

MODELING AND ANALYSIS OF ORDINAL RANKING IN CONTENT FINGERPRINTING

Wei-Hong Chuang, Avinash L. Varna, and Min Wu

Department of Electrical and Computer Engineering, University of Maryland, College Park

ABSTRACT

Content fingerprinting provides a compact representation of multimedia objects for copy detection. This paper analyzes the robustness of the ordinal ranking module frequently used in content fingerprinting by examining the changes in ranks as local variations are introduced in feature values. Closed-form expressions to measure such sensitivity are derived when feature values are jointly Gaussian-distributed. The results show that sensitivity depends on the strength of local variation, the total number of blocks, and the correlations among block-based feature values. Experiments with both synthesized data and image data validate the analysis and provide interesting insights, inspiring an approach to reduce the sensitivity.

Index Terms— content identification, content fingerprinting, block-based fingerprinting, ordinal ranking

1. INTRODUCTION

Over the past several years, online services such as Flickr and YouTube have made content sharing popular. However, as users may upload copyrighted content to these websites, concerns about illegal content modification, duplication and digital rights infringement have also arisen. Content fingerprinting provides a solution to this problem. It extracts a compact representation, or “fingerprint”, from each multimedia object in a database, so that each object may be efficiently identified and distinguished even after moderate processing, format change, or modification. There have been a number of different content fingerprinting schemes proposed and experiments have been conducted to evaluate their performance. Readers may refer to [1] for a survey on video content fingerprinting.

Experimental evaluations, though important, do not provide generalizable and scalable understandings on how various design strategies impact the performance of content fingerprinting, especially for large databases. Theoretical modeling effort has just begun, such as the works in [2] and [3] from the authors’ research group. There is still lack of understanding of the advantages / disadvantages of various fingerprint designs. To facilitate the analysis of fingerprinting schemes, we take on a modular view on the general architecture of content fingerprinting, as shown in Fig. 1. Typically, an input video is first normalized in terms of sampling rate

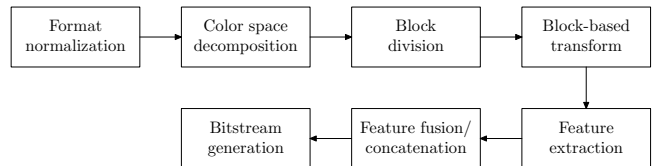


Fig. 1. General flow of fingerprint generation

(such as temporal rate, spatial resolution, and dynamic range). The color space may be transformed into, for example, luminance and chrominance channels, and the video is divided into spatial or temporal blocks in each channel. Block-based transforms, e.g., Fourier analysis or singular value decomposition, may be applied to facilitate the extraction of some continuous-valued features such as the average luminance or salient points from each block. Feature values from different blocks are then combined to form a feature vector which is converted into a bitstream. Such a decomposition process allows us to analyze and better understand the impacts of individual modules on the overall performance.

In this paper, we illustrate such a modular analysis by focusing on one of the digitization modules called “ordinal ranking” that has been widely used in block-based fingerprinting [1]. After a feature value is extracted from each block, all these values are sorted and their ranks are recorded. For example, the feature vector [7.2, 4, 8, 6.3] will be converted into [3, 1, 4, 2] via ordinal ranking. Ordinal ranking has been reported to improve robustness against noise and global transformation, but as the ranking is global, strong local variations such as logo insertion into a block may change the ranks of all blocks. This is the sensitivity issue of ordinal ranking for which extensive experimental studies have been conducted but to the best of our knowledge, there has been no quantitative analysis.

We provide a theoretical analysis of the sensitivity of ordinal ranking along with experimental validation. By assuming that feature values are jointly Gaussian, we give in Section 2 the closed-form expressions for sensitivity. In Section 3, we validate our derivation using synthesized data and image data and present a technique to reduce sensitivity inspired by the analytical modeling. Section 4 concludes this paper.

2. SENSITIVITY FORMULATION AND ANALYSIS

Sensitivity of ordinal ranking can be formulated as follows. Suppose that there are totally N blocks. From each block B_i ,

Email: {whchuang, varna, minwu}@eng.umd.edu.

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

$i = 1, 2, \dots, N$, a real-valued scalar feature X_i is extracted and is normalized to have zero mean and unit variance. The rank of block B_i is denoted by Y_i , and $\{Y_i\}$ form a set taking values from 1 to N .

2.1. Impact of Local Variation in One Block

In order to study the effect of local variation, we first consider the case where the feature value from a single block is modified and consider the general case in Section 2.2. Without loss of generality, we assume that a variational value Δ is added to X_1 , and the block ranks are recalculated and denoted by Z_i , $i = 1, 2, \dots, N$. We focus on the “ensemble behavior” in a statistical sense, to examine how the global ranking changes “on average” as a result of local variation. We define the rank difference of B_i with respect to variation Δ as $D_i(\Delta) = Z_i - Y_i$, and the sensitivity of B_i as the expected value of rank difference, $\mathbb{E}[D_i(\Delta)]$.

2.1.1. Rank Difference Due to Variation in Another Block

We first consider $D_i(\Delta)$ for $i = 2, 3, \dots, N$, and without loss of generality, $\Delta > 0$. With feature value increase in only one block, $D_i(\Delta)$ can only be 0 or -1 , since if X_1 is shifted to $X_1 + \Delta$, only ranks of feature values between X_1 and $X_1 + \Delta$ will be influenced and reduced by 1. $D_i(\Delta)$ can thus be modeled as a Bernoulli random variable, *i.e.*, $D_i(\Delta) = (-1) \times 1\{X_1 < X_i < X_1 + \Delta\}$, where $1(\cdot)$ is an indicator function. In practice, each feature value is usually an average or some combination of sample values from a block, and tends to be Gaussian as the number of values in a block is sufficiently large according to the Central Limit Theorem. If we assume $\{X_i\}$ are i.i.d. Gaussian random variables, then $X_i - X_1$ is a random variable with zero mean and variance of 2, and $\Pr\{X_1 < X_i < X_1 + \Delta\} = \Phi\left(\frac{\Delta}{\sqrt{2}}\right) - \frac{1}{2}$, where $\Phi(\cdot)$ is the c.d.f. of the standard Gaussian distribution. The expected rank difference can be expressed as

$$\begin{aligned}\mathbb{E}[D_i(\Delta)] &= -\Pr\{X_1 < X_i < X_1 + \Delta\} \\ &= -\left(\Phi\left(\frac{\Delta}{\sqrt{2}}\right) - \frac{1}{2}\right).\end{aligned}$$

We refine our model to account for the correlation commonly present between blocks, and use $\rho_{i,j}$ to denote the correlation coefficient between X_i and X_j . If X_1 and X_i are jointly Gaussian and $\rho_{1,i} < 1$, then $X_i - X_1$ is Gaussian with zero mean and variance $2(1 - \rho_{1,i})$. So for $i \geq 2$,

$$\mathbb{E}[D_i(\Delta)] = -\left(\Phi\left(\frac{\Delta}{\sqrt{2(1 - \rho_{1,i})}}\right) - \frac{1}{2}\right). \quad (1)$$

This result describes in closed form the expected rank difference of all other blocks caused by the variation in X_1 in terms of the variational amount Δ and the correlation coefficient $\{\rho_{1,i}\}$. We note that the impact of X_1 on X_i does not depend on the values of other features. Also, the increase in Δ or $\rho_{1,i}$ enlarges the expected rank difference, and $\lim_{\Delta \rightarrow \infty} \mathbb{E}[D_i(\Delta)] = -1/2$.

2.1.2. Rank Difference Due to Variation in The Same Block

We now study $D_1(\Delta)$, the rank difference of the first block. This random variable has minimum and maximum values of 0 and $N - 1$, respectively. As the closed form expression for the distribution $p_k = \Pr\{D_1(\Delta) = k\}$, $k = 0, 1, \dots, (N - 1)$ are more involved, here we derive the expected value of rank difference $\mathbb{E}[D_1(\Delta)]$. Note that when Δ is added to X_1 , if any other block B_i has a rank change $D_i(\Delta)$, block B_1 will have a corresponding rank change of $-D_i(\Delta)$. The overall rank increase of B_1 is the sum of rank changes of all the other blocks, *i.e.*, $D_1(\Delta) = -\sum_{i=2}^N D_i(\Delta)$. The sensitivity is therefore given by

$$\mathbb{E}[D_1(\Delta)] = \sum_{i=2}^N \Phi\left(\frac{\Delta}{\sqrt{2(1 - \rho_{1,i})}}\right) - \frac{N - 1}{2}. \quad (2)$$

As $\Phi(\cdot)$ is a monotonically increasing function, higher Δ and $\rho_{1,i}$'s imply larger sensitivity, and $\lim_{\Delta \rightarrow \infty} \mathbb{E}[D_1(\Delta)] = (N - 1)/2$. An important observation is that the expected difference in B_1 's rank can be seen as the sum of $(N - 1)$ individual terms involving B_1 and one other B_i only. Thus the overall effect on B_1 's rank is composed of $N - 1$ independent “pairwise interactions” with each of the other B_i s.

2.1.3. Variances of Rank Difference

The variance of the rank difference that reflects the dispersion from the expected rank difference can be analyzed similarly. Since for $i \geq 2$, $D_i(\Delta)$ is Bernoulli, and its variance equals $-\mathbb{E}[D_i(\Delta)](1 + \mathbb{E}[D_1(\Delta)])$

$$= \left(\Phi\left(\frac{\Delta}{\sqrt{2(1 - \rho_{1,i})}}\right) - \frac{1}{2}\right) \left(\frac{3}{2} - \Phi\left(\frac{\Delta}{\sqrt{2(1 - \rho_{1,i})}}\right)\right),$$

which is increasing in Δ and $\rho_{1,i}$, and has a maximum value of 1/4. For the case of $i = 1$, since $D_1(\Delta) = -\sum_{i=2}^N D_i(\Delta)$, we have

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{i=2}^N D_i(\Delta)\right)^2\right] &= \mathbb{E}\left[\sum_{i,j=2}^N D_i(\Delta)D_j(\Delta)\right] \\ &= \sum_{i,j=2}^N \mathbb{E}[D_i(\Delta)D_j(\Delta)],\end{aligned}$$

where $D_i(\Delta)D_j(\Delta)$ is again Bernoulli and equals 1 only when both $D_i(\Delta)$ and $D_j(\Delta)$ are non-zero. When $i = j$, this is simply with probability $-\mathbb{E}[D_i(\Delta)]$, and when $i \neq j$, the probability equals

$$\begin{aligned}\Pr\{X_1 < X_i < X_1 + \Delta, X_1 < X_j < X_1 + \Delta\} \\ = \Pr\{0 < X_i - X_1 < \Delta, 0 < X_j - X_1 < \Delta\}.\end{aligned} \quad (3)$$

When X_1 , X_i , and X_j are jointly Gaussian, $X_i - X_1$ and $X_j - X_1$ are also jointly Gaussian, and they are zero-mean with covariance matrix

$$\begin{bmatrix} \tilde{\sigma}_i^2 & \tilde{\rho}_{i,j} \\ \tilde{\rho}_{i,j} & \tilde{\sigma}_j^2 \end{bmatrix},$$

where $\tilde{\sigma}_i^2 = 2(1 - \rho_{1,i})$, $\tilde{\sigma}_j^2 = 2(1 - \rho_{1,j})$, and $\tilde{\rho}_{i,j} = \rho_{i,j} - \rho_{1,i} - \rho_{1,j} + 1$.

Denote the c.d.f. of $(X_i - X_1, X_j - X_1)$ by $\tilde{\Phi}(r, s)$, then

$$\tilde{\Phi}(r, s) = \Phi_{2, \hat{\rho}_{i,j}} \left(\frac{r}{\tilde{\sigma}_i}, \frac{s}{\tilde{\sigma}_j} \right),$$

where $\hat{\rho}_{i,j} = \frac{\tilde{\rho}_{i,j}}{\tilde{\sigma}_i \tilde{\sigma}_j}$ and $\Phi_{2,\rho}$ is the c.d.f. of the standard bivariate Gaussian distribution with zero mean, unit variance, and correlation coefficient ρ . Therefore, for $i \neq j$, we have

$$\mathbb{E}[D_i(\Delta)D_j(\Delta)] = \tilde{\Phi}(\Delta, \Delta) - \tilde{\Phi}(0, \Delta) - \tilde{\Phi}(\Delta, 0) + \tilde{\Phi}(0, 0). \quad (4)$$

Bringing these results together leads to:

$$\text{Var}(D_1(\Delta)) = \sum_{i,j=2}^N \mathbb{E}[D_i(\Delta)D_j(\Delta)] - \mathbb{E}^2[D_1(\Delta)].$$

When Δ is large, we have $\mathbb{E}[D_i^2(\Delta)] \approx 1/2$ for $i > 1$, $\mathbb{E}[D_i(\Delta)D_j(\Delta)] \approx \Phi_{2, \hat{\rho}_{i,j}}(0, 0)$ for $i, j > 1, i \neq j$, and $\mathbb{E}[D_1(\Delta)] \approx (N-1)/2$, so that the variance for large Δ is:

$$\lim_{\Delta \rightarrow \infty} \text{Var}(D_1(\Delta)) \approx \frac{N-1}{2} + 2 \sum_{i,j=2, i < j}^N \Phi_{2, \hat{\rho}_{i,j}}(0, 0) - \frac{(N-1)^2}{4}. \quad (5)$$

$\Phi_{2, \hat{\rho}_{i,j}}(0, 0)$ that takes value between 0 and 0.5 is a monotonically increasing function of $\hat{\rho}_{i,j}$. When all feature values are independent, for large Δ , the relative magnitude of the standard deviation with respect to the expected rank difference is $\sqrt{(N+1)/(3(N-1))} \geq \sqrt{1/3} = 0.5774$. Since the variance of rank difference is not negligible, there will be considerable variations in the rank difference between different realizations.

2.2. Local Variations in More Than One Block

We now consider the general case where each of the feature values may be altered. We model this by the addition of a variational value Δ_i to the feature X_i from the i^{th} block. Assume again that $\{X_i\}$ are jointly zero-mean Gaussian and the correlation coefficient between X_i and X_j is denoted by $\rho_{i,j}$.

Suppose we have stored the feature values according to their original ranks in an array, and we add a value Δ_i to each X_i . Now consider the process of re-sorting this array. In the first step, the values in the first and second positions are compared, say $X_i + \Delta_i$ and $X_j + \Delta_j$. If $X_i + \Delta_i > X_j + \Delta_j$, then the two values are swapped, so that the rank of block B_i increases by 1 and correspondingly the rank of B_j decreases by 1; if $X_i + \Delta_i < X_j + \Delta_j$, then the ranks are not altered. Denote by $D_{i,j}(\Delta_i, \Delta_j)$ the change in the rank of block B_i when compared with block B_j in this resorting process. After this step, the values in the second and third positions are compared and this process is repeated until the array is sorted. The overall change in the rank of B_i after the resorting is the sum of changes that occur due to these pairwise comparisons, $D_i(\Delta_1, \Delta_2, \dots, \Delta_N) = \sum_{j:j \neq i} D_{i,j}(\Delta_i, \Delta_j)$ and the expected value $\mathbb{E}[D_i(\Delta_1, \Delta_2, \dots, \Delta_N)] = \sum_{j:j \neq i} \mathbb{E}[D_{i,j}(\Delta_i, \Delta_j)]$. By following a derivation similar to that in Section 2.1.1, it

can be shown that

$$\mathbb{E}[D_{i,j}(\Delta_i, \Delta_j)] \triangleq \begin{cases} \Phi \left(\frac{\Delta_i - \Delta_j}{\sqrt{2(1 - \rho_{i,j})}} \right) - \frac{1}{2}, & \Delta_i > \Delta_j; \\ -\Phi \left(\frac{\Delta_j - \Delta_i}{\sqrt{2(1 - \rho_{i,j})}} \right) + \frac{1}{2}, & \Delta_i < \Delta_j; \\ 0, & \Delta_i = \Delta_j. \end{cases}$$

For the case when only X_1 is modified, *i.e.*, $\Delta_1 = \Delta$ and $\Delta_i = 0$ for $i > 1$, the above expected rank difference reduces to the results in Section 2.1.1.

The calculation of the variances for the general case could be conducted in a similar manner to Section 2.1.3 and is omitted here due to space limit.

3. EXPERIMENTAL RESULTS

In this section, we verify our derivations using both synthesized data and a practical content fingerprinting scheme [4].

3.1. Verification Based on Synthesized Data

3.1.1. Local Variation in One Block

We generate 10^5 samples of a 4×4 feature matrix with values X_1, X_2, \dots, X_{16} extracted from blocks B_1, B_2, \dots, B_{16} , in raster scan. The whole feature matrix is jointly Gaussian-distributed, each entry has zero mean and unit variance, and the correlation between entries X_i and X_j is set as $\rho_0^{d(B_i, B_j)}$ where ρ_0 is some specified constant and $d(\cdot)$ is the Euclidean distance. This correlation matrix is a good model for the correlation of feature values observed in practice, where blocks that are closer are expected to have higher correlation than blocks that are farther apart.

When only B_1 is influenced by local variation, we plot the sample mean of $D_1(\Delta)$ with respect to different values of Δ and ρ_0 as shown in Fig. 2, where predicted values based on our derivation are plotted in solid lines, and values for synthesized data are plotted with markers. We see that our derivation predicts the expected rank difference perfectly. For all values of ρ_0 , when Δ becomes sufficiently large, the sensitivity converges to $(N-1)/2 = 7.5$, as predicted. It can also be seen that larger values of ρ_0 , or larger correlations among feature values, will cause higher sensitivity.

The standard deviation of $D_1(\Delta)$ in this case is also examined, and the results for different ρ_0 and Δ are given in Fig. 3, where curves associated with theoretical results and synthesized data are plotted without and with markers, respectively. The derivation matches the simulation, too. The figure shows that 1) the standard deviation follows a similar trend to the expected value with respect to Δ and ρ_0 ; 2) an increase in ρ_0 results in an increase in the variance, including the limiting values for large Δ . This suggests that larger ρ_0 not only leads to higher sensitivity, but makes the sensitivity less predictable.

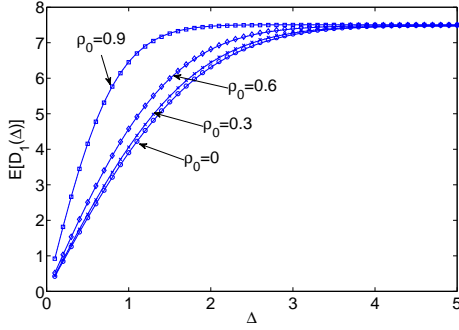


Fig. 2. Sensitivity of the 1st block with respect to Δ for different values of ρ_0

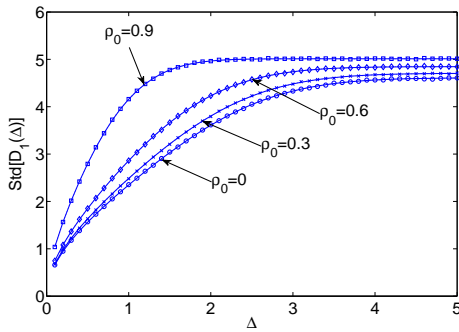


Fig. 3. Standard deviation of rank difference of the 1st block with respect to Δ for different values of ρ_0

3.1.2. Local Variations in More Than One Block

When more than one block is influenced by local variations, the analysis of sensitivity becomes more involved. We consider the scenario where two blocks B_1 and B_2 have local variations Δ_1 and Δ_2 , respectively, in two representative settings. In the first setting, we assume that all feature values are independent, and examine the sensitivity as a function of Δ_1 and Δ_2 . The sensitivity values for the first block B_1 and the third block B_3 that is nearby are shown in Fig. 4 and 5, respectively. $\mathbb{E}[D_1(\Delta_1, \Delta_2)]$ is mainly determined by Δ_1 , and the dependency of $\mathbb{E}[D_3(\Delta_1, \Delta_2)]$ on Δ_1 and Δ_2 are the same.

In the other setting, we assume that blocks are correlated with $\rho_0 = 0.3$, and consider the sensitivity of block B_3 when 1) only B_1 is distorted by Δ ; 2) only B_2 is distorted by Δ ; 3) B_1 and B_2 are equally and respectively distorted by $\Delta/2$. This setting imitates logo insertion that may interfere with video fingerprints in practice. These different cases could occur since logos may not be inserted into a single block but across the boundary, and thus the distortion in each block is reduced. We consider the rank difference of a nearby block, B_3 , due to these local variations, and the well-matched results for theoretical prediction and synthesized data are shown in

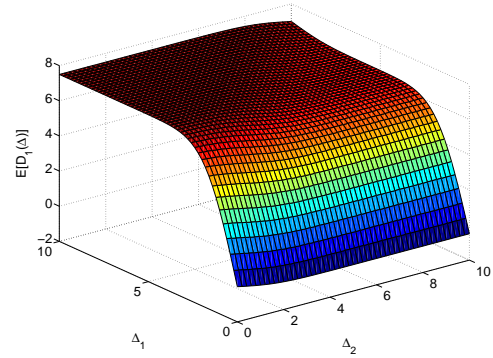


Fig. 4. Sensitivity of the 1st block with respect to Δ_1 and Δ_2

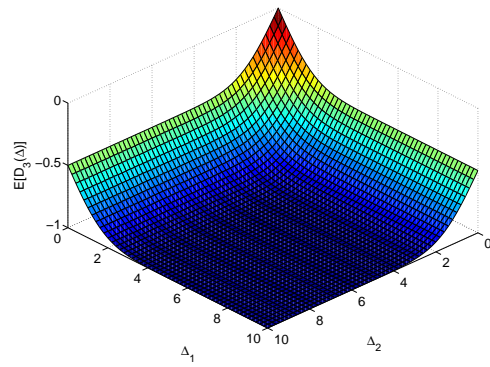


Fig. 5. Sensitivity of the 3rd block with respect to Δ_1 and Δ_2

Fig. 6 with and without markers, respectively. When comparing B_1 and B_2 , adding the same local variation into B_2 will have more impact on B_3 's rank, since B_2 is closer to B_3 than B_1 . Another interesting observation is that, having local variation in a single block has a larger influence on other blocks' ranks than dividing such local variation into two blocks.

3.2. Expected Rank Difference in Practical Fingerprints

To verify how well our analytical modeling characterizes the behavior of ordinal ranking, we examine a practical ordinal-ranking-based video fingerprinting scheme proposed in [4]. For each frame in a video, this scheme partitions the frame into a fixed number of blocks, calculates the average intensity level of each block, and performs ordinal ranking to obtain a robust representation of the frame. In our implementation, we apply the frame-based fingerprint generation to an image set consisting of 1000 images from the Corel image database. We add Δ to each block in turn, calculate the mean of rank difference over these image samples, and take the overall average of these mean rank difference values to obtain the expected rank difference when one block is influenced by local variation Δ . Resulting sensitivity values are plotted in Fig. 7 along with the theoretical predictions calculated using correlation values

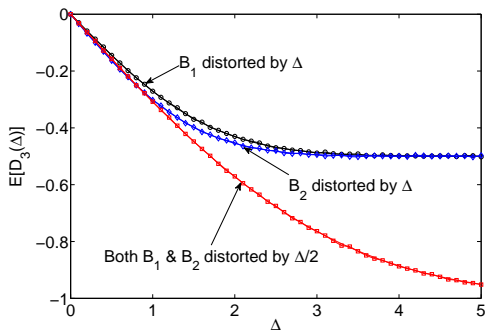


Fig. 6. Different assignments of local variation in two blocks. Well-matched results for theoretical prediction and synthesized data are shown with and without markers, respectively.

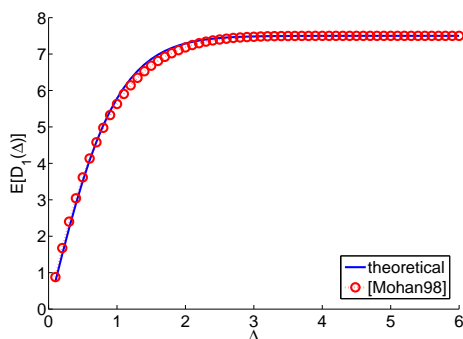


Fig. 7. Sensitivity of [4] compared to theoretical prediction

estimated from the image set. We see that the theoretical prediction matches results for image features extracted by a real fingerprinting scheme.

3.3. Sensitivity Reduction via Feature Masking

From our analysis, a smaller or negative correlation leads to a lower sensitivity. This is because when the correlation coefficient between two feature values is close to 1, it is likely that changing one of them significantly causes their ranks to be exchanged. If the correlation is close to -1 , these two values are statistically farther apart, so that changing one of them will have much less influence on the ranks. Guided by such an observation, we propose a “masking” technique to alleviate the sensitivity. For any given block-based fingerprint, before ordinal ranking is performed, we can do element-wise multiplication of the feature matrix and mask matrix below that contains alternating signs:

$$\begin{bmatrix} -1 & 1 & \cdots & -1 & 1 \\ 1 & -1 & \cdots & 1 & -1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & 1 & \cdots & -1 & 1 \\ 1 & -1 & \cdots & 1 & -1 \end{bmatrix}.$$

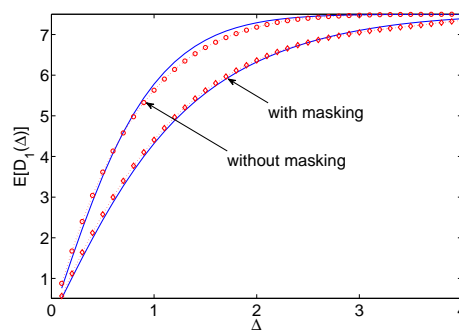


Fig. 8. Sensitivity reduction by the masking technique

The effect of applying such a mask is to change the sign of correlation between adjacent blocks. Since adjacent blocks usually have higher correlations, this step makes the overall correlations among blocks much less positive. We apply the masking technique to the scheme in [4] and show the expected rank differences obtained with and without masking in Fig. 8. As we can see, applying the proposed masking technique indeed significantly reduces the sensitivity.

4. CONCLUSIONS

A modular perspective improves our general understanding of the essential components of various content fingerprinting schemes. In this paper, we take this viewpoint to consider the popular ordinal ranking module used in block-based fingerprinting, and quantify its sensitivity. We provide closed-form expressions for the Gaussian case, and show that these expressions are valid and offer useful insights. Inspired by our analysis, we also provide a masking method to reduce the sensitivity.

5. REFERENCES

- [1] J. Lu, “Video Fingerprinting for Copy Identification: from Research to Industry Applications,” in *Proc. of SPIE Media Forensics and Security*, Jan. 2009.
- [2] A. L. Varna, A. Swaminathan, and M. Wu, “A Decision-Theoretic Framework for Analyzing Binary Hash-based Content Identification Systems,” in *Proc. 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,” to appear in the *IEEE Int’l Conf. on Multimedia and Expo*, Jul. 2009.
- [4] R. Mohan, “Video Sequence Matching,” in *Proc. of Int’l Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Apr. 1998, vol. 6, pp. 3697–3700.