRF consists of an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a set of nodes $\mathcal{V}$ and a set of edges $\mathcal{E}$ between nodes. Each node $X \in \mathcal{V}$ represents a random variable, and we will use X to denote the node and the random variable interchangeably. The vector $\mathbf{X}$ denotes all random variables represented by the MRF. Two nodes X_i and X_j are said to be neighbors if there is an edge between them, i.e. $(i, j) \in \mathcal{E}$. A set of nodes $\mathcal{C}$ is called a maximal clique if every pair of nodes in $\mathcal{C}$ are

Email:{varna,minwu}@umd.edu

This work was supported in part by a grant from Motion Pictures Laboratories, Inc., Palo Alto, CA.



(a) (b)

Fig. 1. Markov Random Field model for (a) fingerprint components and (b) fingerprint and noise.

neighbors and there is no node in $\mathcal{V}\backslash\mathcal{C}$ that is a neighbor of *every* node in $\mathcal{C}$. An energy function $E_{\mathcal{C}}(\{x_{\mathcal{C}}\})$ is associated with every maximal clique $\mathcal{C}$ that maps the values $\{x_{\mathcal{C}}\}$ of the nodes in $\mathcal{C}$ to a real number. The joint probability distribution of all the random variables represented by the MRF is then given as $p(\mathbf{X} = \mathbf{x}) = \frac{1}{Z}\exp\left(-\sum_{\mathcal{C}} E_{\mathcal{C}}(\{x_{\mathcal{C}}\})\right)$, where Z is a normalization constant called the partition function. The term in the exponent, $E(\mathbf{x}) = \sum_{\mathcal{C}} E_{\mathcal{C}}(\{x_{\mathcal{C}}\})$, is sometimes referred to as the energy of the configuration $\mathbf{x}$.

MRFs have been used in several applications in image processing and computer vision [5] as they can represent local correlations among random variables. In the next section, we develop a model for content fingerprints using MRFs to capture local dependencies and examine the performance.

3. MRF MODEL FOR CONTENT FINGERPRINTS

We model content fingerprints as a Markov Random Field to capture correlations among individual fingerprint components. Each fingerprint value is represented as a node in the MRF, and pairs of nodes that have dependencies are joined by edges. We illustrate our model using a representative fingerprinting scheme that partitions each video frame into blocks and extracts one bit from each block [1]. For example, such a scheme could perform thresholding on the average luminance of a block. Alternatively, the differences between the average luminance of neighboring blocks could be quantized to one bit accuracy. While we use a simple 2-D Ising model for ease of illustration, the main principles behind our modeling can be extended to 3-D and more complex models.

3.1. Model for a block-based Fingerprinting scheme

Suppose that each video frame of size $PH_1 \times QH_2$ is partitioned into PQ blocks of size $H_1 \times H_2$ each and one bit of the fingerprint is extracted from each block. Due to underlying correlations among the blocks of the frame, these bits are likely to be correlated. We represent the bit extracted from each block as a node in a graph $\mathcal{G}_0 = (\mathcal{V}_0, \mathcal{E}_0)$, with the node $X_{i,j}$ representing the bit from the $(i,j)^{\text{th}}$ block. Each node may take one of two values ± 1, with bit 'b' represented as $(-1)^b$, and is connected to the four nearest neighbors, so that the overall graph satisfies 4-connectivity as shown in Fig. 1(a). For convenience, we use a vector $\mathbf{X}$ to represent the bits $\{X_{i,j}\}$, which could be obtained by any consistent reordering, such as raster scanning.

As described in Section 2, the joint probability distribution of the fingerprint can be specified by defining an energy function for the model. We use the energy function that has been commonly used for modeling binary images [5]:

$$E_0(\mathbf{x}) = -h\sum_i x_i - \eta \sum_{(j,k)\in\mathcal{E}_0} x_j x_k. \qquad (1)$$

This corresponds to the 2-D Ising model that has been widely used in statistical physics. Here, η controls the correlation between nodes that are connected and h determines the marginal distribution of the individual bits. A higher value for η would increase the correlation among neighboring bits, and large h would bias the bits to be $+1$. The joint distribution can then be written as $p_0(\mathbf{x}) = \frac{1}{Z_0}\exp(-E_0(\mathbf{x}))$.

While the above model suffices to describe the fingerprint bits of the original video frame, in practice, fingerprints are extracted from possibly modified versions of the video and may be noisy. The noise components may be mutually correlated and depend on the fingerprint bits. To accommodate such modifications, we propose a joint model for the noise bits and the fingerprint bits of the original unmodified video, which is shown in Fig. 1(b). The filled circles represent the noise bits and the open circles represent the fingerprint bits. The solid edges capture the dependencies among the fingerprint components, while the dashed and dotted edges represent the local correlations among the noise bits. The dashed edges can be used to model the correlations between the noise bits and the fingerprint bits, but such an undirected edge cannot completely capture the causal nature of this dependance.

In this paper, we consider the case where the noise bits may be mutually dependent, but are independent of the fingerprint bits, implying that the dashed edges are absent. In this case, the model for the noise bits $\{N_{i,j}\}$ reduces to a 2-D Ising model $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ similar to that for the fingerprints. The energy function for a configuration $\mathbf{n}$ can be defined as:

$$E_1(\mathbf{n}) = -\alpha\sum_i n_i - \gamma \sum_{(j,k)\in\mathcal{E}_1} n_j n_k, \qquad (2)$$

and the distribution is specified as $p_1(\mathbf{n}) = \frac{1}{Z_1}\exp(-E_1(\mathbf{n}))$. The parameters α and γ control the marginal distribution and the pairwise correlation among the noise bits, respectively.

The above MRF can be used to model block based binary video fingerprints computed on a frame by frame basis. For other fingerprinting schemes, different graphs can be used to capture the local dependencies among the fingerprint components. Given such a model for the fingerprints and the noise, we analyze the process of matching two fingerprints as a hypothesis testing problem, as illustrated in the next section.

3.2. Hypothesis Testing

Given a query video W and a reference video V in its database, the detector has to decide whether W is derived from V or whether the two videos are unrelated. To do so, the detector computes the fingerprints $\mathbf{y}$ and $\mathbf{x}$ from the videos W and V, respectively. The detector then performs a binary hypothesis test, with the null hypothesis H_0 that the two

fingerprints are independent and the alternate hypothesis H_1 that the fingerprint $\mathbf{y}$ is a noisy version of $\mathbf{x}$:

$$\begin{aligned} H_0 &: (\mathbf{x},\mathbf{y}) \sim p_0(\mathbf{x})p_0(\mathbf{y}), \\ H_1 &: (\mathbf{x},\mathbf{y}) \sim p_0(\mathbf{x})p_1(\mathbf{n}), \end{aligned} \tag{3}$$

where $p_0(\cdot)$ is the distribution of the fingerprints, $p_1(\cdot)$ is the distribution of the noise and the noise is the element-wise product of the two fingerprints $\mathbf{n} = \mathbf{x} \otimes \mathbf{y}$.

We consider a Neyman-Pearson setting, where the detector seeks to maximize the probability of detection P_d under the constraint that the probability of false alarm $P_f \leq \delta$. The optimal decision rule is obtained by comparing the log likelihood ratio (LLR) to a threshold:

$$LLR(\mathbf{x},\mathbf{y}) = E_0(\mathbf{y}) - E_1(\mathbf{n}) \underset{H_0}{\overset{H_1}{\gtrless}} \tau, \tag{4}$$

where the constants have been absorbed into the threshold τ, which is chosen such that $P_f = \delta$. In cases where the LLR is discrete, it may be necessary to incorporate randomization when the LLR equals the threshold.

For the frame-wise block-based binary fingerprinting scheme model described in Section 3.1, the LLR is given by:

$$LLR(\mathbf{x},\mathbf{y}) = -h\sum_i y_i - \eta\sum_{\mathcal{E}_0} y_j y_k + \alpha\sum_i n_i + \gamma\sum_{\mathcal{E}_1} n_j n_k.$$

If the fingerprint bits are i.i.d. and equally likely to be ± 1, corresponding to $\eta = h = 0$, and the noise bits are independent ($\gamma = 0$), the optimum decision rule reduces to a comparison of the Hamming distance between $\mathbf{x}$ and $\mathbf{y}$ to a threshold, as derived in [2]. However, when the bits are correlated, fingerprint matching using the Hamming distance is *suboptimal*.

The probability of detection $P_d = \Pr(LLR(\mathbf{x},\mathbf{y}) > \tau|H_1)$ and the probability of false alarm $P_f = \Pr(LLR(\mathbf{x},\mathbf{y}) > \tau|H_0)$. It is not possible to accurately estimate these tail probabilities using traditional techniques such as Markov Chain Monte Carlo (MCMC) simulations [6], since these events have small probability of occurrence and are rarely observed in a typical MCMC simulation. Instead, we take a different approach inspired by statistical physics to first estimate the so called density of states and then utilize this information to estimate these probabilities.

3.3. Computing P_d and P_f

For ease of illustration, we again use the example of the binary fingerprint model described in Section 3.1. Suppose we define $M(\mathbf{x}) = \sum_i x_i$ and $E_{corr}(\mathbf{x}) = -\sum_{(j,k)\in\mathcal{E}_0} x_j x_k$, the LLR in Eqn. (4) can be written as $LLR(\mathbf{x},\mathbf{y}) = -hM(\mathbf{y}) + \eta E_{corr}(\mathbf{y}) + \alpha M(\mathbf{n}) - \gamma E_{corr}(\mathbf{n})$, since $\mathcal{E}_0 = \mathcal{E}_1$ in this model. Similarly, the energy for the fingerprint bits and the noise , $E_0(\mathbf{x})$ and $E_1(\mathbf{n})$, described in Eqns. (1) and (2) can be rewritten in terms of these functions. Thus, the tuple

$$S(\mathbf{x},\mathbf{y}) = (M(\mathbf{x}), E_{corr}(\mathbf{x}), M(\mathbf{y}), E_{corr}(\mathbf{y}), M(\mathbf{n}), E_{corr}(\mathbf{n})),$$

captures all necessary information regarding the configuration $(\mathbf{x},\mathbf{y})$. Define $g(s) = g(m_x, e_x, m_y, e_y, m_n, e_n)$ as the number of configurations $(\mathbf{x},\mathbf{y})$ that have $M(\mathbf{x}) = m_x, E_{corr}(\mathbf{x}) = e_x, M(\mathbf{y}) = m_y, E_{corr}(\mathbf{y}) = e_y, M(\mathbf{n}) = m_n$, and $E_{corr}(\mathbf{n}) = e_n$. Note that this "density of states" g depends only on the underlying graphical model and is independent of the parameters (h,η,α,γ) of the distributions.

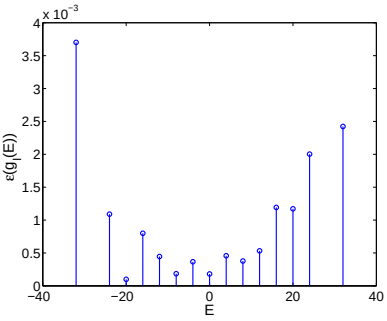


Fig. 2. Relative error in the estimation of density of states for a 4x4 Ising model with periodic boundary conditions.

The probability of detection P_d can then be rewritten as:

$$\begin{aligned} P_d(\tau) &= \sum_{(\mathbf{x},\mathbf{y})} 1\{LLR(\mathbf{x},\mathbf{y}) > \tau\} p_0(\mathbf{x})p_1(\mathbf{n}) \\ &= \sum_s g(s) 1\{LLR > \tau\} p_0(\mathbf{x})p_1(\mathbf{n}), \end{aligned} \tag{5}$$

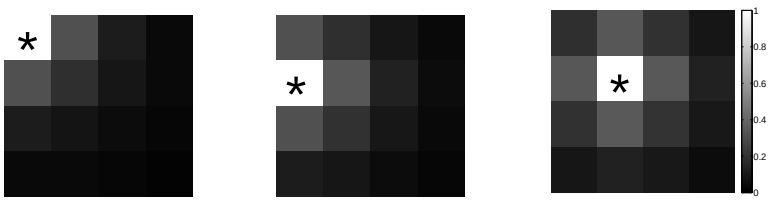
where the summation in the second equation is over all possible values of $s = (m_x, e_x, m_y, e_y, m_n, e_n)$. Similarly,

$$P_f(\tau) = \sum_s g(s) 1\{LLR > \tau\} p_0(\mathbf{x})p_0(\mathbf{y}). \tag{6}$$

As the LLR and the probabilities $p_1(\mathbf{n})$ and $p_0(\mathbf{x})$ depend only on s, knowledge of $g(s)$ allows us to compute P_d and P_f. Thus, the problem of computing P_d and P_f has been converted into one of estimating the density of states $g(s)$. An algorithm to estimate the density of states by constructing a Markov chain that has $\frac{1}{g(s)}$ as its stationary distribution and ensuring that all states are visited approximately equally often was proposed in [7]. An advantage of this "Wang-Landau" algorithm is that states with low probability of occurrence are also visited as often as high probability states, enabling us to estimate their probabilities accurately. We first use this algorithm [7] to estimate the density of states $g(s)$ and then compute P_d and P_f using Eqns. (5) and (6).

4. NUMERICAL RESULTS

We use the MRF model coupled with the technique for computing P_d and P_f described in the previous section to study the influence of correlation among the fingerprint components on the overall detection performance. We focus on binary fingerprinting schemes and provide numerical results for the model described in Section 3.1. As most binary fingerprint schemes generate equally likely (but not independent) bits, we set the parameter $h = 0$ in our simulations. This also has the effect of reducing the parameter space from a 6-D space $(m_x, e_x, m_y, e_y, m_n, e_n)$ to a 4-D space (e_x, e_y, m_n, e_n), as the expressions for the LLR and probability distributions will not involve m_x and m_y.



(a) (b) (c)

Fig. 3. Typical correlation structure among the various fingerprinting bits. Correlation coefficients for the (a) $(1,1)^{\text{th}}$ bit, (b) $(2,1)^{\text{th}}$ bit, (c) $(2,2)^{\text{th}}$ bit and the remaining bits. The '*' denotes the bit under consideration.

4.1. Density of states estimation

We evaluate the accuracy of the estimation algorithm using known exact results for the density of energy states $g_I(E)$ for the 2-D Ising model [8]. To enable comparison, periodic boundary conditions are used - the nodes $X_{1,j}$ in the top row are connected to the corresponding nodes $X_{M,j}$ in the bottom row, and the nodes in the first column are similarly connected to the nodes in the last column, so that every node is 4-connected. 4-connectivity is similarly achieved for the noise nodes $\{N_{i,j}\}$. We use the Wang-Landau algorithm to estimate the density of states $g(s) = g(e_x, e_y, m_n, e_n)$ by performing a random walk in the 4-D parameter space and use the estimated $g(s)$ to estimate $g_I(E)$. In our simulations, we use the parameters suggested in [7] and the maximum number of iterations is capped at 10^{10}.

We measure the accuracy of estimation by computing the relative error $\varepsilon(g_I(E))$ in the estimate of the density of states, defined as $\varepsilon(x) = \frac{|x - x_{est}|}{x}$. Fig. 2 shows the relative error in the estimation of the density of states for a 2-D Ising model of size 4×4 with periodic boundary conditions. From the figure, we observe that the maximum relative error is approximately 0.37%, and the mean relative error is 0.1%. These results demonstrate that accurate estimates of the density of states can be obtained using the Wang-Landau algorithm. The estimation accuracy can be improved by suitably altering parameters in the algorithm as necessary.

4.2. Performance of correlated fingerprints

To examine the performance of correlated fingerprints, we use the model without periodic boundary conditions. The nodes at the corners are only connected to their 2 closest neighbors, the remaining nodes at the borders are connected to their 3 closest neighbors, and all the other nodes are 4-connected.

4.2.1. Correlation among Fingerprint bits

The correlation is estimated from 10^8 MCMC iterations by retaining only 1 out of 100 iterations. Fig. 3 shows the correlation among the fingerprint bits for a 4×4 model, obtained by setting $\eta = 0.3$, $\alpha = 0.3$, and $\gamma = 0.1$. Fig. 3(a) shows the correlation between the $(1,1)^{\text{th}}$ bit (top left corner) and every other bit while Figs. 3(b) and (c) show the same for the $(2,1)^{\text{th}}$ bit and the $(2,2)^{\text{th}}$ bit, respectively. By symmetry, other bits in corresponding positions will have similar correlations. We observe that each bit is correlated with its nearest neighbor with a correlation coefficient $\rho_x \approx 0.3$ and the correlation decays with distance. This is the typical correlation behavior observed in our model and reflects the correlation expected in practice, as bits extracted from adjacent blocks are expected to be more correlated than bits extracted from blocks far apart. We observe a similar correlation structure among the noise bits $N_{i,j}$, as the noise and fingerprint models are similar.

4.2.2. Detection Performance

Using the estimated density of states, the probabilities P_d and P_f are computed as described in Section 3.3 to obtain the Receiver Operating Characteristics (ROC) curves. We examine the influence of different parameters on the detection performance. At the outset, we note that errors in the estimation of the density of states will also affect the accuracy of these estimates. However, as shown in Section 4.1, these errors are small, and the accuracy can be improved by obtaining a better estimate of the density of states.

First, we examine the effect of the noise on the detection accuracy. We characterize the noise by the probability p_n of a noise bit being '-1' - the equivalent of a binary '1' bit, and the correlation among the noise bits ρ_n, which are estimated from the MCMC trials. Fig. 4 shows the ROC curves for a fingerprint of size 4×4 bits with correlation $\rho_x = 0.2$ under two different p_n and fixed $\rho_n = 0.2$, for a detector using the Log Likelihood Ratio (LLR) statistic and a detector using the Hamming distance statistic. As expected, the performance is worse when there is a higher probability of the noise changing the fingerprint bits. We also observe that for a given noise level, the LLR statistic gives $5 - 10\%$ higher P_d at a given P_f compared to the Hamming distance detector.

Fig 5, shows the influence of the noise correlation on the detection performance. From the figure, we infer that for a fixed correlation among the fingerprint bits $\rho_x = 0.2$ and a fixed marginal probability of the noise bits $p_n = 0.3$, detection using the LLR statistic is not significantly affected by the noise correlation. This is due to the fact that the LLR takes into account the correlation among the noise bits. On the other hand, using the Hamming distance leads to a slight degradation in the performance as the correlation increases. This can be explained by the fact that as the noise correlation increases, noise vectors with large Hamming weights become

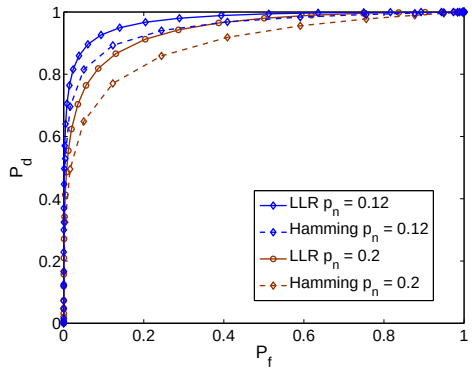


Fig. 4. Influence of p_n on the detection performance for 4×4 bits per frame, $\rho_x = 0.2$ and $\rho_n = 0.2$.

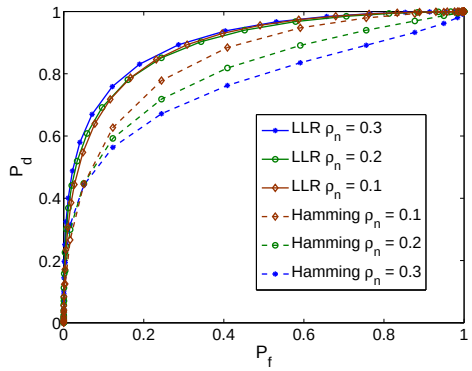


Fig. 5. ROC for different noise correlation ρ_n at fixed $p_e = 0.3$ and $\rho_x = 0.2$.

more probable, leading to higher missed detections.

Next, we examine the influence of the correlation among the fingerprint bits on the detection accuracy. Fig. 6 shows the ROC curves for content identification using fingerprints of size 4×4 for different correlations, where the noise parameters $p_n = \rho_n = 0.2$. From the figure, we again observe that detection using the LLR statistic, which compensates for the correlation among the fingerprint bits is not significantly affected by the correlation. For the Hamming distance statistic, there is an increase in false alarms at a given P_d as the correlation among the fingerprints increases, as similar configurations with smaller distances become more probable.

5. CONCLUSIONS

We have proposed a model for content fingerprints using Markov Random Fields that capture local correlations among individual fingerprint components. We examined the problem of matching two fingerprints as a binary hypothesis testing problem and derived the optimal decision rule. For the case of independent and equally likely bits, the optimum rule reduces to the comparison of the Hamming distance between the fingerprints to a threshold. In a general fingerprinting scheme with correlated bits, however, comparing the Hamming distance to a threshold is suboptimal.

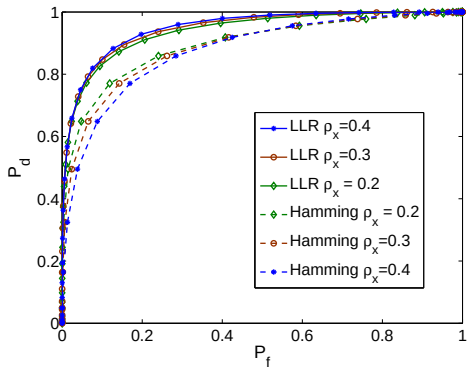


Fig. 6. Influence of correlation of the fingerprint bits on the detection performance ($p_n = \rho_n = 0.2$).

We have also described a technique to compute the probability of correct detection and the false alarm rate by estimating the density of states and have provided numerical results for a model of a block-based fingerprinting scheme. Our results show that the Likelihood Ratio statistic can compensate for correlations in the fingerprint or noise and provides a consistent performance. Increasing the marginal probability of the noise, however, lowers the detection performance. The Likelihood Ratio test also consistently outperforms the Hamming distance statistic, which was found to be sensitive to the correlations among the fingerprint and noise bits.

6. REFERENCES

[1] J. Lu, "Video fingerprinting for copy identification: From research to industry applications," in *SPIE and IS&T Media Forensics and Security*, San Jose, CA, Jan. 2009.

[2] A. L. Varna, A. Swaminathan, and M. Wu, "A Decision-Theoretic Framework for Analyzing Binary Hash-based Content Identification Systems," in *Proceedings of the ACM Workshop on Digital Rights Management*, Oct. 2008, pp. 67–76.

[3] A. L. Varna and M. Wu, "Theoretical Modeling and Analysis of Content Identification," *IEEE International Conference on Multimedia and Expo*, pp. 1529–1531, Jul. 2009.

[4] R. Kinderman and J. L. Snell, *Markov Random Fields and their Applications*, American Mathematical Society, 1980.

[5] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[6] A. Doucet and X. Wang, "Monte Carlo Methods for Signal Processing," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 152–170, Nov. 2005.

[7] F. Wang and D. P. Landau, "Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States," *Physical Review Letters*, vol. 86, no. 10, pp. 2050–2053, Mar. 2001.

[8] P. D. Beale, "Exact Distribution of Energies in the Two-Dimensional Ising Model," *Physical Review Letters*, vol. 76, pp. 78–81, 1996.