SECURITY OF FEATURE EXTRACTION IN IMAGE HASHING

Ashwin Swaminathan, Yinian Mao and Min Wu

ECE Department, University of Maryland, College Park, U.S.A.

ABSTRACT

Security and robustness are two important requirements for image hash functions. In this paper, we introduce "differential entropy" as a metric to quantify the amount of randomness in image hash functions and to study their security. We present a mathematical framework and derive expressions for the proposed security metric for various common image hashing schemes. Using the proposed security metric, we discuss the trade-offs between the security and robustness in image hashing.

1. INTRODUCTION

In the modern era, there is a widespread availability of multimedia data in the digital form. This has led to a tremendous growth of tools to manipulate digital data. To ensure trustworthiness, content based image authentication techniques like image hashing have been proposed. A hash function is a short digital signature of the data [1]. Image hashing has been used in authentication, content based image retrieval (CBIR) and image/video watermarking [2, 3].

For the applications listed above, it is often necessary that the image hash function be robust and secure. By robustness, we mean that the hash function should be resilient to a set of content preserving manipulations such as filtering; geometric and other affine transformations; additive noise; compression; and luminance non-linearities. The hash should also depend on the secret key for applications involving authentication and image/video watermarking. Furthermore, the hash should not be easily forged or estimated without the knowledge of the key.

Traditionally, the cryptographic hash functions have been used in applications involving verification of data integrity and data retrieval. Although they are very secure, these hash functions are not robust as they are very sensitive to every bit of the image data. This is undesirable and inconsistent with human visual perception [3]. As a result, various robust image hashing schemes have been proposed [3, 4, 5]. Many of them follow a three-step framework to attain robustness. This involves extraction of certain invariant features from the image (*Feature Extraction*), quantizing and compressing them [3, 6]. To secure the hash, the key can be employed in any of the three stages. Fridrich et al. have indicated that the feature extraction stage of the algorithm must be key dependent for the hash to be secure [4]. The authors have argued that if the feature extraction stage is not secure, then the output of this stage would be publicly known and therefore an attacker can forge a new image that would give rise to the same features thus defeating the purpose of hashing.

In the present work, we will show that the "differential entropy" can be used as a metric to evaluate the security of the feature extraction stage. We present an evaluation framework and do an information theoretical analysis to obtain the differential entropy of various existing schemes. We then present comparison studies and discuss the trade-offs between the security and robustness for these schemes.

The authors can be contacted via email {ashwins, ymao, minwu} @eng.umd.edu.

2. SECURITY ANALYSIS

There are a number of image hashing schemes presented in the literature. Each of them introduces security in the feature extraction stage in a unique way. For instance, both Venkatesan's scheme [3] and Mihcak's scheme [7] introduce security by the choice of random rectangles from which features are generated; Fridrich et al. introduce security by projecting the image onto random low-pass images [4]; and in our recent work [5], we introduce randomness by performing a weighted summation of the discrete polar Fourier transform over random subsets.

To our best knowledge, there is no metric to compare the degree of security of image hash functions. The only relevant work to characterize security is by Radhakrishnan et al. [8]. In their work, the authors show that the Visual Hash Function (VHF) [4] is not secure and one can create another visually dissimilar image that would give the same hash values. However, their analysis is specific to the VHF and cannot be easily extended to characterize the degree of security of other commonly used hash functions.

In our analysis, we use the differential entropy ($\aleph$) as a metric to characterize the amount of randomness in hash values. The higher the differential entropy of the hash value, the higher the randomness and the larger the number of exhaustive searches required to forge the hash value-h (which is proportional to $\alpha^{\aleph(h)}$ for some $\alpha > 1$). The schemes that do not have any random components in the feature extraction stage have differential entropy of $-\infty$ by definition and the number of exhaustive searches required to forge the

hash is 0 as expected.

In the subsequent analysis, we model the output of the feature extraction stage as random variables and find its degree of uncertainty in terms of differential entropy. We study three existing hashing schemes, namely, (A) Random Discrete Polar FFT based scheme [5], (B) Fridrich's VHF [4] and (C) Venkatesan's robust image hashing [3]. In the analysis, we assume that the attacker has complete information about the image and the hashing algorithm used but not the key used in the generation of the hash.

2.1. Security Analysis of Scheme A

The first scheme we analyze is the hashing scheme-2 presented in [5]. In this scheme, the FFT of the image is first obtained and converted to polar coordinates to obtain $I'(\rho,\theta)$. This is sampled along the ρ-axis and the θ-axis to obtain $I'\left(\rho_i, \frac{(2j+1)\pi}{K}\right)$ (where $1 \le i \le N$ and $0 \le j \le K-1$). A weighted summation is performed along a random subset of the discretized ρ-axis to form the k^{th} hash value h_k. The weights of the summation are Gaussian distributed random variables $\{\beta_{ik}\}$ with mean μ and variance σ^2. The hash values can be expressed as

$$h_k = \sum_{i=1}^{N} \lambda_{ik}\beta_{ik}q_{\rho_i}, \quad \text{where} \tag{1}$$

$$q_{\rho_i} = \sum_{j=0}^{K-1} \left| I'\left(\rho_i, \frac{(2j+1)\pi}{K}\right)\right|, \tag{2}$$

and λ_{ik} are Bernoulli distributed random variables.

To find the differential entropy $\aleph(h_k)$, we first try to obtain the probability density function (PDF) $f_{h_k}(x)$ of the random variables h_k. A detailed analysis of the system reveals that the PDF has rather a complex form and contains 2^N terms. Thus, obtaining $\aleph(h_k)$ from the PDF is impractical. We instead analyze the system to find the lower and the upper bounds. Without loss of generality, we assume that $\rho_1 < \rho_2 < \ldots < \rho_N$ and therefore by the energy compaction property of the Fourier transform we have $q_{\rho_1} \ge q_{\rho_2} \ge \ldots \ge q_{\rho_N}$ for most natural images. We additionally use the fact that the entropy is a concave function to obtain the lower bound

$$\aleph(h_k) \ge \frac{2^N-1}{2^{N+1}}\log_2(2\pi e\,\sigma^2 q_{\rho_N}^2) + \frac{1}{2^N}\sum_{i=1}^{N}\binom{N}{i}\log_2(i) \tag{3}$$

To derive the upper bound, we note that of all distributions with the same variance the Gaussian has the maximum differential entropy. Therefore, we find the variance of the hash values h_k and obtain the upper bound on $\aleph(h_k)$ using the differential entropy of the Gaussian distribution as follows

$$\aleph(h_k) \le \frac{1}{2}\log_2\left((2\pi e)\left(\frac{\sigma^2}{2}+\frac{m^2}{4}\right)\sum_{i=1}^{N} q_{\rho_i}^2\right). \tag{4}$$

2.2. Security Analysis of Scheme B

In Fridrich's scheme, uniformly distributed random images $(X^{(r)})$ are generated using a secret key [4]. The resulting random images are then spatially averaged with a $m \times n$ filter $\{\alpha_{ij}\}$. The output of the filtering operation $Y^{(r)}$ is

$$Y_{kl}^{(r)} = \sum_{i=-\lfloor \frac{m}{2}\rfloor}^{\lfloor \frac{m}{2}\rfloor}\sum_{j=-\lfloor \frac{n}{2}\rfloor}^{\lfloor \frac{n}{2}\rfloor} \alpha_{ij}X_{i+k,j+l}^{(r)}. \tag{5}$$

The image I is then projected on the N-randomly smooth patterns $\{Y^{(r)}, r = 1,2,\ldots,N\}$ to obtain an intermediate hash value h_r. These intermediate values are then quantized to generate the final hash. It can be shown that

$$h_r = \sum_{k=1}^{H}\sum_{l=1}^{W} Y_{kl}^{(r)}I_{kl} = \sum_{i=-\lfloor \frac{m}{2}\rfloor}^{\lfloor \frac{m}{2}\rfloor}\sum_{j=-\lfloor \frac{n}{2}\rfloor}^{\lfloor \frac{n}{2}\rfloor} \alpha_{ij}V_{ij}^{(r)},$$

$$\text{where} \quad V_{ij}^{(r)} = \sum_{k=1}^{H}\sum_{l=1}^{W} X_{i+k,j+l}^{(r)}I_{kl}.$$

We note that $V_{ij}^{(r)}$ is a weighted sum of $W \times H$ uniformly distributed random variables $\{X_{ij}^{(r)}\}$ with the weights determined by the image pixel values I_{ij}. We can therefore assume that $V_{ij}^{(r)}$ are Gaussian distributed. We also note that the variables $V_{ij}^{(r)}$ are highly correlated and

$$E(V_{ij}^{(r)}V_{ab}^{(r)}) = \frac{1}{12}\sum_{k=1}^{H}\sum_{l=1}^{W} I_{kl}I_{i+k-a,j+l-b} + \left(\frac{1}{2}\sum_{k=1}^{H}\sum_{l=1}^{W} I_{kl}\right)^2 \tag{6}$$

Since h_r is a sum of Gaussian distributed variables $V_{ij}^{(r)}$, it would also be Gaussian. Therefore, we compute the differential entropy of h_r from its variance.

$$\aleph(h_r) = \frac{1}{2}\log_2\left(2\pi e\,\frac{1}{12}\sum_{p=1}^{H}\sum_{q=1}^{W} I_{pq}I_{pq}^{(\alpha\alpha)}\right) \tag{7}$$

where $I^{(\alpha\alpha)}$ is obtained by filtering the image I twice with the filter $\{\alpha_{ij}\}$.

2.3. Security Analysis of Scheme C

This scheme first takes a 3-level DWT of the image. A random tiling of each subband of the DWT of the image is generated. The mean (or the variance) of the pixel values in the random rectangle is used to form the feature vector [3]. This feature vector is randomly quantized and compressed.

We present a slightly modified approach to analyze the security of this scheme by taking on the attackers' viewpoint. We shall show that the locations and sizes of the

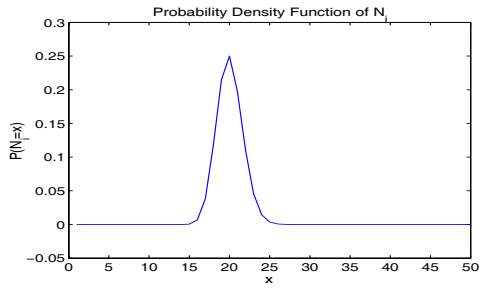

Fig. 1. The plot of the PDF of N_i-the number of blocks in i^{th} row. Parameters $w_{min} = 10$, $w_{max} = 40$ and $W = 512$

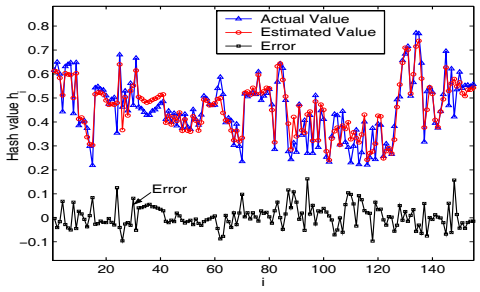

Fig. 2. The plot of actual, estimated and error in estimation of the image statistics vector for the Lena image with $w_{min} = 10$, $w_{max} = 40$ and $W = 512$

exact random partitions are not required to forge the hash. The attacker can instead make an intelligent guess of image statistics based on his/her knowledge of the image and by replacing the random partitions with uniformly spaced, equal sized partitions. If the attacker is correct in estimating the number of partitions N_i in each row and the number of rows M, then he/she can get a good estimate of the hash vector.

As a first step of the analysis, we derive a model for N_i and M by modelling the block partitioning algorithm. The block partitioning algorithm can be approximated as a combination of two 1-D problems - partitioning along the horizontal direction and then along the vertical direction. To partition along the width of the image, we generate random numbers $\{U_i\}$ uniformly distributed in $[w_{min}, w_{max}]$ where w_{min} and w_{max} are the minimum and the maximum widths of the random block. The location of the n^{th} partition is then given by a set of random variables T_n ($T_n = \sum_{i=1}^{n} U_i$). We use a Gaussian approximation for the PDF of T_n to obtain the PDF of N_i. A sample of the PDF of N_i is shown in Fig. 1. It can be shown that the maximum likelihood (ML) estimate of the random variable N_i is the mean (m_{N_i})

$$N_{est} = m_{N_i} = \frac{2W}{w_{max} + w_{min}}. \tag{8}$$

Similarly, the ML estimate of the random variable M can be shown to be

$$M_{est} = \frac{2H}{h_{max} + h_{min}}, \tag{9}$$

where h_{min} and h_{max} are the minimum and the maximum heights of the random block.

Once the number of rows and columns are estimated, the attacker can estimate the image statistics using uniform size partitions of size $(\frac{W}{N_{est}}) \times (\frac{H}{M_{est}})$. We plot the actual hash values, the estimated and the error in Fig. 2. Note that the error has a much lower dynamic range than the actual value. The differential entropy($\aleph$) of the scheme can be estimated using the error. For the Lena image, $\aleph(h_k)$ is around 5.74.

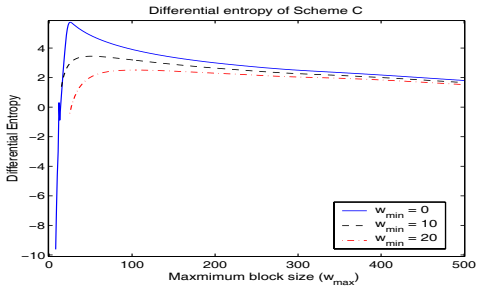

Fig. 3. The entropy of the Venkatesan's scheme plotted for different input values of w_{max} and w_{min}. $W = H = 512$, $w_{min} = h_{min}$ and $w_{max} = h_{max}$.

An alternate way to analyze the security of the scheme is by considering the effects of the synchronization errors introduced by a wrong estimate in N_i (and M). We represent the number of synchronization errors in the n^{th} row in the form of $Y_n = \sum_{i=1}^{n}(N_i - N_{est})$. To obtain the upper bound for the differential entropy, we construct the MxM correlation matrix (R) using $R_{ii} = E(Y_i^2) = i\sigma_N^2$ and $R_{ij} = E(Y_i Y_j) = \min(i,j)\sigma_N^2$ where σ_N^2 is the variance of N_i. It can be shown that $|R| = \sigma_N^{2M}$ and therefore the differential entropy of the hash value is

$$\aleph(h_k) = \frac{1}{2}\log_2(2\pi e\sigma_N^2) + \frac{1}{2M_{est}}\log_2\left(1 + \frac{1}{12\sigma_N^2}\right) \tag{10}$$

We can see that the differential entropy of the scheme heavily depends on the value of the variance σ_N^2. At very low w_{max}, we have $\sigma_N^2 \to 0$ and therefore $\aleph(h_k) \to -\infty$ as shown in Fig. 3. This result is expected because when w_{max} approaches w_{min}, there is no longer any randomness in the choice of the window widths and hence the scheme would no longer be secure.

2.4. Numerical Results and Comparison

We show in Fig. 4 the plot of the derived lower and the upper bounds of the entropy of the *Scheme A* with the number

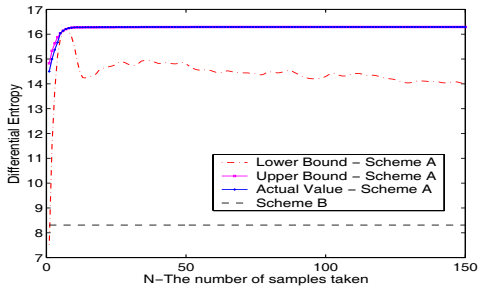

Fig. 4. Entropy of the hash values for the *Scheme A* plotted with respect to the number of sampling points N. Results for *Scheme B* is also shown for comparison

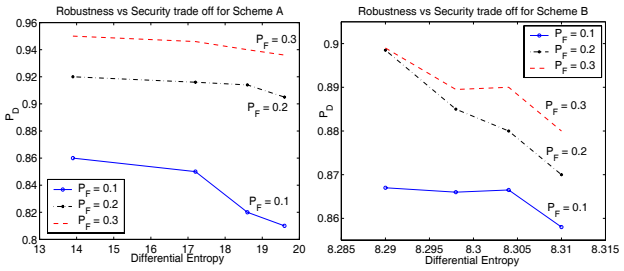

Fig. 5. Robustness and Security trade-off for the Scheme A (left) and B (right)

of sampling points N. To evaluate these bounds, we numerically compute a PDF for the hash values and calculate the true differential entropy of the PDF. Note that the upper bound plotted using equation (4) is a very tight upper bound and is almost equal to the calculated differential entropy. This is because, the true PDF of the hash values is almost Gaussian with the same mean and variance as those used in the upper bound calculation. We observe that the differential entropy of *Scheme A* is greater than that of *Scheme B*. This is a consequence of the filtering operations in *Scheme B*, which reduces the variance of the random variables and hence its entropy. The differential entropy of *Scheme C* is lower than those of *Schemes A* and *B* (compare Fig. 3 and 4). This is because, in *Scheme C*, the image statistics can be estimated to reasonable accuracy without the knowledge of the exact block partitions. However, in other schemes, the attackers need to guess some random variables used in computing features (β_{ik} in *Scheme A*, $Y^{(r)}$ in *Scheme B*).

Notice that we only consider the security of the feature extraction stage in this work. While random permutation or other techniques alike can be applied to almost all feature extraction approaches to bring further randomness, the same type of post processing often enhances the overall security by about the same amount. This does not change the relative security results between the schemes obtained in this work, and thus justifies our focus on the feature extraction stage.

3. DISCUSSIONS AND CONCLUSIONS

In all the three schemes that we have studied, we have observed a trade-off between the security and the robustness in the hashing schemes. We use differential entropy as a security metric; and assess the amount of robustness based on the Receiver Operating Characteristics (ROC) [5]. The probability of correct decision (P_D) is computed for a given probability of false alarm (P_F), based on the performance of the algorithm for various authentic modifications. A higher P_D for the same P_F indicates more robustness. A comparative study of robustness of various schemes is presented in our recent work [5]. In this work, we discuss the trade-off between security and robustness.

For *Scheme A*, we observe from the lower and the upper bounds that as the variance of the random variables $\{\beta_{ik}\}$ is increased, the bounds on differential entropy (hence the security) increase, while the robustness (in terms of P_D for the same P_F) decreases as shown in Fig. 5. Similar trend can also be observed for *Schemes B and C*. For example, in *Scheme B*, it can be shown that as the order of the filter increases, the entropy decreases and robustness increases. This is expected as increasing the order of the filter implies more averaging and therefore less randomness and more robustness.

In summary, in this paper, we have introduced the differential entropy as a metric to study the security in image hashing systems. We then formulate a method and derive explicit expressions for differential entropy for the feature extraction stage of various existing schemes. We have presented comparison studies and discussed the trade-offs between security and robustness for existing schemes.

4. REFERENCES

[1] M. K. Mihcak and R. Venkatesan, "A Tool For Robust Audio Information Hiding: A Perceptual Audio Hashing Algorithm," *Proceedings of 4th Intl. Information Hiding Workshop*, PA, April 2001.

[2] J. Fridrich, "Visual Hash for Oblivious Watermarking," *Proc. of IS&T/ SPIE's 12th Sym. on Electronic Imaging*, Vol. 3971, USA, Jan 2000.

[3] R. Venkatesan, S. M. Koon, M. H. Jakubowski and P. Moulin, "Robust Image Hashing," *IEEE Proc. ICIP*, Vol. 3, pp. 664 - 666, Sep 2000.

[4] J. Fridrich and M. Goljan, "Robust Hash functions for Digital Watermarking," *IEEE Proc. Intl. Conf. on Information Technology: Coding and Computing*, pp. 178 - 183, 27-29 March 2000.

[5] A. Swaminathan, Y. Mao and M. Wu, "Image Hashing Resilient to Geometric and Filtering Operations," *IEEE Workshop on Multimedia Signal Processing*, Sep 2004.

[6] M. Johnson and K. Ramachandran, "Dither-based Secure Image Hashing using Distributed Coding," *Proc. IEEE ICIP*, Spain, Sep 2003.

[7] M. K. Mihcak and R. Venkatesan, "New Iterative Geometric Methods for Robust Perceptual Image Hashing," *Proc. of ACM Workshop on Security & Privacy in Digital Rights Management* PA, Nov. 2001.

[8] R. Radhakrishnan, Z. Xiong and N. Memon, "On the Security of the Visual Hash Function," *Proceedings of SPIE* , Vol. 5020, 2003.