Perceptual Hash Based Blind Geometric Synchronization of Images for Watermarking

Baris Coskun[a] and M. Kıvanç Mıhçak[b]

[a]Polytechnic University, Electrical and Computer Engineering Dept, Brooklyn , NY, USA
[b]Boḡaziçi University, Electrical and Electronics Engineering Dept, Istanbul, Turkey

ABSTRACT

In this work, we consider the problem of blind geometric synchronization of images for watermarking. Existing solutions involve insertion of periodic templates, geometrically-invariant domains, and feature-point-based techniques. However, security leakage and poor watermark detection performance under under lossy geometric attacks are some known disadvantages of these methods. Different from above, recently a perceptual-hash-based secure and robust image synchronization scheme has been proposed. Although it has some promising results, it requires a series of heavy computations which prevents it from being employed in real time (or near real time) applications and from being extended to wider range of geometric attack models. In this paper, we focus on the computational efficiency of this scheme and introduce a novel randomized algorithm which conducts geometric synchronization much faster.

Keywords: Geometric synchronization, watermarking, robust perceptual hashing, computational efficiency.

1. INTRODUCTION

Watermarking targets many applications in the area of content security including authentication and fingerprinting. Most of these applications require the employed watermark to be robust against perceptually insignificant modifications on the content, such as images in this case. In other words, the watermark detector should always be able to detect the watermark as long as the received image is perceptually similar to the original image.

When common signal processing modifications such as compression, image enhancement, independent noise addition, filtering and image smoothing etc. are considered, watermarking research has been quite successful in developing reliable solutions. However, when it comes to geometric modifications, classical watermarking systems usually experience drastic decrease in detection performance even though the modification is perceptually insignificant. Consequently, being an unsolved and important problem, the problem of designing secure watermarking schemes, that are robust to geometric attacks, continues to attract attention of researchers. There have been different strategies proposed for achieving robustness against geometric attacks, including periodic insertion of the mark,[1] template insertion,[2] mark embedding in geometrically invariant domains,[3] and content-based watermarking schemes that extract image feature points.[4]

However, certain shortcomings can be identified for these methods. The major drawback of periodic marks and template-insertion-based approaches is that the redundancy in periodic marks may lead to the leakage of information, and thus may allow attackers to estimate and remove the watermark.[5] Inserting mark in geometrically invariant domains provides robustness only to a fraction of geometric modifications (e.g., global affine transforms only) and therefore the watermark remains fragile against a wide variety of other attacks. In addition, visually pronounced artifacts start to appear for even a small watermark strength in geometrically invariant domains. Although feature-point-based methods have good robustness performance against geometric modifications, their performance strongly depends on the implicit and optimistic assumptions on the feature point detector; a large

Further author information:
Baris Coskun: E-mail: baris@isis.poly.edu, Telephone: +1 (718) 260 41 40
M. Kıvanç Mıhçak: E-mail: kivanc.mihcak@boun.edu.tr, Telephone: +90 (212) 359 64 14
Part of the work was carried out while Baris Coskun and M. Kıvanç Mıhçak were with Microsoft Research.
M. Kıvanç Mıhçak was partially supported by TÜBITAK Career Award, No. 106E177.

fraction of the feature points extracted from the watermarked original image and the received image need to match exactly.

Radically different from above approaches, in[6] Harmanci *et. al.* use perceptual robust hashes to convey information about the geometry of the original image to the watermark detector. The basic idea is to modify the image prior to watermark embedding so that the hash values of the pseudo-random regions of the image meet certain conditions. This later lets receiver search over geometric transform parameters (e.g., under affine transform models), whose inverse transform maps the received image to the original image. During the search process, receiver solves an optimization problem by minimizing a cost function which is the distance between the original hash values at the embedder and their received variants at the receiver. Finally, applying watermark detector to the geometrically synchronized image has been shown to significantly increase the detection performance.

Despite the promising detection performance results of the hash-based method presented in,[6] due to the intense computations in the search process, the scheme can hardly be employed in practical applications yet. Furthermore, although the scheme can be extended to any arbitrary geometric transformation model in theory, practically this is a non-trivial task since the search complexity increases exponentially in the number of model parameters. The number of model parameters required to cover a wide range of geometric modifications may be quite high. Therefore developing a computationally efficient synchronization algorithm is crucial. For this reason we focus on the computational efficiency problem while following a similar philosophy to the one in.[6]

In this work a novel and efficient search algorithm to estimate geometric attack parameters is introduced. The efficiency of the algorithm comes from two novel approaches:

1. A new set of conditions, which the embedder makes the hash values meet, are introduced so that the cost function on the receiver side becomes smoother and more suitable for efficient optimization.
2. A novel randomized search algorithm is introduced for faster convergence.

As a result, we observe that the search algorithm becomes approximately 8 times faster (with respect to the one proposed in[6]) in the expectation sense when a 2-parameter geometric attack model is used. Furthermore, search efficiency gain is expected to increase exponentially as the number of attack parameters increases

2. PRIOR ART AND OVERVIEW OF PROPOSED SCHEME

In this work, we follow a similar philosophy to the one in,[6] where on the embedder side, the image (or its suitably transformed version) is divided into N pseudo-random regions/sub-images using a secret key K_H and their robust hash values are calculated, $\{h_i\}_{i=1}^N$. Then N pseudo-random numbers are generated corresponding to each region, $\{b_i\}_{i=1}^N$. Afterwards, the regions are sorted in increasing order of the deviation of their hash values from the respective b_i's and top M regions are selected. Then, the image is modified to meet a certain relation between b_i and $\tilde{h}_i$ for $1 \leq i \leq M$, where $\tilde{h}_i$ denotes the hash value of the i-th selected region of the modified image. Particularly in,[6] this relation is chosen to be $\tilde{h}_i = b_i$. This modification on the image is referred to as ***Hash Distortion Compensation*** (HDC). Consequently a modified image $\tilde{I}$ is obtained, which is subsequently watermarked. On the other hand, at the receiver N pseudo-random regions $\{R_i\}_{i=1}^N$ are obtained from the received image I_r, and N pseudo-random numbers $\{b_i\}_{i=1}^N$ (which are the same as the ones used at the embedder) are generated. Here, let G denote our geometric attack model. Then, a search is performed for the optimum geometric distortion model G^*, such that the hash values of the top M regions approximately match the previously decided relation with b_i. After that, one can apply $(G^*)^{-1}$ on I_r to obtain the "synchronized image" I_r^* and apply watermark decoding.

In HDC-based synchronization, a hash function having the following properties can be employed:

- **Robustness:** Hash values are robust against common (geometry-preserving) signal processing modifications.
- **Uniqueness:** Perceptually different inputs yield different hash functions.

- **Graceful Degradation:** Hash values slowly (smoothly) deviate from their original values as the magnitude of the geometric attack increases.

Without robustness, matching hash values may possibly not be found since no two corresponding image regions are exactly the same even if they are geometrically synchronized. Without uniqueness, the receiver would get confused since there would be many matching hash values even if their perceptual content is different. Finally, graceful degradation is crucial to solve the resulting optimization problem efficiently at the receiver via searching the correct parameter values. With graceful degradation, the search algorithm can identify and follow the correct search direction such that the hash values get closer to their true values.

Generalized version of the perceptual hash function proposed by Venkatesan *et. al.* in[7] is used in this work since it possesses the aforementioned properties to some extent. In this method, hash values of image regions are given by weighted linear combinations of the DC subband coefficients in these regions after applying the DWT transform. Weights are chosen from a smoothly-varying Gaussian random field in order to provide both robustness and graceful degradation properties. Also, in this work we use the same watermarking algorithm as in,[6] which is originally proposed in.[8] Finally, the geometric transformation model is chosen as the global affine transformation, described in.[6]

The search process at the receiver side is a *gradient-based search*, which is actually a numeric optimization (in the geometric distortion parameter space) procedure trying to minimize a cost function, dependent on the pseudo-random sequence $\{b_i\}$ and the hash values of the received image. For details of the gradient-based search, the reader is referred to.[6] Both in this work and,[6] the cost function is defined as

$$C(p) = \sum_{i=1}^{M} |h'_i - b_i|, \tag{1}$$

where h'_i is the hash value of the i-th region of the received image after applying a geometric transformation, parameterized by p.

In Figure 1, the cost function in[6] is given over the 2-parameter geometric transform space for an image which is HDC-embedded, and then geometrically modified. Although the cost function reaches its global minimum at the correct modification parameters, its wild structure makes it hard for the gradient-based search algorithm to converge to the correct point. Therefore, in,[6] the entire space is divided into smaller subspaces and an individual search algorithm is executed in each of them to find the global minimum. Although this is better than exhaustive search, the total convergence time is still significantly high.

In order to speed up the search process, the cost function has to be improved in terms of smoothness, which would result in: i) fewer local minima and therefore less amount of wasted time on wrong convergence points, ii) ability to select longer gradient-based search step sizes, hence faster convergence. One possible way to obtain a smoother cost function is to increase M, so that more regions are incorporated in the synchronization process, and hence the cost function averages more numbers, thereby becoming smoother. Following this approach, some degree of smoothness is observed, however it drastically the visual quality of the marked image. To overcome this issue, we propose a novel image modification technique for HDC, based on quantization of the error between h_i and b_i (instead of forcing the error to be zero). As a result, we are able to increase M without introducing any further visual distortion and consequently the cost function becomes significantly smoother, which boosts up the convergence speed as presented in Sec. 5.

Having obtained a smoother cost function, we then design a randomized search algorithm, which does not have to search through the entire parameter space unlike in.[6] The idea is based on the observation that the cost function value at the global minimum point is always less than a certain threshold for a typical image. Hence, the algorithm is aware, whether it has converged to a local minimum or the global minimum and therefore able to terminate as soon as it reaches the correct point (global minimum).

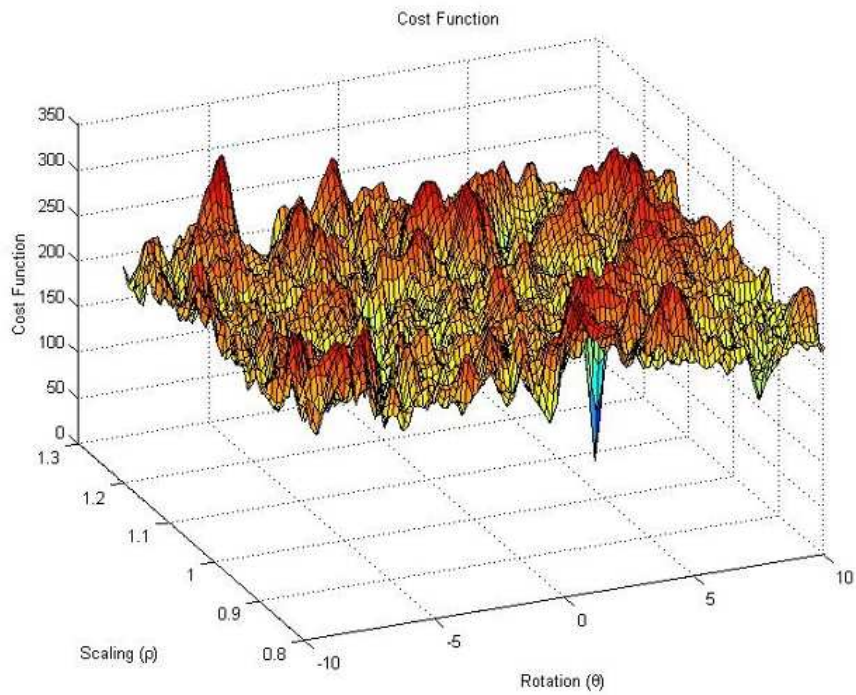

Figure 1. Cost function $C(p)$ in previously proposed algorithm as a function of rotation angle θ and scaling parameter r. The actual attack parameter values are $\theta = -5^o$, $r = 0.909$, where the global minimum of the function is achieved.

3. HASH DISTORTION COMPENSATION (HDC) DESIGN FOR COST FUNCTION IMPROVEMENT

In this section, the details of our proposed image modification scheme for HDC are discussed. The step-by-step process is explained below.

1. Using the secret key K_H, DC subband of L-level DWT of the image is tiled into N possibly overlapping rectangles, where the index set for i-th rectangle is denoted by U_i, $1 \leq i \leq N$.

2. For each U_i, the corresponding hash value, h_i, is computed via (2), where $\{a_{ij}\}$ are weights from a smoothly-varying pseudo-random field and $\{s_j\}$ are coefficients of the DWT transformed image.

$$h_i = \sum_{i=1}^{N} a_{ij} s_j \tag{2}$$

3. Using the secret key K_H, N pseudo-random numbers, $\{b_i\}_{i=1}^{N}$, are generated. Each b_i is i.i.d (independent identically distributed), where the distribution is 0-mean Gaussian, with a suitably-chosen variance* σ^2.

4. First, all $\{(h_i, b_i)\}_{i=1}^{N}$ pairs are sorted in increasing order with respect to the corresponding error values $\{e_i\}_{i=1}^{N}$, where we define $e_i \triangleq |h_i - b_i|$, $1 \leq i \leq N$. Then, top M ($M \ll N$) pairs and the corresponding regions $\left\{U_1^{'}, \ldots, U_M^{'}\right\}$ are selected. Here, without loss of generality, we assume that the chosen pairs are $\{(h_i, b_i)\}_{i=1}^{M}$.

*In practice, the variance is chosen to be approximately the same as the variance of $\{h_i\}$ in order to increase the probability of finding approximately-matching $\{(b_i, h_i)\}$ pairs.

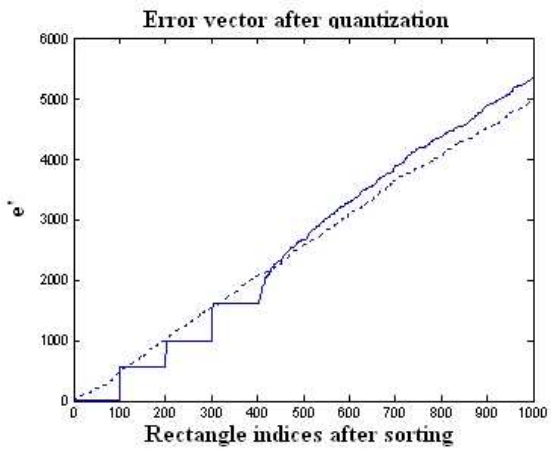

Figure 2. The sorted error values before ($\mathbf{e}$) and after ($\mathbf{e}'$) quantization are depicted in dotted and solid lines, respectively.

5. Finally, the image is modified, such that the distance between $\{b_i\}_{i=1}^{M}$ and the hash values of the received image are equal to $\{e_i'\}_{i=1}^{M}$, where e_i' is the quantized version of e_i, $1 \leq i \leq M$. Note that, such a quantization rule enables the receiver to "search" for the resulting "quantization pattern condition", thereby achieving geometric synchronization. The quantization procedure is explained in detail below.

Step 4 above ensures that we have $e_1 \leq e_2 \leq \ldots \leq e_M$. We define $\mathbf{e} \triangleq [e_1\ e_2\ \ldots\ e_M]^T$, $\mathbf{e}' \triangleq [e_1'\ e_2'\ \ldots\ e_M]^T$. Steps 1–4 are also applied in,[6] but at step 5, HDC enforces the condition $\mathbf{e}' = [0\ 0\ \ldots\ 0]^T$. However, as discussed earlier, this leads to a significant amount of visual distortion in the host image for large M. Therefore, in,[6] M is kept sufficiently small, which inevitably results in a wild cost function that hinders the efficiency of the search at the receiver.

In this work, at step 5, we propose to apply the quantization rule

$$e_i' = e_{\lfloor \frac{i}{K} \rfloor K}, \quad 1 \leq i \leq M,$$

where K is chosen such that M is an integer multiple of K. Since this results in a lower value of $\|\mathbf{e}' - \mathbf{e}\|$ than the one in,[6] we consequently achieve a smoother cost function and maintain the visual quality with reasonably high values of M. Note that, here, K is also an important algorithmic parameter, which, together with m, determines the visual quality and the smoothness of the cost function. In Fig. 2, $\mathbf{e}'$ and $\mathbf{e}$ are plotted for $M = 400$, $K = 100$. Note that, the method of[6] is a special case of this algorithm (the two are equivalent for $K = M$). Our experiments reveal that the proposed strategy allows us to choose M almost as four times as the one in[6] without introducing further distortion.

In order to obtain $\mathbf{e}'$, the original image $\mathbf{s}$ is modified by an additive perturbation $\mathbf{d}$ (HDC), that produces $\mathbf{s}' = \mathbf{s} + \mathbf{d}$, so that the hash values of $\mathbf{s}'$ are $\left\{h_i'\right\}_{i=1}^{M}$, which are given by

$$h_i' = \begin{cases} b_i + e_i' & \text{, if } h_i \geq b_i, \\ b_i - e_i' & \text{, if } h_i < b_i. \end{cases}$$

The minimum-distortion (in L_2 sense) additive perturbation, d_{min}, is given by the well-known minimum norm solution

$$\mathbf{d}_{min} \ = \ \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\left(\mathbf{h}' - \mathbf{h}\right),$$

where $\mathbf{A}$ is $M \times n$ matrix (assumed to be non-singular) containing pseudo-randomly-chosen weights corresponding to regions $\{U_i'\}_{i=1}^{M}$, $\mathbf{h}$ is the $M \times 1$ vector of hash values corresponding to the same regions, $\mathbf{h}'$ is the $M \times 1$

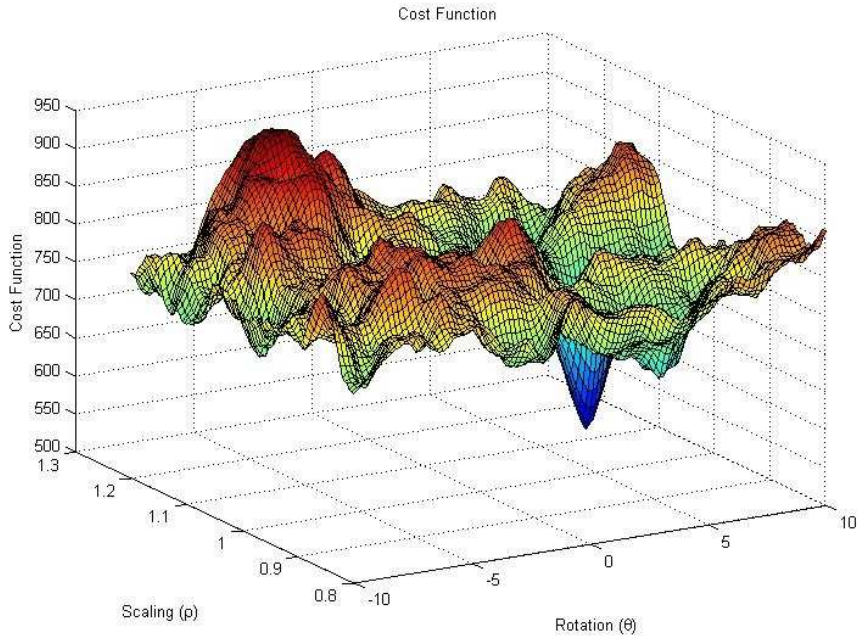

Figure 3. Proposed cost function $C(p)$ for $M = 320$ and $K = 80$ as a function of rotation angle θ and scaling parameter r. The actual attack was $\theta = -5^o$, $r = 0.909$. Note the function achieves global minimum at that point.

vector of desired hash values, and $\mathbf{s}$ is the $n \times 1$ vector representation of the image consisting of n coefficients. For further details, we refer the interested reader to[6] and.[8]

The resulting cost function at the receiver side, $C(p)$, is given in Fig. 3 over a two-parameter space of the geometric transformation model. It is observed that, the cost function is significantly improved in terms of smoothness, thereby letting a gradient-based search algorithm run more efficiently. Furthermore, our experiments reveal no signs of visual distortion more than in.[6] Specifically, the PSNR values of the HDC-embedded images are around 40 dB in both schemes.

4. GEOMETRIC SYNCHRONIZATION ALGORITHM

Geometric synchronization of the received image corresponds to finding the global minimum of the cost function over the entire parameter space. For this purpose we develop a randomized algorithm, which is basically a steepest-descent gradient-search, initiated from randomly selected initial points. In Sec. 2, it is discussed that gradient-based search runs more efficiently on smoother cost functions, which we achieve as shown in Sec. 3. Moreover, the following two (experimentally-justified) properties of our cost function can be employed in further improving the efficiency of overall synchronization algorithm: i) Cost function value at the global minimum is always less than a threshold (Th_{Global}). ii) Cost function values for the points within the convergence region are less than another threshold ($Th_{Initial}$) with high probability. Having defined the threshold values, the proposed synchronization algorithm steps are given below:

1. Pseudo-randomly choose an initial point p^0. If $C\left(p^0\right) > Th_{Initial}$, then it means we are not in the global convergence region; restart search by choosing another initial point randomly. Otherwise proceed to the next step.

2. Start a steepest-descent gradient-search initiated from the initial point p^0.

3. Let p^* be the solution (stopping point) of a gradient-based search in Step 2. If $C\left(p^*\right) > Th_{Global}$, then p^* is not the global minimum of $C(p)$; hence, go back to Step 1. Otherwise p^* is the global minimum and proceed to the next step.

(a) Original (b) HDC Embedded and Attacked (c) Synchronized

Figure 4. An example of original, attacked and synchronized image. The geometric attack parameters are $\theta = 7^o$ and $r = 0.9$

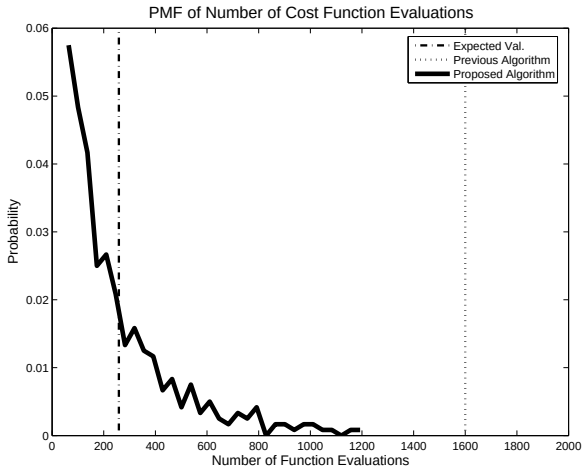

Figure 5. PMF of number of cost function evaluations. Since previous algorithm is not randomized it doesn't have a PMF. Approzimately 8-fold improvement over the previously proposed algorithm is observed in terms of expected values.

4. Apply transformation parameters in p^* to the received image to obtain the geometrically-synchronized image.

5. EXPERIMENTAL RESULTS AND DISCUSSION

In all of our experiments, we confine ourselves to global affine transformations within the class of geometric attacks. We observed that in all cases, all geometric attack parameters are estimated with sufficient accuracy, which was the case in[6] as well.

Hence, the experimental results presented in this section mainly focus on measuring the computational efficiency. During synchronization, the most costly operation is to calculate $C(p)$ since the image has to be geometrically transformed according to the parameters defined by p prior to hash computation. Since the gradient-based search algorithm evaluates $C(p)$ many times at different points, we regard the number of cost function evaluations until convergence as the "computational efficiency measure".

In order to obtain quick results, we restrict ourselves to 2-parameter modification space consisting of rotation (θ) and scaling (r) only. We fix the attack parameters to $\theta = 7^o$ and $r = 0.9$. We use grayscale test images of size 512×512. A sample image before and after geometric synchronization is given in Figure 4. For robust hash, 3-level DWT is used and the region sizes vary pseudo-randomly between 32 and 64. Weights $\{a_{ij}\}$ are chosen

as i.i.d. zero-mean, unit variance Gaussian numbers, filtered by an ideal low-pass filter with cutoff frequency 0.8π. Parameters M, K and N, that were defined in Section 3, are chosen as 320, 80 and 10000, respectively. At the receiver, prior to applying our detection algorithm, the received image is resized to 512×512 via bi-cubic interpolation. At the receiver side, we consider $r \in [0.8, 1.2]$ and $\theta \in [-10, 10]$ as the range of possible geometric attack parameters for scaling and rotation, respectively. We run the geometric synchronization algorithm 100 times, each starting from a randomly selected initial point and collect the number of cost function evaluations. We repeat this experiment for 5 different images and collect 500 values in total. The distribution of the number of function calls is given in Fig. 5. It can be observed that randomized search algorithm performs almost 8 times better than that of,[6] in the expectation sense. Note that, this performance improvement is expected to increase exponentially as the search space dimension increases.

6. CONCLUSION

In this work, we propose a randomized synchronization algorithm, which significantly increases the computational efficiency of the geometric image synchronization technique presented in.[6] The computational efficiency is crucial for real time watermarking applications and also for extending HDC-based method to higher-dimensional geometric modification models. We introduce a new image modification scheme for HDC, which results in much smoother cost functions, thereby yielding a more efficient search algorithm. As a result, we observe that the proposed method runs approximately 8 times faster than that of[6] in the expectation sense. In the future, we plan to analyze the proposed randomized algorithm and develop probabilistic bounds on its runtime.

REFERENCES

1. M. Kutter, "Watermarking resistant to translation, rotation and scaling," *Proc. SPIE Multimedia Systems and Applications* **3528**, pp. 423–431, Nov. 1998.
2. S. Pereira and T. Pun, "Fast robust template matching for affine resistant watermarking," *Int. Workshop on Information Hiding, Lecture Notes in Computer Science* **1768**, pp. 200–210, 1999.
3. J. K. O. Ruanaidh and T. Pun, "Rotation, scale and translation invariant spread spectrum image watermarking," *Signal Processing: Imag. Comm.* **66**, pp. 303–317, May 1998.
4. M. Alghoniemy and A. H. Tewfik, "Geometric distortion correction in image watermarking," *Proc. SPIE Symp. on Electronic Imaging* , pp. 82–89, Jan. 2000.
5. S. Voloshynovskiy, A. Herrigel, and Y. B. Rystar, "Watermark template attack," *Proc. SPIE Annu. Symposium on Electronic Imaging* , Jan. 2001.
6. O. Harmanci, V. Monga, and M. K. Mihcak, "Geometrically invariant image watermarking via robust perceptual hashes," *ICIP, Atlanta, GA* , pp. 8–11, Oct. 2006.
7. R. Venkatesan, S. M. Koon, M. H. Jakubowski, and P. Moulin, "Robust image hashing," pp. 664–666, Sept. 2000.
8. K. Mihcak, R. Venkatesan, and T. Liu, "Watermarking via optimization algorithms for quantizing randomized semi-global image statistics," *ACM Multimedia Systems Journal* , Apr. 2005.