IMAGE AUTHENTICATION UNDER GEOMETRIC ATTACKS VIA STRUCTURE MATCHING

Vishal Monga, Divyanshu Vats and Brian L. Evans

Embedded Signal Processing Laboratory, Center for Perceptual Systems
The University of Texas at Austin, Austin, TX 78712
{vishal, vats, bevans}@ece.utexas.edu

ABSTRACT

Surviving geometric attacks in image authentication is considered to be of great importance. This is because of the vulnerability of classical watermarking and digital signature based schemes to geometric image manipulations, particularly local geometric attacks. In this paper, we present a general framework for image content authentication using salient feature points. We first develop an iterative feature detector based on an explicit modeling of the human visual system. Then, we compare features from two images by developing a generalized *Hausdorff* distance measure. The use of such a distance measure is crucial to the robustness of the scheme, and accounts for feature detector failure or occlusion, which previously proposed methods do not address. The proposed algorithm withstands standard benchmark (e.g. Stirmark) attacks including compression, common signal processing operations, global as well as local geometric transformations, and even hard to model distortions such as print and scan. Content changing (malicious) manipulations of image data are also accurately detected.

1. INTRODUCTION

Traditionally, the methods used for media verification can be classified into two categories: digital signature-based and watermark-based. A digital signature is a set of features extracted from the media that sufficiently represents the content of the original media. Watermarking, on the other hand, is a media authentication/protection technique that embeds invisible (or inaudible) information into the media. For content authentication, the embedded watermark can be extracted and used for verification purposes. The major difference between a watermark and a digital signature is that the embedding process of the former requires the content of the media to change. However, for content authentication, both the watermark-based approach and the digital signature-based approach are expected to be sensitive to any malicious modification of the media while being able to tolerate incidental modifications such as JPEG compression, enhancement, and common signal processing operations.

An important subset of allowable distortions on an image is geometric manipulations. These can further be decomposed into two classes: global transformations such as scaling, rotations and translations, and local transformations such as random bending and shearing (e.g. the StirMark attack). One major drawback of classical watermarking [1, 2, 3, 4] as well as digital signature schemes [5, 6, 7, 8] is the lack of robustness to geometric distortions. For this reason, significant attention has been devoted in recent years towards developing geometrically invariant watermarking schemes. This includes periodic insertion of the mark [9, 10, 11], template insertion [12], mark embedding in geometrically invariant domains [13, 14], and content based watermarking schemes that extract image feature points [15, 16, 17, 18].

A common shortcoming of the methods in [9]-[15] is that they are not robust to local geometric transformations. While the methods in [16] - [18] exhibit good robustness to both global and local distortions, they implicitly make very strong assumptions of the feature point detector. In other words, feature points from the watermarked original image and a candidate image are required to exactly match (under a model of the geometric distortion) for the mark to be successfully detected. In practice, under arbitrary geometric distortions, such an assumption often proves too optimistic. Also, feature detection is seldom perfect. Feature points that are detected in the original copy may not be present in the version that has undergone a (perceptually insignificant) geometric transformation.

We present a framework for image authentication using visually significant feature points. However, unlike the aforementioned methods, our approach is signature (and not watermark) based. We extract significant image features by using a wavelet based feature detection algorithm based on the characteristics of the visual system [19]. The key component of our scheme that enables robustness to geometric transformations is the use of a generalized *Hausdorff* distance to match geometric structures. Experimental results show that such a distance more accurately captures visual changes in image content, and also compensates for occasional failure of the feature detector. Finally, we propose randomized feature extraction to enhance security against maliciously generated geometric attacks.

MATLAB code for the authentication scheme described in this paper is available at:
`www.ece.utexas.edu/~bevans/papers/2005/authentication`

Research supported by a gift from the Xerox Foundation.

0-7803-9332-5/05/$20.00 ©2005 IEEE

2. FEATURE EXTRACTION

2.1. End-Stopped Wavelets

Psychovisual studies have identified the presence of certain cells, called hypercomplex or end-stopped cells, in the primary visual cortex [19]. For real-world scenes, these cells respond strongly to extremely robust image features such as corner like stimuli and points of high curvature in general [20], [21]. Bhattacherjee *et al.* [21] constructed "end-stopped" wavelets to capture this behavior. Morlet wavelets can be used to detect linear structures having a specific orientation. In spatial domain, the two dimensional (2-D) Morlet wavelet is given by [22]

$$\psi_M(\mathbf{x}) = (e^{j\mathbf{k}_0 \cdot \mathbf{x}} - e^{-\frac{1}{2}|\mathbf{k}_0|^2})(e^{-\frac{1}{2}|\mathbf{x}|^2}) \tag{1}$$

where $\mathbf{x} = (x, y)$ represents 2-D spatial coordinates, and $\mathbf{k}_0 = (k_0, k_1)$ is the *wave-vector* of the mother wavelet, which determines scale-resolving power (SRP) and angular-resolving power (ARP) of the wavelet [22]. The frequency domain representation, $\psi_M(\mathbf{k})$, of a Morlet wavelet is

$$\hat{\psi}_M(\mathbf{k}) = (e^{-\frac{1}{2}|\mathbf{k}-k_0|^2} - e^{-\frac{1}{2}|\mathbf{k}_0|^2})(e^{-\frac{1}{2}|\mathbf{x}|^2}) \tag{2}$$

Here, $\mathbf{k}$ represents the 2-D frequency variable (u, v). In two dimensions, the end points of linear structures can be detected by applying the first-derivative of Gaussian (FDoG) filter parallel to the orientation of structures in question. The two filtering stages, the first to detect lines having a specific orientation and the second to detect the end-points of such lines, can be combined into a single filter. This results in an "end-stopped" wavelet [21]. An example of an end-stopped wavelet and its 2-D Fourier transform follow:

$$\psi_E(x, y) = \frac{1}{4} y e^{-\quad \frac{x^2+y^2}{4} + \frac{k_0}{4}(k_0 - 2jx)} \tag{3}$$

$$\hat{\psi}_E(u, v) = 2\pi \quad e^{-(\frac{(u-k_0)^2+(v)^2}{2})} \qquad jve^{-(\frac{u^2+v^2}{2})} \tag{4}$$

Equation (4) shows that $\hat{\psi}_E$ is a product of two components. The first is a Morlet wavelet oriented along the $u-$axis. The second factor is a FDoG operator applied along the frequency-axis v, that is in the direction perpendicular to the Morlet wavelet. Hence, this wavelet detects line ends and high curvature points in the vertical direction.

2.2. Proposed feature detection method

Our approach to feature detection computes a wavelet transform based on an *end-stopped* wavelet obtained by applying the FDoG operator to the Morlet wavelet:

$$\psi_E(x, y, \theta) = (FDoG)\ o(\psi_M(x, y, \theta)) \tag{5}$$

Orientation tuning is given by $\theta = \tan^{-1}(\frac{k_1}{k_0})$. Let the orientation range $[0, \pi]$ be discretized into M intervals and the scale parameter α be sampled exponentially as α^i, $i \in Z$. This results in the wavelet family

$$\psi_E(\alpha^i(x, y, \theta_k) \quad , \alpha \in \mathcal{R},\ i \in Z \tag{6}$$

where $\theta_k = (k\pi)/M$, k = 0,...,M-1. The transform is

$$W_i(x, y, \theta) = \int f(x_1, y_1) {\psi_E}^* \quad \alpha^i(x - x_1, y - y_1), \theta \quad dx_1 dy_1 \tag{7}$$

The sampling parameter α is chosen to be 2.

Fig. 1 describes the proposed feature detection method. To obtain the final set of feature points, we use an algorithm that iteratively employs the feature detector in Fig. 1, until a fixed point is reached. The iterative procedure is based on strengthening strong geometrical components of the image while eliminating weak isolated geometry. Details may be found in [23].

1. Compute the wavelet transform in $W_i(x, y, \theta)$ (7) at a suitably chosen scale i for several different orientations θ. The coarsest scale ($i = 1$) is not selected as it is too sensitive to global variations. Finer the scale, the more sensitive it is to distortions such as quantization noise. We choose $i = 3$.
2. Locations (x, y) in the image that are identified as candidate feature points satisfy
$$W_i(x, y, \theta) = \max_{(x', y') \in N_{(x,y)}} |W_i(x', y', \theta)| \tag{8}$$
where $N_{(x,y)}$ represents the local neighborhood of (x, y) within which the search is conducted.
3. From the candidate points selected in step 2, qualify a location as a final feature point if
$$\max_{\theta} W_i(x, y, \theta) > T \tag{9}$$
where T is a user-defined threshold.

Figure 1: Feature detection method that preserves significant image geometry feature points of an image.

3. PROPOSED SCHEME FOR IMAGE AUTHENTICATION

Our proposed image authentication scheme is illustrated in Fig. 2. The set of feature points $\mathbf{N}$ extracted from a candidate image (using the feature extractor described in Section 2.2) is transformed by a suitable model $\mathbf{T}$, of the geometric distortion. The transformed set of points is then compared against the (pre-computed) set of feature points $\mathbf{M}$ from a reference image using a robust distance measure $\mathbf{D}(\cdot, \cdot)$. The transformation $\mathbf{T}$ is updated using an intelligent search strategy until a local minima of the distance function is reached. Based on the value of this minimum distance, we declare the image to be credible or tampered. Next, we detail the particular choice of various components in our proposed authentication framework.

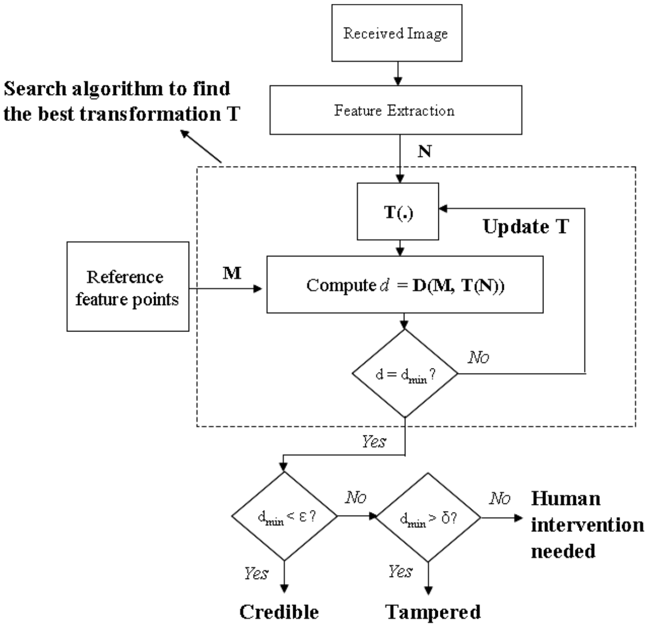


Figure 2: Flow chart of the image authentication scheme

3.1. Distortion Modeling

We model the geometric distortion on the feature points via an affine transformation $\mathbf{T}$ such that

$$\mathbf{T}(\mathbf{x}) = \mathbf{y} = \mathbf{R}\mathbf{x} + \mathbf{t} \tag{10}$$

where $\mathbf{x} = (x_1, x_2)$, $\mathbf{y} = (y_1, y_2)$, $\mathbf{R}$ is a 2×2 matrix and $\mathbf{t}$ denotes a 2×1 vector. Using an affine transform permits us to exactly model distortions such as rotation, scaling, translation, and shearing effects. Also, under a robust distance measure several other geometric distortions are well approximated via the affine transform.

3.2. Robust Distance Measure on Image Features

3.2.1. Hausdorff Distance

Given two finite point sets $\mathbf{M} = \{m_1, ..., m_p\}$ and $\mathbf{N} = \{n_1, ..., n_q\}$, the Hausdorff distance is defined as

$$H(\mathbf{M}, \mathbf{N}) = \max(h(\mathbf{M}, \mathbf{N}), h(\mathbf{N}, \mathbf{M})) \tag{11}$$

where

$$h(\mathbf{M}, \mathbf{N}) = \max_{m \in \mathbf{M}} \min_{n \in \mathbf{N}} \parallel m - n \parallel \tag{12}$$

and $\parallel \cdot \parallel$ is the underlying norm on the points of $\mathbf{M}$ and $\mathbf{N}$. The function $h(\mathbf{M}, \mathbf{N})$ is called the *directed* Hausdorff distance from $\mathbf{M}$ to $\mathbf{N}$. $h(\mathbf{M}, \mathbf{N})$ in effect ranks each point of $\mathbf{M}$ based on its distance to the nearest point of $\mathbf{N}$ and then uses the largest ranked such point as the distance. The Hausdorff distance $H(\mathbf{M}, \mathbf{N})$ is the maximum of $h(\mathbf{M}, \mathbf{N})$ and $h(\mathbf{N}, \mathbf{M})$. Thus it measures the degree of mismatch between any two shapes described by the sets $\mathbf{M}$ and $\mathbf{N}$. Our choice of Hausdorff distance is based on its relative insensitivity to perturbations in feature points, and robustness to occasional feature detector failure or occlusion [24].

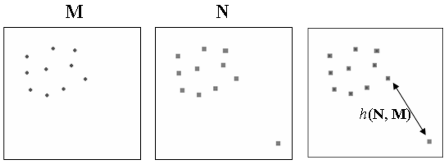


Figure 3: The directed Hausdorff distance is large just because of a single outlier

3.2.2. Modifying the Hausdorff Distance

The original Hausdorff distance in (11) is of limited utility in a robust authentication application because of its sensitivity to outliers. This is illustrated in Fig. 3. Therefore, we develop a generalized directed distance given by

$$h_g(\mathbf{M}, \mathbf{N}) = \sum_{i=1..|\mathbf{M}|} \alpha_i \min_{n \in \mathbf{N}} \parallel m_i - n \parallel, \quad where \quad \sum_i \alpha_i = 1 \tag{13}$$

The generalized Hausdorff distance $H_g(\mathbf{M}, \mathbf{N})$ is the maximum of $h_g(\mathbf{M}, \mathbf{N})$ and $h_g(\mathbf{N}, \mathbf{M})$.

Note this distance is generalized[1] because for the case that only one of the α_i's is equal to one (corresponding to $m_i \in \mathbf{M}$ that is farthest away from the closest point in $\mathbf{N}$) and rest are zero, (13) reduces to the directed Hausdorff distance in (12). Also, if each of the $\alpha_i = \frac{1}{|\mathbf{M}|}$ then this reduces to an average Hausdorff distance proposed for pattern matching by Jain *et al.* [25].

3.3. Authentication Procedure

After extracting the feature point set $\mathbf{N}$ from a received image, we find the affine transformation $\mathbf{T}^*$ that best approximates the geometric distortion. That is,

$$\mathbf{T}^* = \arg \min_{\mathbf{T}} H_g(\mathbf{M}, \mathbf{T}o\mathbf{N}) \tag{14}$$

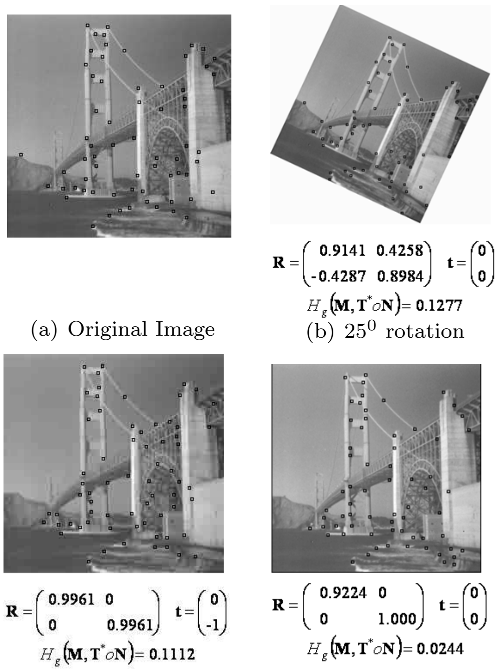
The search strategy to find $\mathbf{T}^*$ is based on a divide and conquer rule and is detailed in [26].

Finally, $H_g(\mathbf{M}, \mathbf{T}^*o\mathbf{N})$ is compared against predefined thresholds ϵ and δ $(0 < \epsilon < \delta)$ to determine the credibility of image content. Note that to be able to fix ϵ and δ, we need a normalized distance (between zero and a constant). However, there is no natural way to normalize the distance in this case. For this reason, we normalize the data sets $\mathbf{M}$ and $\mathbf{N}$, i.e. recompute their coordinates such that the mean is zero and variance is set to unity. Then, we determine empirically $\epsilon = 0.15$ and $\delta = 0.2$.

4. RESULTS

Fig. 4 (a) shows the original *bridge* image with the extracted feature points overlayed. Three modified versions of this image under both global and local geometric distortions are shown in Figs. 4 (b) though (d). From a visual inspection of Figs. 4 (a)-(d) it can be ascertained that the features largely follow the geometric transformation on the image. This validates the capability of the feature detector to successfully capture information about the geometric distortion on the

[1] The α_i's in (13) were empirically optimized.



(a) Original Image

$$\mathbf{R} = \begin{pmatrix} 0.9141 & 0.4258 \\ -0.4287 & 0.8984 \end{pmatrix} \quad \mathbf{t} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$H_g(\mathbf{M}, \mathbf{T}^* o\mathbf{N}) = 0.1277$$

(b) 25^0 rotation

$$\mathbf{R} = \begin{pmatrix} 0.9961 & 0 \\ 0 & 0.9961 \end{pmatrix} \quad \mathbf{t} = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

$$H_g(\mathbf{M}, \mathbf{T}^* o\mathbf{N}) = 0.1112$$

(c) JPEG, QF = 10

$$\mathbf{R} = \begin{pmatrix} 0.9224 & 0 \\ 0 & 1.000 \end{pmatrix} \quad \mathbf{t} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$H_g(\mathbf{M}, \mathbf{T}^* o\mathbf{N}) = 0.0244$$

(d) Stirmark local geometric attack

Figure 4: Examples of geometrically distorted images. Feature points are overlayed.

image. For each of the distorted images, we also show in Fig. 4, the estimate of the geometric transformation as determined by our authentication procedure, and the final generalized Hausdorff distance between image features under this estimated transformation. Table 1 then tabulates this distance for three different images across several different (allowable) geometric distortions. The distorted images were generated using the Stirmark benchmark software [27]. The deviation is less than 0.15 except for very large cropping (more than 25%).

We also tested under several content changing attacks including object insertion and removal, addition of excessive noise, alteration of the position of image elements, and alteration of a significant image characteristic such as texture and structure. In all cases, the detection was accurate. That is, the generalized Hausdorff distance between the features of original and attacked images was greater than 0.2. Visual as well as quantitative results for many more images, and attacks may be found at
`www.ece.utexas.edu/~bevans/projects/hashing/geometric`

5. SECURITY VIA RANDOMIZATION

We propose to enhance algorithm security by using a *randomized subspace projection* scheme. In particular, we first extract a large feature set $A = \{a_1, ..., a_Q\}$, and then (pseudo) randomly project it to a much smaller feature space spanned by the set $B = \{b_1, ..., b_P\}, P < Q$, which is finally used in image comparisons. This is accomplished via a secret key K which is used as a seed to a random number generator. This ensures that with high probability, the features that are extracted will not be the same unless the secret key is available. In practice, this significantly reduces the vulnerability to attacks by an adversary who attempts to generate malicious inputs (images) that defeat the authentication scheme.

Attack	*Lena*	*Bridge*	*Peppers*
JPEG, QF = 10	0.0857	0.1112	0.105
Scaling by 50%	0.0000	0.0020	0.1110
Rotation by 15°	0.0030	0.1277	0.0078
Random Bending	0.0345	0.0244	0.0866
Print and Scan	0.0905	0.1244	0.1091
Cropping by 10%	0.0833	0.0025	0.1117
Cropping by 25%	0.2414	0.2207	0.2766

Table 1: Generalized Hausdorff distance ($H_g(\mathbf{M}, \mathbf{T}^* o\mathbf{N})$) between features of original and distorted images.

6. REFERENCES

[1] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. on Imag. Proc.*, vol. 6, no. 12, pp. 243–246, Dec. 1996.

[2] E. T. Lin and E. J. Delp, "A review of fragile image watermarks," *Proc. ACM Multimedia and Security Workshop*, vol. 1, pp. 25–29, Oct. 1999.

[3] R. B. Wolfgang and E. J. Delp, "Fragile watermarking using the VW2D watermark," *Proc. SPIE/IS&T Int. Conf. Security and Watermarking of Multimedia Contents*, pp. 204–213, Jan. 1999.

[4] L. Xie and G. R. Arce, "A class of authentication digital watermarks for secure multimedia communication," *IEEE Trans. on Imag. Proc.*, vol. 10, pp. 1754–1764, Nov. 2001.

[5] M. Schneider and S. F. Chang, "A robust content based digital signature for image authentication," *Proc. IEEE Conf. on Image Processing*, vol. 3, pp. 227–230, Sept. 1996.

[6] C. Y. Lin and S. F. Chang, "A robust image authentication system distingushing JPEG compression from malicious manipulation," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 11, pp. 153–168, Feb. 2001.

[7] S. Bhatacherjee and M. Kutter, "Compression tolerant image authentication," *Proc. IEEE Conf. on Image Processing*, vol. 1, pp. 435–439, Sept. 1998.

[8] J. Dittman, A. Steinmetz, and R. Steinmetz, "Content based digital signature for motion picture authentication and content-fragile watermarking," *Proc. IEEE Int. Conf. on Multimedia Comput. and Sys.*, pp. 209–213, June 1999.

[9] M. Kutter, "Watermarking resistant to translation, rotation and scaling," *Proc. SPIE Multimedia Systems and Applications*, vol. 3528, pp. 423–431, Nov. 1998.

[10] T. Kalker, G. Depovere, J. Haitsma, and M. Maes, "A video watermarking system for broadcast monitoring," *Proc. SPIE Symp. on Electronic Imaging*, pp. 103–112, Jan. 1998.

[11] D. Delanay and B. Macq, "Generalized 2-d cyclic patterns for secret watermark generation," *Proc. IEEE Conf. on Image Processing*, pp. 77–80, Sept. 2000.

[12] S. Pereira and T. Pun, "Fast robust template matching for affine resistant watermarking," *Int. Workshop on Information Hiding, Lecture Notes in Computer Science*, vol. 1768, pp. 200–210, 1999.

[13] J. K. O Ruanaidh and T. Pun, "Rotation, scale and translation invariant spread spectrum image watermarking," *Signal Processing: Imag. Comm.*, vol. 66, no. 3, pp. 303–317, May 1998.

[14] C. Y. Lin, M. Wu, A. Bloom J, M. L. Miller, I. Cox, and Y. M. Lui, "Rotation, scale, and translation resilient public watermarking for images," *IEEE Trans. on Imag. Proc.*, vol. 10, pp. 767–782, May 2001.

[15] M. Alghoniemy and A. H. Tewfik, "Geometric distortion correction in image watermarking," *Proc. SPIE Symp. on Electronic Imaging*, pp. 82–89, Jan. 2000.

[16] Q. Sun, J. Wu, and R. Deng, "Recovering modified watermarked image with reference to original image," *Proc. SPIE Symp. on Electronic Imaging*, pp. 415–424, Jan. 1999.

[17] Z. Duric and N. F. Johnson, "Recovering watermarks from images," *Information and Software Engineering Technical Report*, Apr. 1999.

[18] P. Bas, J. M. Chassery, and B. Macq, "Geometrically invariant watermarking using feature points," *IEEE Trans. on Imag. Proc.*, vol. 11, pp. 1014 – 1028, Sept. 2002.

[19] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture in two nonstriate visual areas of the cat," *J. Neurophysiology*, pp. 229–289, 1965.

[20] A. Dobbins, S. W. Zucker, and M. S. Cynader, "End-stopping and curvature," *Vision Research*, pp. 1371–1387, 1989.

[21] S. Bhatacherjee and P. Vandergheynst, "End-stopped wavelets for detection low-level features," *Proc. SPIE, Wavelet Applications in Signal and Image Processing VII*, pp. 732–741, 1999.

[22] J.-P. Antoine and R. Murenzi, "Two-dimensional directional wavelets and the scale-angle representation," *Signal Processing*, pp. 259–281, Aug. 1996.

[23] V. Monga and B. L. Evans, "Robust perceptual image hashing using feature points," *Proc. IEEE Conf. on Image Processing*, vol. 1, pp. 677–680, Oct. 2004.

[24] W. J. Rucklidge, *Efficient Computation of the Minimum Hausdorff Distance for Visual Recognition*, Ph.D. thesis, Cornell University, 1995.

[25] M. P. Dubuisson and A. K. Jain, "A modified hausdorff distance for object matching," *Proc. IEEE Int. Conf. on Pattern Recognition*, pp. 566–568, Sept. 1994.

[26] W. J. Rucklidge, "Locating objects using the hausdorff distance," *IEEE Int. Conf. on Computer Vision*, 1995.

[27] "Fair evaluation procedures for watermarking systems," http://www.petitcolas.net/fabien/watermarking/stirmark, 2000.