Image Watermarking Based on Invariant Regions of Scale-Space Representation

Jin S. Seo, *Associate Member, IEEE,* and Chang D. Yoo, *Member, IEEE*

***Abstract*—This paper proposes a novel content-based image watermarking method based on invariant regions of an image. The invariant regions are self-adaptive image patches that deform with geometric transformations. Three different invariant-region detection methods based on the scale-space representation of an image were considered for watermarking. At each invariant region, the watermark is embedded after geometric normalization according to the shape of the region. By binding watermarking with invariant regions, resilience against geometric transformations can be readily obtained. Experimental results show that the proposed method is robust against various image processing steps, including geometric transformations, cropping, filtering, and JPEG compression.**

***Index Terms*—Content-based synchronization, feature point, geometric distortion, invariant region, scale space, watermarking.**

I. Introduction

With the advent of Internet, the use of digital media through electronic commerce and on-line services has grown rapidly. Since digital media is easily reproduced and manipulated, anyone is potentially capable of incurring considerable financial loss to the media producers and content providers. Digital watermarking is introduced to safeguard against such loss. While the most prominent application of watermarking is copyright protection [1], others including fingerprinting [2], broadcast monitoring [3], image authentication [4], and copy protection [5] are important research areas.

With the development of watermarking technologies, attacks against watermarking systems have become more sophisticated. In general, the attacks on watermarking systems can be categorized into noise-like signal processing and geometric distortions. While the noise-like signal processing, such as lossy compression, denoising, noise addition and lowpass filtering, reduces watermark energy, geometric distortions induce synchronization errors between the original and the embedded watermark pattern and therefore can mislead the watermark detector. Most of the previous methods have shown robustness against noise-like signal processing attacks, and only a few specialized watermarking methods have addressed the geometric distortions. These few can be classified into *nonblind scheme*, *invariant transform*, *embedding-based synchronization*, and *content-based synchronization*.

- *Nonblind scheme*: Nonblind schemes use the original image to synchronize watermark. In [6] and [7], the meshes of the original image and the watermarked image are compared to recover synchronization errors. Nonblind scheme is effective for local distortions, but an extension to global affine transformations is computationally demanding [8].
- *Invariant transform*: The most obvious way to achieve resilience against geometric distortions is to use an invariant transform. In [9]–[11], the watermark is embedded in an affine-invariant domain such as the Fourier–Mellin transform. Despite its robustness against affine transformations, those techniques involving invariant domain suffer from implementation issues and are vulnerable to cropping [10].
- *Embedding-based synchronization*: Another way to cope with geometric distortions is to use template [12] or periodic insertion [13] of the watermark pattern. In [12], the template is embedded in the discrete Fourier transform (DFT) domain as local peaks in predefined positions. The embedded local peaks are searched during watermark detection process in order to yield information about the affine transformations that the image has undergone [14]. In [13], it is shown that the periodic insertion of the watermark pattern can give synchronization information. The periodically embedded watermark pattern appears as a lattice of peaks in the autocorrelation domain, and careful analysis of both the orientation of the lattice and the distance between the peaks can reveal vital information about the affine transformations. However, this kind of approach can be tampered by the malicious attack stated in [15] since anyone can access the peaks in the DFT or the autocorrelation domain and easily eliminate them.
- *Content-based synchronization*: By binding the watermark synchronization with the image characteristics, watermark detection can be done without synchronization error. In [16]–[19], moment-based normalization approaches are proposed. In spite of the robustness against

Manuscript received June 28, 2004; revised May 23, 2005. This work was supported in part by the Ministry of information & Communications, Korea, under the Information Technology Research Center (ITRC) Support Program, the Grant no. R01-2003-000-10829-0 from the Basic Research Program of the Korea Science & Engineering Foundation, and the Brain Korea 21 Project, the School of Information Technology, KAIST, in 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ton Kalker.

J. S. Seo was with the Division of Electrical Engineering, Department of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Korea. He is now with the Digital Contents Research Division, Electronics and Telecommunications Research Institute, Daejeon 305-700, Korea (e-mail: jsseo@kaist.ac.kr).

C. D. Yoo is with the Division of Electrical Engineering, Department of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, Korea (e-mail: cdyoo@ee.kaist.ac.kr).

Digital Object Identifier 10.1109/TSP.2006.870581

affine transformations, they are highly vulnerable against cropping attacks. In [20], a region-based watermarking method is proposed by segmenting an image into a number of regions. The two largest regions are watermarked with the affine normalized watermark. In [8], [21], and [22], the feature points of an image are used as a reference for retaining synchronization of the watermark. The previous feature-point-based watermarking methods achieved robustness against various geometric attacks using a geometric normalization method in the form of tessellation [8], self-spanning pattern [21], and moments [22]. The robustness of the resulting watermark strongly depends on the repeatability of the geometric normalization method. The weak point of the content-based approaches is the added computational complexity in calculating the features of an image before watermark detection.

This paper presents a novel content-based watermarking method based on scale-space feature points. Although there are approaches that deal with the geometric distortions, the problems with random bending attack [23], geometric distortion with cropping, nonisotropic scaling, and projective transformation still remain difficult. By binding the watermark synchronization with *invariant regions* of an image, watermark detection can be performed without synchronization error. More specifically, the *invariant regions* attained from the feature points of the scale-space representation of an image are used for referencing the watermark. Since a feature point falls short of providing sufficient information about the geometric transformations that the image has undergone, additional information about the geometric transformations has to be somehow obtained. We considered three different methods for obtaining the additional information: *characteristic scale* [24]–[26], *shape adaptation* of the Gaussian kernel [27], [28], and *feature-point sets* [29]–[31]. Feature points with the additional information lead to geometrically invariant regions. The invariant-region detection methods were originally proposed as an effort to find correspondences between images in which there are large changes in viewpoint, scale, and illumination. The invariant regions are self-adaptive image patches that deform with geometric transformations as to keep on covering identical physical parts of a scene [32]. There have been approaches that use feature points of an image as a reference for watermarking [8], [21], [22]; however, these approaches, which do not explicitly consider the variation of feature points regarding to the scaling of an image, have limited robustness against scaling. Recent literature has shown that this problem can be mitigated by using the scale-space feature points [25], [26], [33]. The details of scale-space feature points and invariant-region detection methods are described in Section II. By synchronizing watermarking with invariant regions, the proposed watermarking method achieves robustness against both noise-like and geometric attacks.

This paper is organized as follows. Section II describes the scale-space feature points and the three invariant-region detection methods. Section III describes the proposed watermarking method. Section IV evaluates the performance of the proposed method. Section V summarizes the performance and the limitations of the proposed method.

II. Invariant-Region Detection Based on Scale-Space Feature Points

To develop a content-based method, image characteristics appropriate for watermarking should be carefully selected. Feature points can be a good candidate for content-based watermarking since there exist feature point detection methods that are proven to be robust against many image processing steps such as sharpening, blurring, compression, and geometric transformations [34], [26], [25], [35]. However, a feature point provides only the position information. To cope with affine or projective transformations, we need additional information about the geometric transformations, which can be acquired from the neighborhood of the feature point. Three different methods are considered for that purpose, such as *characteristic scale*, *shape adaptation* of the Gaussian kernel, and *feature-point sets*. In [24]–[26], the characteristic scale of a feature point is used for scale-invariant pattern matching. The characteristic scale is the scale at which the normalized scale-space representation of an image attains a maximum value. In [27], [28], and [36], the *second-moment matrix* around a feature point is used to form an affine-invariant region. It is based on the *shape adaptation* [27] procedure that adapts the Gaussian smoothing kernel to the local image structures around the feature point. The shape adaptation process is iteratively performed until the shape of the kernel converges. In [29], [30], and [37], feature-point sets obtained by grouping feature points are used for invariant pattern matching since the relative positions of the feature points are geometrically invariant [31]. The details of the scale-space feature points and the three invariant-region detection methods are described in the next subsections.

A. Detection of Harris Points Through Scale Space

The detection of feature points, which are invariant to geometric transformations, has been one of the main issues in pattern recognition and computer vision. A feature point is the point where the characteristics of an image change in multiple directions, such as corners, high curvature points and junctions. Haralick and Shapiro [38] consider a point in an image interesting if it has two main properties: distinctiveness and invariance. This means that a point should be distinguishable from its immediate neighbors, and the position as well as the selection of the interesting point should be invariant with respect to the expected geometric distortions [35]. Among various feature point detectors, the Harris detector [39] showed the best performance in the evaluation test [34], [8] in terms of repeatability. The Harris point is geometrically stable under various image processing steps that include rotation, noise addition and illumination change. However, it is unstable under scaling with a large ratio since the window size used in extracting Harris points is fixed. The image structures in a fixed-size window can change substantially after scaling. Without prior knowledge of the image structures, there is no reason to favor any particular window size [40]. A main argument behind the scale-space representation is that if no prior information is available about what is the appropriate window size (scale) for a given data, then the only reasonable approach is to represent the input data using various window sizes (at multiple scales) [24]. The scale-space representation

is a set of images represented at different levels of resolutions. Given a scale s, the *uniform Gaussian scale-space representation* L: $\mathbb{R}^2 \times \mathbb{R}_+ \to \mathbb{R}$ of an image $f : \mathbb{R}^2 \to \mathbb{R}$ is defined by

$$L(\mathbf{x}; s) = g(\mathbf{x}; s) * f(\mathbf{x}) \tag{1}$$

where $\mathbf{x} = (x, y)$ refers to the image spatial coordinate, $g(\mathbf{x}; s)$ is the associated uniform Gaussian kernel with standard deviation s and mean zero, and $*$ denotes linear convolution. It was proven in [41], [42] that the Gaussian kernel is the unique kernel for generating scale-space representation under some postulated conditions. This mathematical result is also in accordance with the biological experimental results in [43] stating that the measured response at the optic nerve of the human eye is similar to the derivative of the Gaussian [24].

The Harris detector is based on a specific image descriptor called the second-moment matrix, which reflects the local distribution of gradient directions in the image [27]. The scale-normalized second-moment matrix $\mu(\mathbf{x}, s)$ is given by

$$\mu(\mathbf{x}, s) = s^2 g(\mathbf{x}; s) * \begin{bmatrix} L_x^2(\mathbf{x}; t) & L_x L_y(\mathbf{x}; t) \\ L_x L_y(\mathbf{x}; t) & L_y^2(\mathbf{x}; t) \end{bmatrix} \tag{2}$$

where $L_x = (\partial/\partial_x)L(\mathbf{x}; s)$ and t is set to $(s/2)$ in this paper. This matrix represents the statistics of gradient directions in the neighborhood of the point $\mathbf{x}$. Harris and Stephens [39] claim that the point $\mathbf{x}$ is a feature (corner) point if the matrix $\mu(\mathbf{x}, s)$ has two significant eigenvalues. To avoid explicit eigenvalue decomposition of $\mu(\mathbf{x}, s)$, the trace and the determinant of $\mu(\mathbf{x}, s)$, which are the summation and the multiplication of the two eigenvalues, respectively, are used in the formulation. Then the scale-normalized Harris corner strength (SHCS) measure $U(\mathbf{x}, s)$ at scale s and point $\mathbf{x}$ is given in terms of $\mu(\mathbf{x}, s)$ as follows:

$$U(\mathbf{x}, s) = \det \mu(\mathbf{x}, s) - 0.04 * (\operatorname{trace} \mu(\mathbf{x}, s))^2. \tag{3}$$

At each level of the scale space, Harris points are detected as the local maxima in the image plane as follows:

$$\begin{aligned} &U(\mathbf{x}, s) > U(\mathbf{x_a}, s) \quad \forall \mathbf{x_a} \in A \\ &U(\mathbf{x}, s) > t_u \end{aligned} \tag{4}$$

where A and t_u denote the neighborhood of the point $\mathbf{x}$ and the detection threshold respectively. Corners are placed at local maxima of the corner strength measure. The corner strength measure then can be used to order the corners in order of significance. In general, the number of the Harris points decreases with increasing scale s [25].

B. Automatic Scale Selection and Scale-Invariant Feature Point

Although the scale-space representation presented in the previous subsection provides a well-founded framework for representing and detecting image structures at multiple scales, it does not address the problem of how to select locally appropriate scales for further analysis. Lindeberg proposed a general method for feature detection with automatic scale selection in [24]. The basic idea is to apply the feature detector at all scales and then select scale levels at which normalized measures of feature strength assume local maxima [40].

A well-known property of the scale-space representation is that the amplitude of spatial derivatives in general *decreases with scale* [24], i.e., if a signal is subjected to scale-space smoothing, then the numerical values of spatial derivatives computed from the smoothed data can be expected to decrease. Thus, to select scale that reflects local characteristics of an image, a scale-normalized derivative ∂^N is introduced [24] as follows:

$$\partial^N_{x^{\alpha_1}, y^{\alpha_2}} = s^{\alpha_1+\alpha_2} \partial_{x^{\alpha_1}} \partial_{y^{\alpha_2}} \tag{5}$$

where α_1 and α_2 are the order of differentiation. To give a formal characterization of the characteristic scale, consider two images f and $\tilde{f}$ related by $f(\mathbf{x}) = \tilde{f}(t\mathbf{x})$. The scale-space representations L and $\tilde{L}$ of f and $\tilde{f}$ are defined by (1), respectively. Then the scale-space representations are related by

$$L(\mathbf{x}; s) = \tilde{L}(\tilde{\mathbf{x}}; \tilde{s}) \tag{6}$$

where $\tilde{\mathbf{x}} = t\mathbf{x}$ and $\tilde{s} = ts$. Differentiation of (6) by x and y with order α_1 and α_2 gives the following relationship:

$$\partial_{x^{\alpha_1}} \partial_{y^{\alpha_2}} L(\mathbf{x}; s) = t^{\alpha_1+\alpha_2} \partial_{\tilde{x}^{\alpha_1}} \partial_{\tilde{y}^{\alpha_2}} \tilde{L}(\tilde{\mathbf{x}}; \tilde{s}) \tag{7}$$

where $\partial_{x^{\alpha_1}} \partial_{y^{\alpha_2}} L(\mathbf{x}; s) = (\partial^{\alpha_1+\alpha_2}/\partial_{x^{\alpha_1}} \partial_{y^{\alpha_2}}) L(\mathbf{x}; s)$. By multiplying $s^{\alpha_1+\alpha_2}$ to (7), we can obtain the *scale invariance* from (5) as follows:

$$\partial^N_{x^{\alpha_1}, y^{\alpha_2}} L(\mathbf{x}; s) = \partial^N_{\tilde{x}^{\alpha_1}, \tilde{y}^{\alpha_2}} \tilde{L}(\tilde{\mathbf{x}}; \tilde{s}). \tag{8}$$

The scale level, at which a combination of normalized derivatives attains a local maximum over scales, can be treated as reflecting a characteristic length of corresponding structure. Thus, the normalized scale-space maxima have been considered as *characteristic scale* of the image. The scale-invariance property of normalized derivatives ensures the invariance of scale-space maxima [24]. If a normalized scale-space maximum is at $(\mathbf{x_0}; s_0)$ in the scale-space representation of an image f, then the corresponding scale-space maximum is assumed at $(t\mathbf{x_0}; ts_0)$ in the scale-space representation of $\tilde{f}$. Fig. 1 shows an example of scale selection. For the same point in the original and the cropped and scaled image (scale factor is 2.0), we compute the amplitude of normalized derivatives (Laplacian) [44] over scales. The figure shows that the characteristic scale is relatively invariant to scaling. The ratio of the scales at corresponding points in the two images, at which the maxima were found, is equal to the scale factor between the two images. In the comparative tests [25], [34], the Harris points proved to be the most reliable under various image processing steps, and the Laplacian was determined to have the best repeatability for scale selection under large scale changes. Thus, the scale selected by the Laplacian at each Harris point is used as the characteristic scale. The scale-invariant feature points selected by the Harris measure and the Laplacian exhibit invariance to scaling (preserving aspect ratio), rotation and translation as well as robustness to illumination changes [25].

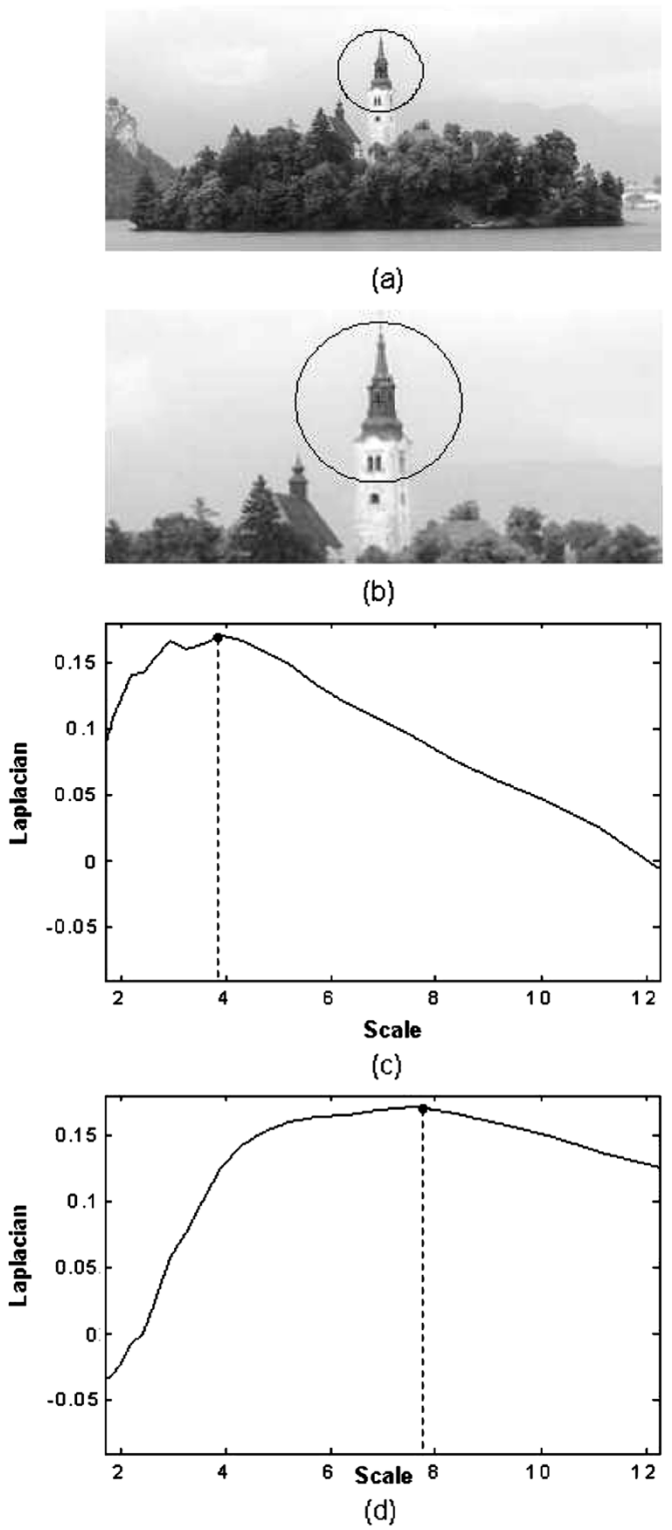

Fig. 1. (a) Original image. (b) 50% cropped and 200% scaled image. (c) Scale-normalized Laplacian of the image (a) around the feature point at the center of circle. (d) Scale-normalized Laplacian of the image (b) around the feature point at the center of circle. The radius of the circle is $10s$, where s is the characteristic scale at that point.

C. Shape Adaptation of Scale-Space Operator

Scale selection based on the uniform Gaussian scale space in (1) cannot provide invariance when there is nonisotropic scaling and skew. For such cases, the window shape used in the feature detection should adapt differently along different directions to keep on covering the identical region of an image. When dealing with the linear transformations of the spatial domain, a natural generalization of the uniform Gaussian scale-space representation in (1) is the *affine Gaussian scale-space representation* [27] generated by convolution with nonuniform Gaussian kernel:

$$g(\mathbf{x};\Sigma)=\frac{1}{2\pi\sqrt{\det\Sigma}}\exp\left(\frac{-\mathbf{x}^T\Sigma^{-1}\mathbf{x}}{2}\right) \tag{9}$$

where Σ is a symmetric positive-definite 2×2 matrix. Then the *affine Gaussian scale-space representation* is defined by

$$L(\mathbf{x};\Sigma)=g(\mathbf{x};\Sigma)*f(\mathbf{x}). \tag{10}$$

The second-moment matrix $\mu(\mathbf{x};\Sigma_s,\Sigma_t)$ of the *affine Gaussian scale space* is given by

$$\mu(\mathbf{x};\Sigma_s,\Sigma_t)=g(\mathbf{x};\Sigma_s)*((\nabla L)(\mathbf{x};\Sigma_t)(\nabla L)(\mathbf{x};\Sigma_t)^T) \tag{11}$$

where Σ_s and Σ_t are the covariance matrices corresponding to the integration and the local scale respectively. The computation of the second-moment matrix involves the integration of the local image statistics (corresponding to Σ_t) over a finite-sized image region (corresponding to Σ_s) [27]. By comparing (10) with (1), we can easily realize that the scale selection in the uniform Gaussian kernel in (1) is corresponding to the selection of the covariance matrix Σ in the nonuniform Gaussian kernel in (10). The selection of the covariance matrix Σ is usually performed iteratively [27], [28] until the second-moment matrix computed at a certain point is proportional to the Σ of the Gaussian kernel. The iteration process is called the *shape adaptation of the Gaussian kernel*.

To give a formal characterization of the shape adaptation of the Gaussian kernel and its invariance against linear transformations, consider two images f and $\tilde{f}$ related by $f(\mathbf{x})=\tilde{f}(B\mathbf{x})$. In [27], Lindeberg showed that the second-moment matrices μ and $\tilde{\mu}$ of f and $\tilde{f}$, respectively, satisfy the following *transformation property*:

$$\mu(\mathbf{x};\Sigma_s,\Sigma_t)=B^T\tilde{\mu}(B\mathbf{x};B\Sigma_sB^T,B\Sigma_tB^T)B \tag{12}$$

where B^T refers to the transpose of the matrix B. From the transformation property, it is shown that if the second-moment matrix at a point $\mathbf{x}_L$ of the image f satisfies the following *fixed-point* condition:

$$\mu(\mathbf{x}_L;\Sigma_{s,L},\Sigma_{t,L})=M_L,\quad \Sigma_{s,L}=sM_L^{-1},\quad \Sigma_{t,L}=tM_L^{-1} \tag{13}$$

then the second-moment matrix at the point $\tilde{\mathbf{x}}_L(=B\mathbf{x}_L)$ of the image $\tilde{f}$ also satisfies the following conditions:

$$\tilde{\mu}(\tilde{\mathbf{x}}_L;\tilde{\Sigma}_{s,L},\tilde{\Sigma}_{t,L})=\tilde{M}_L,\quad \tilde{\Sigma}_{s,L}=s\tilde{M}_L^{-1},\quad \tilde{\Sigma}_{t,L}=t\tilde{M}_L^{-1} \tag{14}$$

where $s,t\in\mathbb{R}_+$. A fixed-point condition is a natural indicator of successful shape adaptation. More generally, a fixed-point condition reflects the fact that the shape of the smoothing kernel agrees with the shape of the image structure [27]. Since the fixed

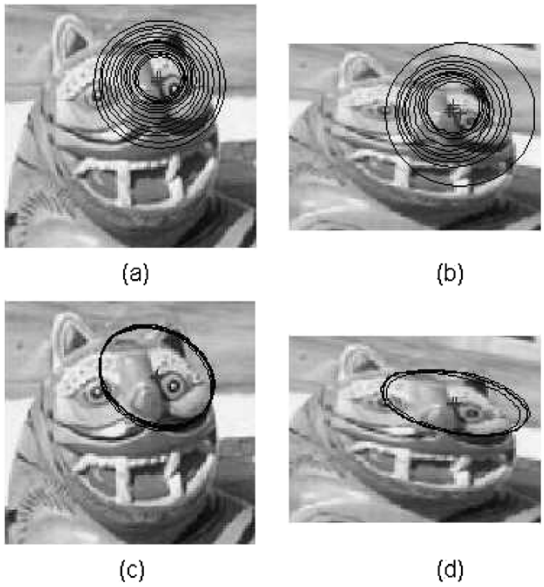

Fig. 2. (a), (b) Scale-space Harris points detected from the original and the nonisotropically scaled and rotated image respectively. (c), (d) Shape of the converged Gaussian kernels from the original and the nonisotropically scaled and rotated image, respectively. The radius of the circles in (a) and (b) and the major and minor axes of the ellipses in (c) and (d) are five times that of the original value.

point is preserved under linear transformations as shown above, regions invariant against the linear transformations can be derived from the converged second-moment matrix in (13) and (14). Baumberg [36] transforms an image to a normalized domain using the square root of the second-moment matrix. The normalized images f' and $\tilde{f}'$ of f and $\tilde{f}$, respectively, are given by

$$f'\left(M_L^{-\frac{1}{2}}\mathbf{x}\right) = f(\mathbf{x}), \quad \tilde{f}'\left(\tilde{M}_L^{-\frac{1}{2}}\hat{\mathbf{x}}\right) = \tilde{f}(\hat{\mathbf{x}}) \qquad (15)$$

where $M_L^{-(1/2)}$ is the square root matrix of M_L. From the transformation property in (12), the second-moment matrices μ' and $\tilde{\mu}'$ of f' and $\tilde{f}'$, respectively, are given by

$$\mu'\left(M_L^{-\frac{1}{2}}\mathbf{x}_L; sI, tI\right) = \tilde{\mu}'\left(\tilde{M}_L^{-\frac{1}{2}}\hat{\mathbf{x}}_L; sI, tI\right) = I \qquad (16)$$

where I is the 2×2 identity matrix. In this normalized domain, the second-moment matrix calculated using the uniform Gaussian smoothing kernel is the identity for both images [36]. Thus, the shape adaptation process is iteratively performed until the second-moment matrix of the normalized frame converges to the identity. The details of the iteration can be found in [28]. Fig. 2 shows an example of shape adaptation. For both the original and the affine-transformed image, the initial scale-space Harris points converge to the ellipses (the shape of the Gaussian kernel) that cover nearly identical regions.

D. Feature-Point Sets

Scale-space feature points have been used for image matching and recognition [25], [26], [32] thanks to the high repeatability [45], [34] against various image processing steps. Assuming that the feature points are highly repeatable, grouping the feature points into *feature-point sets* can provide geometrically invariant regions. The relative positions of all the pixels in the image to the points in the feature-point set are invariant against geometric transformations. This means that the points in the feature-point set are sufficient in defining a *basis* when an image is allowed to undergo geometric transformations [31]. This concept has been utilized in extracting geometric invariants of an image [30], [46], [37], [29]. A detailed survey on the geometric invariance from point sets can be found in [47], [31]. By using a feature-point set, the following invariance can be obtained for 2-D images [30], [29]:

- *one-point basis*: translation;
- *two-point basis*: similarity transformation (translation, rotation, and aspect-ratio preserving scaling);
- *three-point basis*: affine transformation;
- *four-point basis*: projective transformation.

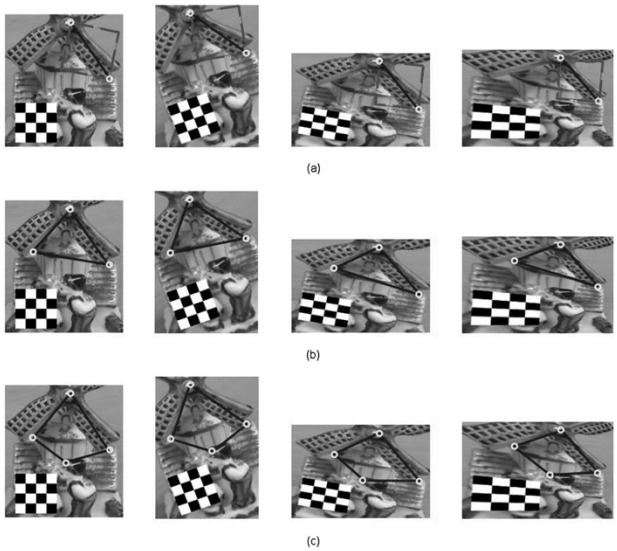

Fig. 3. (a), (b), (c) Invariant regions from two-point set, three-point set and four-point set, respectively; the original image (the first column) is subjected to rotation (the second column), nonisotropic scaling and rotation (the third column), and projective transformation (the fourth column).

Fig. 3 shows the invariant regions obtained from the feature-point sets consisting of two, three and four Harris points of an image, respectively. Some selected geometric transformations are applied to an image to show the invariance of the feature-point sets. The distortion to the checkerboard helps to visualize the applied geometric transformations. After the similarity transformation, the squares in the checkerboard remain as square. After the affine transformation, the parallelism of the checkerboard is preserved. After the projective transformation, the parallel lines in the checkerboard are changed to converging lines. In the case of a two-point set, a square with the two points at the opposite corners is used as an invariant region. In the case of a three-point set, a triangle with the three points as its corners is used as an invariant region. In the case of a four-point set, a quadrangle with the four points as its corners is used as an invariant region (we only consider convex quadrangles). The scale-space Harris points of an image are used for constructing feature-point sets. Since the scale-space Harris points are highly

TABLE I
SHAPE OF THE WATERMARK AND THE GEOMETRIC INVARIANCE [31]

Shape of watermark (invariant region)	Degree of freedom	Geometric invariance	Homography
Circle (+char. orientation)	3+1	Similarity	$\begin{bmatrix} sr_{11} & sr_{12} & t_x \\ sr_{21} & sr_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$
Ellipse (+char. orientation)	5+1	Affine	$\begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$
Square (two-point set)	4	Similarity	$\begin{bmatrix} sr_{11} & sr_{12} & t_x \\ sr_{21} & sr_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$
Triangle (three-point set)	6	Affine	$\begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix}$
Quadrangle (four-point set)	8	Projective	$\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$

repeatable for most image processing steps, the invariant regions obtained from the feature-point sets of the scale-space Harris points are also highly repeatable. At each scale s, the Harris points are grouped into feature-point sets whose cardinality is Q. We considered two, three, and four cases ($Q = 2, 3, 4$, respectively). Since grouping the Harris points of an image into all possible feature-point sets is computationally demanding, the Harris points with the locally largest SHCS measure in (3) are only considered based on the assumption that the Harris point with larger SHCS measure is more likely to be repeatable. In other words, if the Harris point has the largest SHCS measure at its neighborhood (disk of radius $5s$), it survives; otherwise, it is erased. From the neighborhood of each surviving Harris point (disk of radius $30s$), we select up to P Harris points (typically $P = 12$) with the P largest SHCS measures. By selecting $Q-1$ points from the P points, we can construct feature-point sets with cardinality Q (including the one in the center of the disk). Then the number of possible invariant regions for each surviving Harris point is given as $\binom{P}{Q-1}$.

III. PROPOSED WATERMARKING SCHEME

One of the major difficulties for watermarking is that watermark embedding and detection should be performed over the same regions of an image. In this respect, the watermark synchronization problem can be regarded as finding correspondences between the original and the processed image in which there are various image processing steps, including geometric transformations, filtering and compression. Finding correspondences between images in which there are large changes in scale, viewpoint, and illumination has been a difficult problem. In an attempt to solve the problem, local invariant matching has been introduced. The idea is to extract invariant regions of an image, which are then used for feature matching [48], [26], [29]. In this paper, we apply the same idea to watermark synchronization. The watermark is shaped adaptively based on the invariant regions of an image. We considered three invariant-region detection methods for watermarking. By using the local invariant regions of an image, the proposed method achieves resilience against both removal and geometric attacks.

A. Watermark Embedding

The Harris points appropriate for watermarking as stated in Section II-A are extracted from the scale-space representation of an image. Invariant regions are generated at each scale based on *characteristic scale*, *shape adaptation* of the Gaussian kernel, and *feature-point sets*. As shown in Table I, the invariant regions represent circle, ellipse, triangle, and quadrangle respectively. In (1), the scale refers to the standard deviation of the associated Gaussian kernel. The scale range, which was considered in this paper, is between 2.2 and 12.5. In general, the size of the watermark is greater than the scale values. To compensate for this mismatch, the radius of a circle is set to be eight times the characteristic scale of the Harris point, and the major and minor axes of an ellipse are set to be 11 times that of the original value. To select nonoverlapping invariant regions for watermark embedding out of many invariant regions available throughout the scale space, it is assumed that the Harris point with larger SHCS measure is more likely to be repeatable. In (3), the determinant and the trace of the scale-normalized second-moment matrix $\mu(\mathbf{x}, s)$ [24], [36] are used in calculating $U(\mathbf{x}, s)$. This means that corner strengths can be compared across different scales using the value of the SHCS measure $U(\mathbf{x}, s)$ [36]. The *repeatability* of invariant regions depends on the *repeatability* of the Harris points. For an invariant region from either the *characteristic scale* or the *shape adaptation* of

the Gaussian kernel, the *repeatability measure* is assumed to be the SHCS measure of the Harris point. For an invariant region from the *feature-point set*, the *repeatability measure* is assumed to be the minimum of the SHCS measures of the Harris points in the feature-point set. Since the watermarked image can be scaled up or down, we first consider the invariant regions in the middle-scale band $(4.5 \leq s \leq 8)$. When a number of invariant regions are overlapping, only the one with the largest repeatability measure survives. By repeating the same process for all the invariant regions in the middle-scale band, we can select invariant regions that have the largest repeatability measure in their neighborhood. While retaining the selected invariant regions in the middle-scale band, the same process is performed in the high-scale $(8 < s < 12.5)$ and the low-scale $(2.2 < s < 4.5)$ band, respectively. In other words, when the invariant regions in the low-scale and the high-scale band are overlapping with the selected invariant regions in the middle-scale band, they are not considered. Otherwise, the same selection process used in the middle-scale band is applied to both the low-scale and the high-scale band. Fig. 4 shows the surviving invariant regions for Lena image. The watermark is embedded additively inside the surviving invariant regions in Fig. 4.

The circular and elliptical regions lack the invariance against rotation [28], [36]. The characteristic orientation of these regions can be decided by the *principal axes* moments [49]. The watermark should be synchronized by the characteristic orientation to achieve resilience against rotation. In case of an elliptical region, the characteristic orientation is calculated after transforming the region to a circular region of a fixed size. After rotating a circular region by an angle θ, ρ_{pq}, the p, qth moment of the circular region, is changed to ρ'_{pq} as follows:

$$\rho'_{pq} = \sum_{r=0}^{p}\sum_{s=0}^{q}\binom{p}{r}\binom{q}{s}(\cos\theta)^{p-r+s} \cdot (\sin\theta)^{q+r-s}\rho_{p+q-r-s,r+s}. \quad (17)$$

The principal axis is obtained by rotating the axis of the moments until ρ'_{11} is zero. Then the characteristic orientation θ_c, measured from the original axis, is defined by

$$\tan 2\theta_c = \frac{2\rho_{11}}{\rho_{20} - \rho_{02}}. \quad (18)$$

More details of the principal axes moments can be found in [49]. Table I shows the geometric invariance related to the shape of the watermark in terms of homography[1] matrix of the geometric transformations [31]. If we want to detect the watermark after geometric transformation, the degree of freedom of the watermark shape should correspond to that of the geometric transformation.

The watermark is embedded in the selected invariant regions after geometric normalization according to the shape of the regions. A binary zero-mean pseudorandom sequence $C = (c_1, c_2, \ldots, c_M)$ is generated by a secure key. The construction of a circular and an elliptical watermark is shown in Fig. 5. The watermark sequence C is spread over the first quadrant. The quadrant is rotated and copied into the other quadrants, and then the circular prototype watermark pattern W is obtained. From the circular watermark, the elliptical watermark can be obtained by the affine transformation according to the shape of the ellipse. The construction of a triangular and a quadrangular watermark is shown in Fig. 5. We divide the invariant regions into three and four equi-areal regions for the triangle and the quadrangle respectively. The two-dimensional (2-D) square prototype watermark W ($\sqrt{M}$ by $\sqrt{M}$ pattern) from the watermark sequence C is transformed to each invariant region as shown in Fig. 5. By embedding the same watermark redundantly at each vertex of the triangle and the quadrangle, the watermark detector need not consider the orientation of them. However, we note that the simple redundant embedding does not guarantee robustness against mirroring of an image, for which the vertices need to be ordered by either the magnitude of the angles [8] or the moment normalization [49]. In the spatial domain, the watermark is embedded additively

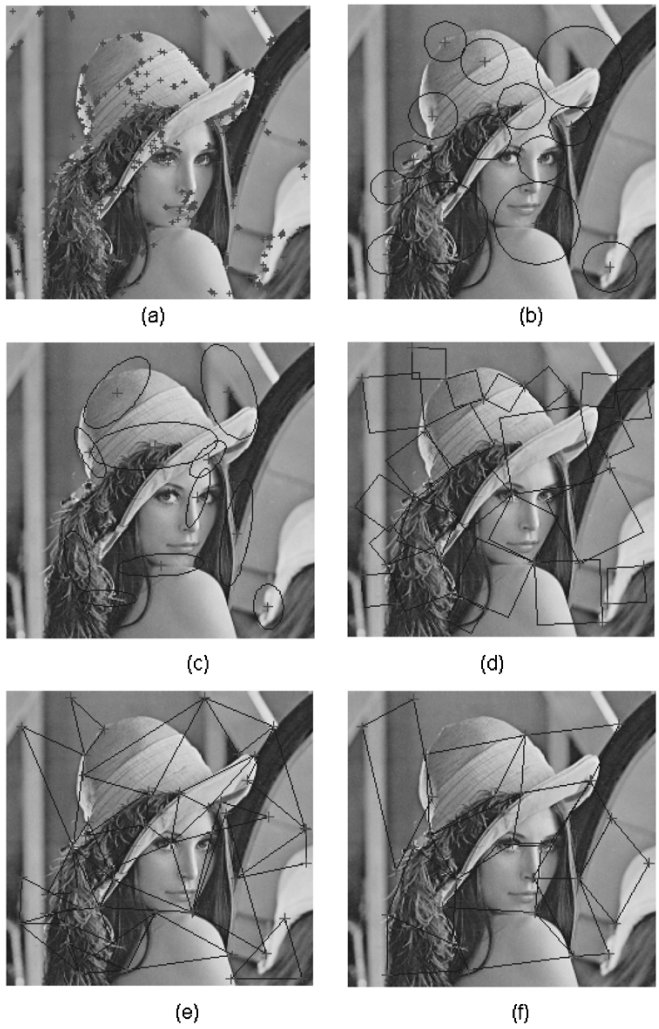

Fig. 4. (a) Scale-space Harris points. Finally selected invariant regions for watermarking using (b) characteristic scale, (c) shape adaptation of the Gaussian kernel, (d) feature-point set with $Q = 2$, (e) feature-point set with $Q = 3$, (f) feature-point set with $Q = 4$. The radius of the circles in (b) is eight times the characteristic scale of the Harris point. The major and minor axes of the ellipses in (c) are 11 times that of the original value.

[1]Homography refers to an invertible point-to-point mapping that maps lines to lines and is synonymous with the projective transformation [31].

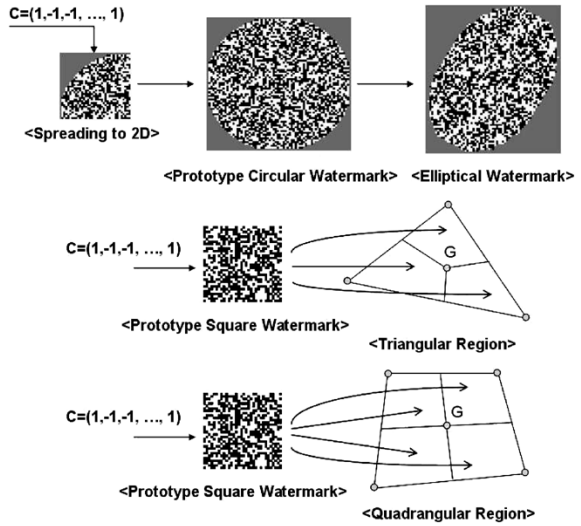

Fig. 5. Construction of the prototype watermark W from one-dimensional binary random sequence C and embedding W to invariant regions; white, gray and black refer to 1, 0 and -1 respectively. G refers to the center of mass of the triangle and the quadrangle.

after geometric normalization according to the invariant region. Mathematically it can be written as follows:

$$f'(\mathbf{x}) = f(\mathbf{x}) + \alpha(\mathbf{x})W(\mathbf{H}^{-1}\mathbf{x}^T) \tag{19}$$

where $\alpha(\mathbf{x})$ is the local masking function calculated from the Human Visual System (HVS) [50] to make the embedded watermark imperceptible, and $\mathbf{H}$ is the mapping matrix [31] representing the homography between the prototype watermark W and the invariant region.

The watermark embedding procedure is summarized as follows:

1) construct invariant regions at each scale from the scale-space Harris points of an image;
2) select invariant regions for watermarking, considering the spatial positions and the repeatability measures;
3) embed the watermark additively into the selected regions after geometric normalization according to the shape of the regions as shown in Fig. 5.

B. Watermark Detection

It is assumed that the watermark exists at the invariant regions with the locally largest *repeatability measure* as the watermark was embedded. After extracting invariant regions from an image, the same region-selection process used in the watermark embedding is performed. Since the selection process after watermark embedding and attacks may not be repeatable, the region-selection process is performed iteratively after excluding the already selected regions until N regions are selected (typically $N = 50$). At each selected region, the detection mask D is obtained by mapping the invariant region to the watermark coordinate as follows:

$$D = f_w(\mathbf{H}\mathbf{x}^T) \tag{20}$$

where f_w is the estimated watermark from an image by using Wiener filter [8]. Note that for each invariant region the same watermark is embedded four times for circular, elliptical and quadrangular regions and three times for triangular regions as shown in Fig. 5. The redundantly embedded parts of a mask are added and then converted into the sequence $V = (v_1, v_2, \ldots, v_M)$. Watermark detection is based on normalized correlation [51] between the original watermark C and the estimated watermark V as follows:

$$Z = \frac{C \cdot \tilde{V}}{\sqrt{(C \cdot C)(\tilde{V} \cdot \tilde{V})}} \tag{21}$$

where $\tilde{V} = (v_1 - \bar{V}, v_2 - \bar{V}, \ldots, v_M - \bar{V})$, $\bar{V}$ is the mean value of V and $C \cdot \tilde{V} = \sum_{i=1}^{M} c_i \tilde{v}_i$. Using the normalized correlation Z, the watermark detection problem can be formulated as the following hypothesis testing:

- H_0: an image is watermarked by C if $Z \geq T$;
- H_1: an image is not watermarked by C if $Z < T$

where T is the watermark detection threshold. We will discuss how to determine the threshold T in Section IV-A.

The watermark detection procedure is summarized as follows:

1) construct invariant regions at each scale from the scale-space Harris points of an image;
2) find N most probably watermarked candidate invariant regions across the scale space;
3) apply the normalized correlation detector to each candidate invariant region after geometric normalization according to the shape of the regions;
4) make a decision whether the image is watermarked or not by testing hypothesis with a threshold T.

IV. Experimental Results

In the experiments, we have used 1024-length watermark pattern $(M = 1024)$ and only 50 invariant regions $(N = 50)$ in the watermark detection. For the characteristic-scale selection and the shape adaptation of the Gaussian kernel, we have used scale-space representation with 36 scale levels (initial scale: 2.3 and the scale factor between two levels of resolution: 1.05). For the feature-point set, we have used scale-space representation with 15 scale levels (initial scale: 3.1 and the scale factor between two levels of resolution: 1.1). The scale-space representation and Harris points can be computed within a minute on a 2.4-GHz Pentium 4 processor for 512 by 512 images. As stated in Section II-D, feature-point sets with cardinality Q $(Q = 2, 3, 4)$ are constructed by selecting Q-1 points out of the 12 strongest Harris points $(P = 12)$ at the neighborhood of each locally strongest Harris point.

Since the watermark pattern is pseudorandomly generated, watermark correlation detector is highly sensitive to the synchronization errors. The pseudorandom watermark pattern was vulnerable to geometric distortions involving more than 0.5 pixel translation, 5% of the characteristic scale and 3° of characteristic orientation. However, the feature-point detector cannot have such a high accuracy since we are dealing with sampled spatial (digital images) and scale spaces. To resolve this mismatch, we performed correlation detection K times by changing the position, the characteristic scale, and orientation of each invariant region in the amount of ± 0.5 pixel, ± 0.025

of a scale, and $\pm 1.5°$, respectively. We call this process *local search*. The two neighborhoods of one-dimensional (1-D) parameters, such as the radius of circles, the tilt angle of ellipses, and the characteristic orientation, and the four neighborhoods of 2-D parameters, such as the position of feature points and the axes (major and minor) of ellipses, are considered in the search. For the circle and ellipse, the local search space amounts, respectively, to 45 (5, 3 and 3 for position, radius, and characteristic orientation, respectively) and 225 (5, 5, 3 and 3 for position, axes, tilt angle, and characteristic orientation, respectively). For the two-, three-, and four-point set, the local search space amounts to 25, 125, and 625 (5 for position of each feature point), respectively. The execution time for the local search was a few minutes for the 50 invariant regions on a 2.4-GHz Pentium 4 processor.

A. False Alarm Analysis

To determine the watermark detection threshold T, the false alarm rate $P_{\rm FA}$ and the false rejection rate $P_{\rm FR}$ should be considered. The false alarm rate $P_{\rm FA}$ is the probability to declare an unmarked image as *marked*. The false rejection rate $P_{\rm FR}$ is the probability to declare a marked image as *unmarked*. There is a tradeoff between the two probabilities in selecting threshold T. In practice $P_{\rm FR}$ is difficult to analyze since there are plenty of image processing steps of those we do not know the exact characteristics. Thus, it is common to select a threshold T of minimizing $P_{\rm FR}$ subject to a fixed $P_{\rm FA}$. First we examine the false alarm rate of each invariant region. By assuming that the watermark has zero mean and unit variance and is independent with the image, the mean $\bar{Z}$ and the standard deviation σ_Z of the normalized correlation Z for unmarked images are given by [51]

$$\bar{Z} = 0, \quad \sigma_Z = \frac{1}{\sqrt{M}}. \tag{22}$$

By the Gaussian assumption of Z, the false alarm rate of the correlation output $P_{\rm FA-cor}$ is given for a certain value of the threshold T as follows:

$$\begin{aligned} P_{\rm FA-cor} &= \int_T^\infty \frac{\sqrt{M}}{\sqrt{2\pi}} \exp\left(\frac{-x^2 M}{2}\right) dx \\ &= \frac{1}{2}\mathrm{erfc}\left(\frac{\sqrt{M}T}{\sqrt{2}}\right). \end{aligned} \tag{23}$$

To compensate for the inaccuracy of the position, the characteristic scale and orientation of the feature point after attacks, the correlation detection is performed at each invariant region K times (*local search*). If at least one out of the K correlation outputs is greater than the threshold T, the region is claimed to be watermarked. Thus, the false alarm rate of each region $P_{\rm FA-reg}$ is given by

$$\begin{aligned} P_{\rm FA-reg} &= \sum_{i=1}^{K} \binom{K}{i} P_{\rm FA-cor}^i (1 - P_{\rm FA-cor})^{(K-i)} \\ &= 1 - (1 - P_{\rm FA-cor})^K. \end{aligned} \tag{24}$$

By viewing all the watermarked regions as independent communication channels, we claim the existence of the watermark

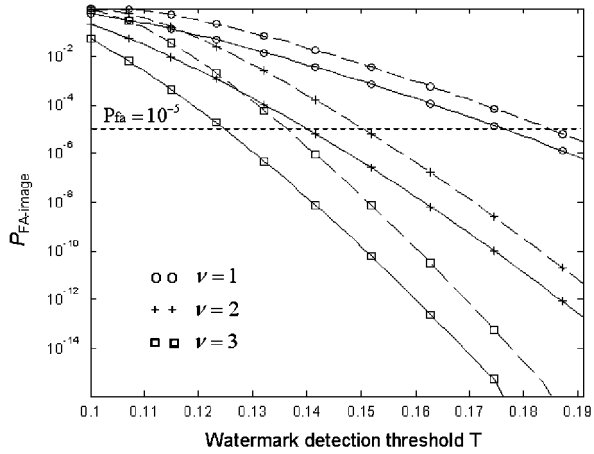

Fig. 6. $P_{\rm FA-image}$ (in log scale) versus watermark detection threshold T for $N = 50, M = 1024$, and $\nu = 1, 2, 3$; solid line: $K = 25$; dashed line : $K = 125$.

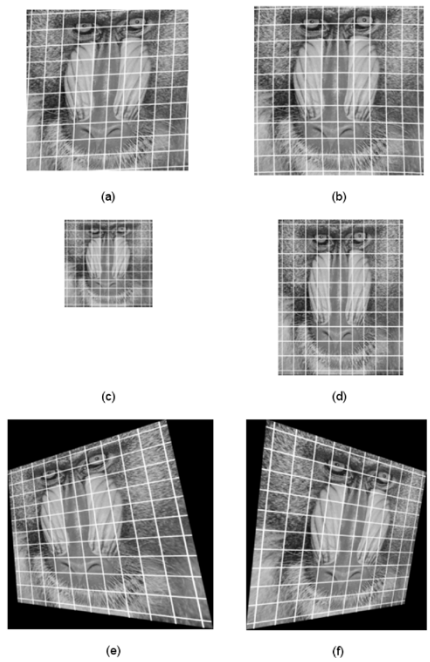

Fig. 7. Examples of attacks on watermarking. (a) Shearing (5% both in x and y directions). (b) Stirmark random-bending attack. (c) Scaling $(x-0.5, y-0.5)$. (d) Nonisotropic scaling $(x-0.7, y-0.9)$. (e) Projective transformation H_1. (f) Projective transformation H_2.

if the same watermark is detected from at least ν number of the invariant regions [22]. Then the false alarm probability for an image $P_{\rm FA-image}$ is given by

$$P_{\rm FA-image} = \sum_{i=\nu}^{N} \binom{N}{i} P_{\rm FA-reg}^i (1 - P_{\rm FA-reg})^{(N-i)} \tag{25}$$

where N is the number of invariant regions used in the watermark detection. In our case, we choose 50 most probably watermarked candidate regions based on the SHCS measure through the scale space as in the watermark embedding and try to detect watermark. All the tests in Section IV-B are done with the threshold at $P_{\rm FA-image} = 10^{-5}$. In practice, many invariant

TABLE II
WATERMARK DETECTION RESULTS UNDER NONGEOMETRIC ATTACKS, CROPPING, AND STIRMARK RANDOM-BENDING ATTACK. THE THREE NUMBERS IN THE TABLES INDICATE THE NUMBER OF THE INVARIANT REGIONS, WHERE THE WATERMARK WAS DETECTED, FOR $\nu = 1, 2, 3$, RESPECTIVELY. THE NUMBER IN THE PARENTHESES REFERS TO THE NUMBER OF WATERMARKED REGIONS

Methods	Image	Gaussian filter	Sharpening filter	Median filter	Uniform noise	JPEG (Q=50%)	JPEG (Q=30%)	Cropping (15% off)	Cropping (25% off)	Stirmark attack
Char. Scale	AI (14)	8, 8, 8	5, 5, 5	4, 4, 6	4, 4, 4	9, 11, 11	5, 7, 8	5, 5, 5	1, 1, 1	6, 7, 7
	BA (22)	6, 7, 8	2, 2, 2	2, 3, 3	1, 1, 1	5, 6, 6	1, 3, 3	6, 6, 6	5, 5, 5	1, 3, 5
	BO (16)	10, 11, 12	6, 6, 6	6, 8, 8	4, 4, 4	8, 9, 9	2, 5, 5	8, 8, 8	8, 8, 8	5, 7, 7
	LA (19)	13, 16, 16	4, 6, 6	7, 8, 8	8, 8, 9	8, 10, 11	2, 10, 11	7, 7, 7	6, 6, 6	7, 10, 10
	LE (14)	12, 12, 12	8, 9, 9	7, 8, 9	5, 5, 5	8, 11, 11	5, 7, 8	6, 6, 6	3, 3, 3	7, 7, 7
	PE (14)	11, 11, 11	6, 6, 6	10, 10, 10	4, 5, 6	12, 12, 12	5, 10, 11	2, 2, 2	3, 3, 3	5, 7, 7
Shape Adapt.	AI (14)	2, 2, 2	1, 1, 1	0, 0, 0	0, 0, 0	0, 0, 0	0, 1, 1	1, 2, 3	1, 2, 3	2, 3, 3
	BA (13)	0, 1, 1	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 2	0, 0, 0	1, 1, 1	1, 1, 1	0, 0, 0
	BO (20)	0, 0, 0	0, 1, 1	0, 0, 0	0, 0, 1	0, 1, 1	0, 1, 2	3, 3, 3	3, 3, 3	0, 1, 1
	LA (16)	1, 2, 5	1, 1, 1	0, 1, 2	1, 1, 1	0, 2, 2	0, 0, 2	3, 7, 7	3, 7, 7	4, 5, 6
	LE (11)	1, 1, 2	0, 0, 0	2, 2, 2	0, 0, 1	1, 1, 2	0, 1, 1	2, 4, 4	2, 4, 4	1, 3, 4
	PE (13)	1, 2, 2	1, 1, 1	2, 2, 2	0, 0, 0	1, 1, 2	0, 1, 1	0, 1, 1	0, 0, 1	2, 2, 3
Two-pt Set	AI (24)	10, 11, 12	11, 12, 12	7, 7, 7	9, 10, 11	7, 9, 12	3, 7, 8	11, 11, 11	10, 10, 10	7, 9, 9
	BA (31)	7, 9, 9	7, 8, 9	6, 7, 7	6, 6, 6	7, 13, 14	1, 4, 6	12, 12, 12	8, 8, 8	5, 8, 10
	BO (22)	12, 15, 16	8, 8, 8	12, 13, 14	9, 10, 10	10, 16, 17	5, 8, 9	7, 7, 7	4, 4, 4	7, 10, 10
	LA (26)	14, 17, 19	12, 12, 12	10, 10, 10	9, 9, 9	11, 15, 16	4, 9, 15	7, 7, 7	7, 7, 7	13, 17, 18
	LE (23)	12, 14, 14	7, 7, 7	11, 12, 12	7, 10, 10	8, 12, 12	2, 6, 10	7, 7, 7	4, 4, 4	11, 12, 13
	PE (28)	14, 15, 15	8, 8, 8	13, 14, 15	7, 8, 9	7, 12, 12	0, 4, 6	6, 6, 7	2, 2, 3	10, 12, 12
Three-pt Set	AI (26)	9, 11, 13	9, 9, 9	6, 6, 6	5, 9, 12	6, 15, 15	1, 10, 11	6, 6, 6	6, 6, 6	12, 14, 14
	BA (36)	4, 5, 7	1, 2, 2	2, 2, 3	0, 3, 4	2, 4, 5	2, 2, 2	8, 9, 9	3, 3, 3	3, 6, 7
	BO (35)	9, 13, 13	8, 8, 8	7, 9, 9	2, 5, 11	3, 6, 9	0, 2, 3	4, 4, 4	5, 5, 5	7, 9, 10
	LA (31)	10, 11, 11	12, 12, 12	10, 11, 12	2, 8, 10	7, 12, 14	1, 1, 8	6, 6, 6	4, 4, 4	7, 9, 9
	LE (27)	9, 9, 9	3, 3, 3	8, 8, 8	0, 4, 4	1, 5, 6	2, 3, 3	4, 4, 4	6, 6, 6	4, 4, 4
	PE (23)	8, 9, 9	7, 7, 7	7, 7, 8	0, 2, 4	3, 4, 5	0, 3, 5	1, 1, 1	2, 2, 2	6, 7, 7
Four-pt Set	AI (12)	7, 9, 9	6, 6, 6	8, 8, 8	7, 7, 8	5, 7, 8	2, 3, 6	2, 2, 3	0, 0, 1	6, 7, 7
	BA (15)	3, 4, 6	2, 4, 4	2, 2, 2	0, 3, 3	2, 6, 7	1, 1, 1	5, 5, 5	0, 0, 0	3, 6, 6
	BO (16)	4, 6, 6	1, 1, 1	6, 6, 7	5, 6, 6	6, 9, 10	1, 4, 8	2, 2, 2	3, 3, 3	7, 7, 7
	LA (14)	5, 5, 5	4, 4, 5	4, 4, 4	1, 2, 2	4, 4, 5	1, 3, 4	0, 0, 0	0, 0, 0	1, 1, 1
	LE (13)	8, 8, 9	5, 5, 5	9, 9, 9	3, 4, 5	5, 7, 8	0, 3, 3	5, 5, 5	2, 2, 2	7, 9, 9
	PE (15)	10, 11, 11	11, 11, 11	5, 5, 5	6, 6, 6	9, 11, 11	3, 4, 7	2, 2, 2	1, 1, 1	5, 5, 5
Tessell-ation [8]	AI (38)	10, 13, 15	9, 9, 10	21, 25, 27	1, 3, 4	8, 13, 18	4, 7, 8	10, 10, 10	2, 2, 2	10, 16, 17
	BA (40)	2, 3, 4	0, 0, 0	0, 0, 0	1, 4, 5	0, 0, 4	1, 2, 3	0, 0, 0	0, 0, 0	2, 5, 5
	BO (57)	8, 12, 13	4, 4, 4	4, 5, 6	4, 11, 11	1, 9, 14	0, 0, 4	5, 5, 5	1, 1, 1	11, 13, 16
	LA (44)	10, 18, 22	15, 15, 15	11, 14, 14	3, 5, 8	5, 17, 25	0, 0, 2	6, 6, 6	4, 4, 4	20, 21, 21
	LE (39)	11, 16, 18	5, 7, 7	10, 13, 13	0, 0, 1	2, 3, 8	0, 2, 4	2, 2, 2	1, 1, 1	12, 15, 16
	PE (60)	20, 23, 25	9, 9, 10	20, 21, 23	1, 6, 11	4, 13, 16	0, 1, 3	2, 2, 3	0, 0, 0	16, 17, 19

regions are overlapping. If we detect watermark at regions significantly overlapping one another, that is counted one. Thus, we can expect that the actual $P_{\text{FA}-\text{image}}$ is lower than that in (25). Fig. 6 shows $P_{\text{FA}-\text{image}}$ versus T for various values of ν and K. With larger ν, the threshold T is smaller for the same $P_{\text{FA}-\text{image}}$. As the local search space (K) grows, the threshold of watermark detection for the same $P_{\text{FA}-\text{image}}$ increases.

B. Robustness Test

The proposed watermarking scheme is tested on the six popular test images: 512 × 512 Airplane, Baboon, Boat, Lake, Lena, and Peppers. The test images can be downloaded at USC-SIPI Image Database.[2] The original watermark pattern C is prepared as a zero-mean pseudorandom 1024-length pattern. The peak signal-to-noise ratio (PSNR) of the watermarked regions was between 34 and 43 dB after masking [50]. Then the resulting PSNRs of the watermarked test images were between 38 and 45 dB. Through the masking, the embedded watermark is hardly perceptible. To test the robustness of the proposed method, all the watermarked images were subjected to various kinds of image processing steps (see [23] for the detailed description of the processing steps). We considered the following two projective transformations H_1 and H_2:

$$H_1 = \begin{bmatrix} 0.62 & -0.23 & -0.0007 \\ 0.69 & 0.69 & -0.0002 \\ -30 & 50 & 1 \end{bmatrix}$$
$$H_2 = \begin{bmatrix} 1.6286 & 0.4537 & 0.0012 \\ -0.1256 & 1.3936 & 0.0002 \\ 55.1381 & -56.0687 & 1.0274 \end{bmatrix}. \quad (26)$$

Examples of the processed images are illustrated in Fig. 7. The effects of the geometric transformations can be easily visualized by the grid in the image.

Watermark detection results for various signal processing steps are shown in Tables II and III. The three numbers in the Tables indicate the number of the invariant regions where the watermark was detected with the three values of threshold (obtained from Fig. 6 for $P_{\text{FA}-\text{image}} = 10^{-5}$) corresponding to $\nu = 1, 2, 3$, respectively. We note that only ν detected regions are needed for claiming watermark existence. For most of the attacks, the detection results were best at $\nu = 1$. However, for uniform noise, JPEG compression, and shearing, the detection results at $\nu = 2$ or 3 were better than those at

[2]http://sipi.usc.edu/database/

TABLE III
WATERMARK DETECTION RESULTS UNDER AFFINE AND PROJECTIVE TRANSFORMATIONS. THE THREE NUMBERS IN THE TABLES INDICATE THE NUMBER OF THE INVARIANT REGIONS, WHERE WATERMARK WAS DETECTED, FOR $\nu = 1, 2, 3$, RESPECTIVELY. NOTE THAT AUTOCROP REFERS TO CROPPING TO ORIGINAL SIZE AFTER ROTATION. THE NUMBER IN THE PARENTHESES REFERS TO THE NUMBER OF WATERMARKED REGIONS

Methods	Image	Rot. 45°	Rot. 45° autocrop	Shearing (5%)	Scaling (0.75, 0.75)	Scaling (0.5, 0.5)	Scaling (1.0, 0.8)	Scaling (0.7, 0.9)	Proj. Tr. H_1	Proj. Tr. H_2
Char. Scale	AI (14)	4, 4, 5	4, 5, 5	3, 5, 5	3, 4, 4	0, 0, 1	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0
	BA (22)	1, 1, 1	2, 2, 2	0, 0, 1	1, 1, 1	0, 0, 1	0, 0, 0	0, 0, 0	0, 0, 1	0, 0, 0
	BO (16)	5, 5, 5	4, 5, 6	0, 0, 1	7, 7, 7	0, 2, 2	0, 0, 0	0, 0, 0	0, 0, 0	0, 1, 2
	LA (19)	8, 8, 8	7, 9, 9	1, 2, 3	1, 3, 5	0, 4, 4	0, 0, 0	0, 0, 0	0, 1, 1	0, 1, 1
	LE (14)	6, 6, 6	5, 5, 5	0, 3, 5	2, 2, 2	2, 3, 3	0, 0, 0	0, 0, 1	0, 0, 0	0, 0, 0
	PE (14)	10, 10, 10	4, 4, 4	0, 1, 3	5, 5, 5	2, 3, 3	0, 0, 1	0, 0, 0	0, 0, 0	1, 1, 1
Shape Adapt.	AI (14)	0, 0, 0	0, 0, 1	1, 3, 3	3, 5, 5	0, 0, 0	0, 2, 3	0, 0, 0	0, 0, 1	2, 3, 3
	BA (13)	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0
	BO (20)	0, 0, 0	0, 0, 0	0, 1, 1	1, 1, 1	0, 0, 0	1, 2, 4	0, 1, 1	1, 2, 3	0, 1, 1
	LA (16)	1, 2, 2	2, 2, 2	2, 2, 2	1, 2, 2	0, 1, 1	1, 4, 5	0, 0, 1	3, 5, 5	1, 3, 3
	LE (11)	0, 1, 1	0, 0, 0	1, 2, 2	1, 2, 2	0, 0, 0	1, 1, 1	0, 1, 1	0, 1, 1	0, 1, 1
	PE (13)	1, 1, 1	0, 0, 0	0, 0, 0	1, 2, 3	0, 0, 0	1, 1, 1	0, 0, 0	0, 0, 0	0, 0, 1
Two-pt Set	AI (24)	7, 8, 8	7, 7, 7	0, 3, 3	5, 5, 5	1, 1, 1	0, 0, 0	0, 0, 0	1, 1, 2	1, 1, 1
	BA (31)	4, 6, 6	4, 6, 7	0, 0, 0	4, 4, 4	1, 1, 1	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0
	BO (22)	6, 6, 6	7, 7, 7	0, 2, 2	5, 5, 5	1, 1, 1	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0
	LA (26)	6, 6, 7	9, 10, 10	2, 3, 3	8, 9, 9	0, 1, 1	0, 0, 0	0, 0, 0	0, 0, 0	1, 1, 1
	LE (23)	6, 6, 7	6, 6, 6	0, 2, 2	8, 11, 11	2, 2, 2	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0
	PE (28)	4, 4, 4	5, 6, 6	0, 1, 4	7, 7, 7	0, 1, 1	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0
Three-pt Set	AI (26)	8, 9, 10	3, 3, 3	5, 5, 5	2, 2, 2	0, 0, 0	7, 7, 7	1, 1, 1	1, 1, 1	2, 3, 3
	BA (36)	2, 2, 3	5, 6, 8	2, 3, 3	1, 1, 1	1, 1, 1	0, 1, 1	1, 1, 1	2, 2, 2	0, 0, 1
	BO (35)	6, 7, 7	4, 4, 4	4, 5, 5	2, 3, 3	0, 0, 0	6, 6, 6	3, 3, 3	0, 3, 4	0, 0, 0
	LA (31)	5, 6, 7	8, 8, 8	4, 4, 4	2, 2, 4	0, 0, 0	1, 5, 5	1, 1, 1	0, 1, 1	1, 1, 1
	LE (27)	2, 2, 2	2, 2, 2	5, 5, 5	4, 4, 4	0, 0, 0	3, 3, 4	1, 1, 1	2, 2, 2	0, 0, 0
	PE (23)	3, 3, 3	3, 4, 4	4, 4, 4	4, 4, 4	0, 0, 0	1, 1, 1	1, 2, 2	0, 1, 1	1, 2, 2
Four-pt Set	AI (12)	7, 7, 7	3, 3, 3	4, 4, 4	1, 1, 1	0, 0, 0	0, 0, 0	0, 0, 0	2, 2, 2	0, 0, 0
	BA (15)	3, 3, 4	2, 3, 4	1, 1, 1	1, 1, 1	0, 0, 0	1, 1, 1	0, 0, 0	0, 0, 0	1, 2, 2
	BO (16)	6, 6, 6	5, 5, 5	2, 2, 2	2, 2, 2	0, 0, 0	3, 3, 3	1, 1, 1	1, 1, 1	3, 3, 3
	LA (14)	3, 3, 3	1, 1, 1	0, 0, 0	3, 3, 3	1, 1, 1	0, 0, 0	1, 1, 1	1, 1, 1	0, 0, 0
	LE (13)	2, 2, 2	4, 4, 4	4, 4, 4	2, 2, 2	0, 0, 0	2, 2, 2	2, 2, 2	2, 2, 2	0, 0, 0
	PE (15)	3, 3, 3	4, 4, 4	5, 5, 5	1, 1, 1	0, 0, 0	4, 4, 4	1, 1, 1	1, 1, 1	0, 0, 0
Tessell-ation [8]	AI (38)	1, 1, 1	11, 13, 14	10, 11, 13	3, 3, 3	0, 0, 0	3, 3, 3	2, 2, 2	0, 0, 2	3, 4, 4
	BA (40)	0, 0, 0	0, 1, 2	1, 3, 3	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0	0, 1, 2
	BO (57)	0, 0, 0	11, 15, 15	13, 14, 15	2, 2, 2	0, 0, 0	9, 10, 10	3, 3, 3	1, 2, 2	0, 0, 0
	LA (44)	0, 0, 0	6, 6, 6	13, 14, 14	1, 1, 1	0, 0, 0	0, 0, 0	0, 0, 0	0, 0, 0	0, 1, 1
	LE (39)	0, 0, 0	7, 8, 9	11, 13, 13	0, 0, 0	0, 0, 0	5, 5, 5	2, 2, 2	1, 1, 1	0, 0, 0
	PE (60)	0, 0, 0	5, 5, 5	4, 5, 5	0, 0, 0	0, 0, 0	6, 6, 6	1, 1, 1	0, 1, 1	1, 1, 1

$\nu = 1$. In practice, employing multiple values of ν rather than one is beneficial in reducing false rejection rate with a nominal increase in computation (i.e., comparison with multiple thresholds). Among the three invariant-region detection methods, the shape adaptation of the Gaussian kernel showed the worst performance. The iteration process in the shape adaptation was not stable enough to withstand most attacks. Thus, the other two methods, which do not rely on iteration, showed better robustness. As shown in Table I, the watermark shape should correspond to the expected geometric transformations. In general, the regions with lower degree of freedom have higher repeatability. Thus, the watermark of simple geometric shapes, such as circle and square, is more robust against nongeometric signal processing steps and similarity transformations. However, if we want to make the watermark robust against nonisotropic scaling and projective transformations, the watermark of complex geometric shapes, such as ellipse, triangle, and quadrangle, should be used. In practice, the watermark-detection performance strongly depends on the location and the image characteristics of the watermarked regions. If a region is located at the border of an image, it might be removed after cropping. The watermarked regions in the textured areas are less robust since many new feature points may show up, and the feature points of the watermarked regions may shift considerably (more than several pixels) after either geometric distortions or noise-like signal processing attacks [8], [22]. The feature points in the intersection of strong edges show higher repeatability. These facts are reflected in the results, and the textured images Baboon and Lake showed less robustness than the other images. The robustness of the proposed method based on the scale-space representation was compared with that of the previous watermarking method based on the tessellation of the Harris points [8]. We tried to detect the watermark at each triangle obtained through the tessellation of the Harris points of an image. For a fair comparison, the same amount of local search was performed at each vertex of the triangles as used in the testing of the proposed method. Even without the scale-space representation, the method in [8] showed notable robustness against most attacks. However, as we expected, the method was susceptible to the scaling attack. In this paper, we show that this problem can be resolved using the scale-space Harris points. We note that watermarking methods based on the characteristic scale and the two-point set are robust against scale changes due to the scale-space representation. The price

paid for the improved performance is the increased complexity in computing the scale-space representation. However, note that there exists an efficient way to compute the scale-space representation of an image [26].

V. Conclusion

This paper shows that the resilience of the watermark against geometric distortions can be significantly improved by synchronizing watermarking with the invariant regions of the scale-space representation of an image. The robustness against geometric distortions is essential because it is quite easy to impose geometric distortions to images with modern computers. The original image is not needed at the watermark detector. Two limitations are commonly quoted on the content-based watermarking methods: the inaccuracy of feature point detector [8], [22] and the added computational complexity [20]. In this paper, the inaccuracy was overcome by using the local search. Obviously, the proposed method is computationally more demanding than watermarking methods that do not consider geometric distortions. However, only *50* invariant regions, selected based on the SHCS measure, are used in the watermark detector. It is noticeably less burdensome than exhaustive search. The experimental results show that the robustness of the proposed content-based watermarking method, especially to scaling and cropping, is improved by using the invariant regions of the scale-space Harris points. This implies that there may be room for improvement of watermarking performance by using refined pattern recognition techniques. The watermark of simple geometric shapes, such as circle and square, is certainly more robust against both the nongeometric signal processing steps and the similarity transformations than that of complex geometric shapes, such as ellipse, triangle and quadrangle. Moreover, the complex geometric shapes require more computation in both the invariant-region detection and the local search for watermark detection. However, the watermarks with complex geometric shapes perform better for the nonisotropic scaling and the projective transformations. Future work includes the development of more robust and computationally efficient invariant-region detector, especially on textured areas, and the extension of the proposed method to video.

Acknowledgment

The authors would like to thank the anonymous reviewers for their detailed and helpful comments.

References

[1] F. Hartung and M. Kutter, "Multimedia watermarking technique," *Proc. IEEE*, vol. 87, pp. 1079–1107, 1999.

[2] D. Boneh and J. Shaw, "Collusion-secure fingerprinting for digital media," *IEEE Trans. Inf. Theory*, vol. 44, no. 5, pp. 1897–1905, Sep. 1998.

[3] T. Kalker, G. Depovere, J. Haitsma, and M. Maes, "A video watermarking system for broadcast monitoring," in *Proc. IS&T/SPIE Electronic Imaging*, 1999, pp. 103–112.

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

[5] M. Maes, T. Kalker, J.-P. M. Linnartz, J. Talstra, G. F. Depovere, and J. Haitsma, "Digital watermarking for DVD video copy protection," *IEEE Signal Process. Mag.*, vol. 17, no. 5, pp. 47–57, Sep. 2000.

[6] P. Dong, J. Brankov, N. Galatsanos, and Y. Yang, "Geometric robust watermarking based on a new mesh model correction approach," in *Proc. IEEE Int. Conf. Image Processing*, Jun. 2002, pp. 493–496.

[7] F. Davoine, "Triangular meshes: A solution to resist to geometric distortions based watermark-removal softwares," in *Proc. Eur. Signal Processing Conf.*, vol. 3, Jun. 2000, pp. 493–496.

[8] P. Bas, J.-M. Chassery, and B. Macq, "Geometrically invariant watermarking using feature points," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 1014–1028, Sep. 2002.

[9] J. Ruanaidh and T. Pun, "Rotation, scale and translation invariant spread spectrum digital image watermarking," *Signal Process.*, vol. 66, no. 3, pp. 303–317, 1998.

[10] C.-Y. Lin, M. Wu, J. Bloom, M. Miller, I. Cox, and Y.-M. Lui, "Rotation, scale, and translation resilient public watermarking for images," *IEEE Trans. Image Process.*, vol. 10, no. 5, pp. 767–782, May 2001.

[11] D. Zheng, J. Zhao, and A. Saddik, "RST-invariant digital image watermarking based on log-polar mapping and phase correlation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 8, pp. 753–765, Aug. 2003.

[12] S. Pereira and T. Pun, "Robust template matching for affine resistant image watermarks," *IEEE Trans. Image Process.*, vol. 9, no. 6, pp. 1123–1129, Jun. 2000.

[13] M. Kutter, "Watermarking resisting to translation, rotation and scaling," in *Proc. SPIE Multimedia Systems Applications*, vol. 3528, Boston, MA, Jan. 1998, pp. 423–431.

[14] S. Pereira and T. Pun, "An iterative template matching algorithm using the Chirp-Z transform for digital image watermarking," *Pattern Recognit.*, vol. 33, pp. 173–175, 2000.

[15] A. Herrigel, S. Voloshynovskiy, and Y. Rytsar, "The watermark template attack," in *Proc. IS&T/SPIE Electronic Imaging*, San Jose, CA, Jan. 2001, pp. 394–405.

[16] M. Alghoniemy and A. Tewfik, "Geometric distortion correction through image normalization," in *Proc. IEEE Int. Conf. Multimedia Expo.*, vol. 3, Jul.–Aug. 2000, pp. 1291–1294.

[17] P. Dong and N. Galatsanos, "Affine transformation resistant watermarking based on image normalizaton," in *Proc. IEEE Int. Conf. Image Processing*, Jun. 2002, pp. 489–492.

[18] H. Kim and H.-K. Lee, "Invariant image watermark using Zernike moments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 8, pp. 766–775, Aug. 2003.

[19] M. Alghoniemy and A. H. Tewfik, "Geometric invariance in image watermarking," *IEEE Trans. Image Process.*, vol. 13, no. 2, pp. 145–153, Feb. 2004.

[20] A. Nikolaidis and I. Pitas, "Region-based image watermarking," *IEEE Trans. Image Process.*, vol. 10, no. 11, pp. 1726–1740, Nov. 2001.

[21] J. Dittmann, T. Fiebig, and R. Steinmetz, "A new approach for transformation invariant image and video watermarking in the spatial domain: SSP-self spanning patterns," in *Proc. IS&T/SPIE Electronic Imaging*, Jan. 2000, pp. 176–185.

[22] C.-W. Tang and H.-M. Hang, "A feature-based robust digital image watermarking scheme," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 950–959, Apr. 2003.

[23] F. Petitcolas, "Watermarking schemes evaluation," *IEEE Signal Process. Mag.*, vol. 17, no. 5, pp. 58–64, Sep. 2000.

[24] T. Lindeberg, "Feature detection with automatic scale selection," *Int. J. Comput. Vis.*, vol. 30, no. 2, pp. 77–116, 1998.

[25] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," in *Proc. IEEE Int. Conf. Computer Vision*, vol. 1, Jul. 2001, pp. 525–531.

[26] D. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Computer Vision*, vol. 2, Sep. 1999, pp. 1150–1157.

[27] T. Lindeberg and J. Garding, "Shape-adapted smoothing in estimation of 3-D shape cues from affine deformations of local 2-D brightness structure," *Image Vis. Comput.*, vol. 15, pp. 415–434, 1997.

[28] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *Proc. Eur. Conf. Computer Vision*, May 2002, pp. 128–142.

[29] M. Brown and D. Lowe, "Invariant features from interest point groups," in *Proc. British Machine Vision Conf.*, Sep. 2002, pp. 656–665.

[30] H. Wolfson and I. Rigoutsos, "Geometric hashing: An overview," *IEEE Comput. Sci. Eng. Mag.*, vol. 4, no. 4, pp. 10–21, Dec. 1997.

[31] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[32] T. Tuytelaars and L. V. Gool, "Matching widely separated views based on affine invariant regions," *Int. J. Comput. Vis.*, vol. 59, no. 1, pp. 61–85, 2004.

[33] J. Seo and C. Yoo, "Image watermarking based on scale-space representation," in *Proc. IS&T/SPIE Electronic Imaging*, Jan. 2004, pp. 560–570.
[34] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *Int. J. Comput. Vis.*, vol. 37, no. 2, pp. 151–172, 2000.
[35] N. Sebe, Q. Tian, E. Loupias, M. Lew, and T. Huang, "Evaluation of salient point techniques," *Image Vis. Comput.*, vol. 21, no. 13–14, pp. 1087–1095, 2003.
[36] A. Baumberg, "Reliable feature matching across widely separated views," in *Proc. IEEE Int. Conf. Computer Vision Pattern Recognition*, vol. 1, Jun. 2000, pp. 774–781.
[37] S. Linnainmaa, D. Harwood, and L. Davis, "Pose determination of a three-dimensional object using triangle pairs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 5, pp. 634–647, Sep. 1988.
[38] R. Haralick and L. Shapiro, *Computer and Robot Vision*. Reading, MA: Addison-Wesley, 1992.
[39] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vision Conf.*, 1988, pp. 147–151.
[40] T. Lindeberg, *Scale-Space Theory in Computer Vision*. Norwell, MA: Kluwer , 1994.
[41] J. Babaud, A. Witkin, M. Baudin, and R. Duda, "Uniqueness of the Gaussian kernel for scale-space filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 1, pp. 26–33, 1986.
[42] J. Koenderink, "The structure of images," *Biol. Cybern.*, vol. 50, pp. 363–370, 1984.
[43] R. Young, "The Gaussian derivative model for spatial vision: I. Retinal mechanisms," *Spatial Vis.*, vol. 2, no. 4, pp. 273–293, 1987.
[44] A. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[45] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
[46] T. Suk and J. Flusser, "Point-based projective invariants," *Pattern Recognit.*, vol. 33, pp. 251–261, 2000.
[47] J. Mundy and A. Zisserman, *Geometric Invariance in Computer Vision*. Cambridge, MA: MIT Press, 1992.
[48] C. Schmid and R. Mohr, "Local gray value invariants for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 5, pp. 530–535, May 1997.
[49] A. Reeves, R. Prokop, S. Andrews, and F. Kuhl, "Three-dimensional shape analysis using moments and Fourier descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 6, pp. 937–943, Nov. 1988.
[50] B. Girod, "The information theoretical significance of spatial and temporal masking in video signals," presented at the SPIE/SPSE Conf. Human Vision, Visual Processing Digital Display, 1989.
[51] I. Cox, M. Miller, and J. Bloom, *Digital Watermarking*. San Mateo, CA: Morgan Kaufmann, 2002.

Jin S. Seo (S'00–A'05) received the B.S., M.S., and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1998, 2000, and 2005, respectively, all in electrical engineering.

While working toward the Ph.D. degree, he was a thesis trainee at the Philips Research Laboratories, Eindhoven, The Netherlands, in 2002. He joined the Digital Contents Research Division at Electronics and Telecommunications Research Institute (ETRI), Daejeon, Korea, in 2006. His research interests are multimedia watermarking and retrieval, pattern recognition, and signal enhancement.

Chang D. Yoo (S'92–M'96) received the B.S. degree in engineering and applied science from the California Institute of Technology, Pasadena, in 1986, the M.S. degree in electrical engineering from Cornell University, Ithaca, NY, in 1988, and the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1996.

From January 1997 to March 1999, he worked as a Senior Researcher at Korea Telecom, Seoul. In April 1999, he joined the Department of Electrical Engineering at the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea. His current research interests include the application of digital signal processing theory in speech and audio, speech recognition, digital communication, and adaptive signal processing.

Dr. Yoo is a Member of Tau Beta Pi and Sigma Xi.