Rotation, scale, and translation resilient public watermarking for images

Ching-Yung Lin[a], Min Wu[b], Jeffrey A. Bloom[c],
Ingemar J. Cox[c], Matt L. Miller[c], Yui Man Lui[d]

[a]Columbia University, Dept. of Electrical Engineering, New York, NY, USA
[b]Princeton University, Dept. of Electrical Engineering, Princeton, NJ, USA
[c]NEC Research Institute, 4 Independence Way, Princeton, NJ, USA
[d]Signafy, Inc., 4 Independence Way, Princeton, NJ, USA

ABSTRACT

Many electronic watermarks for still images and video content are sensitive to geometric distortions. For example, simple rotation, scaling, and/or translation (RST) of an image can prevent detection of a public watermark. In this paper, we propose a watermarking algorithm that is robust to RST distortions. The watermark is embedded into a 1-dimensional signal obtained by first taking the Fourier transform of the image, resampling the Fourier magnitudes into log-polar coordinates, and then summing a function of those magnitudes along the log-radius axis. If the image is rotated, the resulting signal is cyclically shifted. If it is scaled, the signal is multiplied by some value. And if the image is translated, the signal is unaffected. We can therefore compensate for rotation with a simple search, and for scaling by using the correlation coefficient for the detection metric.

False positive results on a database of 10,000 images are reported. Robustness results on a database of 2,000 images are described. It is shown that the watermark is robust to rotation, scale and translation. In addition, the algorithm shows resistance to cropping.

Keywords: watermarking, RST, Fourier-Mellin

1. INTRODUCTION

We examine a watermarking scheme's robustness against the geometric distortions of rotation, scale, and translation (RST). The problem of watermark detection after geometric distortion has traditionally been approached by preceding the detection with a registration of the suspect image in an attempt to invert the unknown distortion. When the original image is available to the detector, registration is a well studied problem.

In a public watermark, detection must be performed without the original image. Two approaches for image registration have been proposed in this case. One solution involves the insertion of a second watermark, which is a specifically designed spatial pattern that is easily detected after geometric distortion.[1,2] From the distorted registration pattern, the RST parameters can be determined.

A second solution is to impart the recognizable structure to the data-carrying watermark itself. For example, as suggested in [3], the watermark might be encoded with a small, rectangular pattern, and embedded several times in the image in a tiled grid. Then, regardless of the watermark pattern, the grid structure can be recognized by looking at the autocorrelation function of the image, which would contain a corresponding grid of peaks. These peaks can be analyzed to identify any affine distortions.

In this paper, we propose a 1D watermark signal derived from the image so as to be invariant to RST distortions. This type of signal, or descriptor, is well studied in the pattern recognition community. The current approach is similar to the development of "strong" or "absolute" invariants based on or related to the Fourier-Mellin transform.[4–12] The terms absolute and strong refer to the fact that all information about an image except that of position, orientation or scale is preserved. This may be important for recognition tasks, especially if the library of objects is large.

O'Ruanaidh and Pun[13] first suggested a watermarking method based on the Fourier-Mellin transform. However, they note very severe implementation difficulties which we suspect have hampered further work in this area. Many of these difficulties are related to the fact that they have chosen to use a transformation that is strongly invariant.

This work was performed while all authors were employed by NEC Research Institute or Signafy Inc.

The watermark signal proposed in the current work is not based on a strong invariant, as we believe that strong invariance is not necessary for watermarking applications.

There have been a number of other recent watermarking algorithms designed to deal with geometric distortions. Of particular note is the recent work of Bas *et al*[14] that inserts a signal relative to the salient features of an image. A somewhat related set of methods is described by Maes and van Overveld[15] and Rongen *et al.*[16] These methods are based on geometrically warping local regions of an image onto a set of random lines.

In Section 2 we describe our algorithm. It differs from that of [13] in two primary ways. First, we choose to watermark a projection of the transform space. Second, the watermark embedding is based on the principle of communication with side information.[17] This is described in more detail, including the iterative procedure for dealing with the one-to-many mapping from watermark to image space. Section 3 describes the results of a broad range of experiments on a large database.

Before proceeding further, it is important to define what we mean by the geometric distortions of rotation, scale and translation. Specifically, we are interested in the situation in which a watermarked image undergoes an *unknown* rotation, scale and/or translation prior to the detection of the watermark. The detector should detect the watermark if it is present. This definition is somewhat obvious, so it may be more useful to describe what we are not interested in. In particular, some watermark algorithms claim robustness to scale changes by first embedding a watermark at a canonical scale, then changing the size of the image and finally, at the detector, scaling the image back to the canonical size prior to correlation. In our opinion, the detector does not see a scale change. Rather, the process is more closely approximated by a low pass filtering operation that occurs when the image is reduced in size. Similarly, tests that rotate an image by some number of degrees and subsequently rotate the image by the same amount in the opposite direction are not adequate tests of robustness to rotation. The same is true for translation. The common situation we are concerned with occurs when a watermarked image is printed and then cropped or padded and scanned back into the digital domain. In these circumstances, the image dimensions have changed both because of cropping and possibly scaling. There is also likely to be an associated translational shift. In this example, scaling to a canonical size does not undo the scaling. Rather, if the cropping is not symmetric in both the rows and columns, then scaling to a canonical size will result in a change in the image's aspect ratio. Changes in aspect ratio are not addressed in this paper. Application of the current watermarking method to the print and scan process has been discussed elsewhere.[18]

2. ALGORITHM

It is well known that the magnitude of the Fourier transform of an image after rotation, scaling, and translation is given in as

$$|I'(f_x, f_y)| = |\sigma|^{-2} \left| I\left(\sigma^{-1}(f_x cos\alpha + f_y sin\alpha), \sigma^{-1}(-f_x sin\alpha + f_y cos\alpha)\right)\right| \tag{1}$$

where α and σ are the rotation and scale parameters, respectively, and $I(f_x, f_y)$ is the Fourier transform of the original image. Rewriting in log-polar frequency coordinates yields

$$|I'(\rho, \theta)| = |\sigma|^{-2} \left| I\left(\rho - \log\sigma, \theta - \alpha\right)\right|, \tag{2}$$

where ρ is along the log radius axis and θ is along the angle axis.

Equation 2 demonstrates that the amplitude of the log-polar spectrum is scaled by $|\sigma|^{-2}$, that image scaling results in a translational shift of $\log\sigma$ along the ρ axis, and that image rotation results in a rotational shift of α along the θ axis. These facts are well known in the pattern recognition community and are the basis for the Fourier-Mellin transform. The scaling will not affect the detection value (see Section 2.1), so the RST geometric distortion appears as a 2D shift of the magnitude spectrum.

Lacking a fast, efficient technique for direct computation of the log-polar Fourier transform, an interpolation from the Cartesian Fourier transform is performed. In the current implementation, the image is padded with black, prior to calculation of the DFT. This results in a denser sampling of the Fourier transform. The magnitudes of the log-polar Fourier sample points are then interpolated with an inexpensive, linear interpolation from the magnitudes of the four nearest neighbors.

The log magnitudes of the 2D log-polar spectrum are then summed over a band of radii to yield a 1D signal,

$$g(\theta) = \sum_j \log\left(|I(\rho_j, \theta)|\right) \qquad 0° \le \theta < 180°, \tag{3}$$

where log values have been used to allow some equalization of the component contributions over the frequency range. The band of radii excludes the very high and very low frequencies.

The energy in an image is seldom evenly distributed in angular frequency. Images frequently have a large amount of energy in one group of directions, while having much lower energy in an orthogonal group of directions (consider images containing buildings or trees and those containing seascapes or sunsets). This suggests an uneven visual masking ability in orthogonal directions.

To minimize the impact on fidelity, $g(\theta)$ is divided into two halves which are added together,

$$g_1(\theta') = g(\theta') + g(\theta' + 90^\circ) \qquad 0^\circ \leq \theta < 90^\circ. \tag{4}$$

Modification of an element of $g_1(\theta)$, can be accomplished by hiding noise that's oriented along either angle θ or angle $\theta + 90^\circ$. This increases the likelihood that each element of the watermark can be embedded within the fidelity constraints. Clearly, $g_1(\theta)$ is invariant to both translation and scaling while rotations result in a circular shift of the values of $g_1(\theta)$.

2.1. Watermark detection

In principle, detectors may be built that can handle watermarks encoding several bits. However, the present detector determines only whether or not a given watermark has been embedded in a given image.

The watermark, w, is expressed as a vector of length N. To determine whether the watermark is present, an "extracted signal", $v = g_1(\theta)$, is computed from the image, for N values of θ evenly spaced between 0° and 90°. The extracted signal is then compared to the watermark using the correlation coefficient and this result is compared to a threshold.* The correlation coefficient is defined as

$$C = \frac{w \cdot v}{\sqrt{(w \cdot w)(v \cdot v)}}. \tag{5}$$

Correlation coefficient is most effective when the elements of the two vectors, v and w, are drawn from white distributions. For natural images, $g_1(\theta)$ is likely to vary smoothly as a function of θ. To improve the detection measure, the watermark and extracted vector are whitened prior to calculation of the correlation coefficient.[19] The whitening filter was empirically derived from signals extracted from 10,000 images in the database.[20] (These images were then excluded from use in subsequent experiments).

It is well known that the rectangular boundary of an image usually causes a "cross" artifact in the image's energy spectrum. The cross, unrelated to the frequency content of the underlying image data, may not be subject to the same geometric transformations as is the image data. Since the energy corresponding to this artifact lies in perpendicular directions, it will all be projected to one small neighborhood in the extracted signal, $g_1(\theta)$.

Our present solution to this problem is to exclude the neighborhood around the highest-valued element of the extracted signal. Alternative solutions that appear in the literature include multiplication of the image by a circularly-symmetric window[21] and blurring of the image edges.[22] These solutions are currently under investigation.

2.2. Watermark embedding process

Watermark embedding is viewed as communications with side information at the embedder.[17] Knowing the detection process, the role of the embedder is to modify the image so as to maximize the correlation coefficient between the extracted signal and the target watermark vector.

To apply the concept of using side information at the embedder, while maintaining acceptable fidelity, Cox *et al*[17] introduces the idea of a *mixing function*, $s = f(v, w)$. This generates a *mixed signal*, s, which is perceptually similar to v, and has a high correlation with w. The mixing function used in the current implementation simply computes a weighted average of w and v, which is a highly sub-optimal approach. More sophisticated methods will be explored in the future. The embedder adds the difference between the extracted signal and the mixed signal to the host image.

*The use of correlation coefficient as a detection measure is recommended in [17]. One benefit of this metric is its independence to scaling of the signal amplitudes.

In the current implementation the difference signal, at each angle, is uniformly distributed throughout all log radii elements of the log-polar Fourier magnitudes. The modified Cartesian Fourier magnitudes are then obtained by inverting the log-polar resampling of the Fourier magnitudes. Finally, the complex terms of the original Fourier transform are scaled to have the new magnitudes that were computed in the modified Fourier transform, and the inverse Fourier transform is applied to obtain the watermarked image.

The main implementation issue in such an approach is the inherent instability in inverting the log-polar resampling. We therefore approximate this step with an iterative method in which a local inversion of the interpolation function is used for the resampling.

Recall that, in the current implementation of the log-polar resampling, each element of the log-polar Fourier magnitude array is a weighted average of up to four elements of the Cartesian Fourier magnitude array. Thus, the change in the magnitude of a log-polar coefficient can be represented as a weighted change in the magnitudes of each of four Cartesian coefficients. Where multiple changes to a single Cartesian coefficient are required, a weighted average is applied.

Successively better approximations are obtained by applying this operation iteratively. The results presented in the next section of this paper were obtained with three iterations.

3. EXPERIMENTAL RESULTS

The following results were obtained by extracting a length 90 vector from the image and neglecting the 16 samples surrounding the peak (assumed to correspond to the DFT cross artifact). This leaves a descriptor that is 74 samples in length. The detection process involves a comparison of the watermark with all 90 cyclic rotations of the extracted descriptor. In this section we examine the false positive behavior, effectiveness, and robustness of the proposed scheme. False positive measurements were collected on 10,000 unwatermarked images[†], and effectiveness and robustness measurements were collected on 2,000 watermarked images except that scale up with cropping used only 947 images. Watermarks were embedded with a strength that resulted in a mean SNR of 40dB.[‡]

3.1. Probability of False Positive

A false positive or false detection occurs when the detector incorrectly concludes that an unwatermarked image contains a given watermark. Thus, the probability of false positive is defined as

$$P_{fp} = P\{C_{max} > T\} \tag{6}$$

where C_{max} is a detection value obtained by running the detector on a randomly selected, unwatermarked image and T is the detection threshold. The subscript max specifies the maximum detection value from all of the cyclical shifts examined.

This probability is estimated empirically by applying the detector to 10,000 unwatermarked images from [20], testing for 10 different binary watermarks in each. The probability of detection (in this case false detection) is shown as a function of threshold in Figure 1 (a). Each trace corresponds to all 900,000 correlation coefficients obtained using one of the 10 watermarks. The highest detection value obtained in this experiment was 0.55.

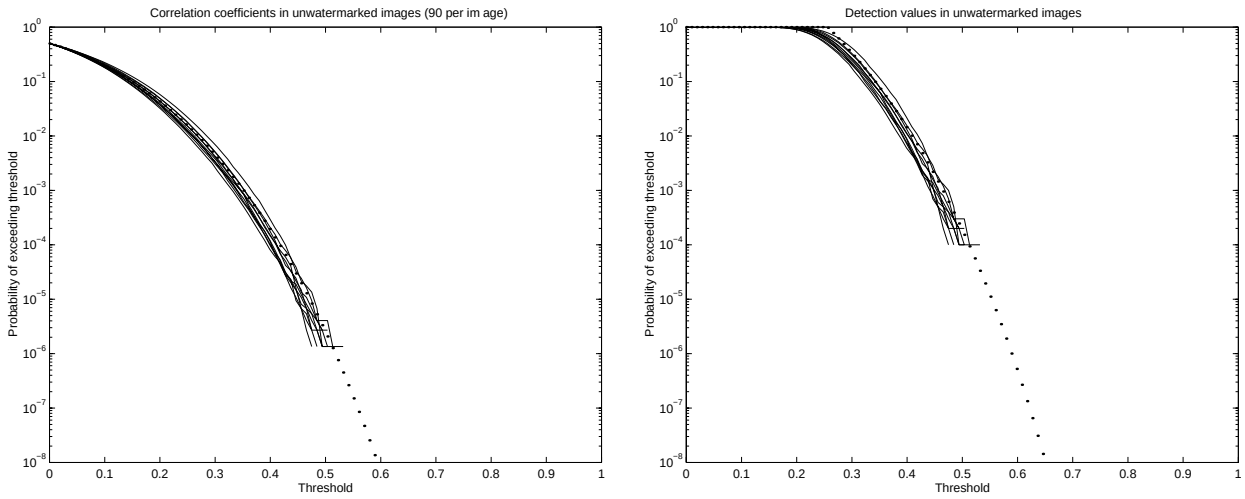
To estimate P_{fp} for $T > 0.55$, we employ the theoretical model described by Miller & Bloom.[23] This model predicts the false positive probability when an unwatermarked, extracted vector is drawn from a radially-symmetric distribution and the correlation coefficient, C is used as the detection metric. For a d-dimensional watermark, this probability is

$$P\{C > T\} = R(T,d) \equiv \frac{\int_0^{cos^{-1}(T)} \sin^{d-2}(u)du}{2\int_0^{\pi/2} \sin^{d-2}(u)du}. \tag{7}$$

The whitening filter employed in the detector makes the distribution roughly spherical, so this model is expected to apply with $d = 74$. The resulting false positive rate prediction is shown as a dotted line in Figure 1 (a).

[†]The images used in this test were all different from, but from the same database as the 10,000 images that were used to generate the whitening filter.

[‡]Here the "signal" is the image, and the "noise" is the watermark pattern.



(a) (b)

Figure 1. Measured false positive rates plotted with theoretical estimates. (a) Individual correlation coefficients. (b) Final detection value

Ultimately, however, we are concerned with the probability that, for each watermark and unwatermarked image, the *maximum* among 90 correlation coefficients is greater than the threshold. These maximum detection values are used to generate Figure 1 (b).

The model can be used to estimate an upper bound on $P\{C_{max} > T\}$ by observing that

$$P\{Q_0 \ or \ Q_1 \ or \ldots Q_{n-1}\} \leq \min\left(1, \sum_i P\{Q_i\}\right). \tag{8}$$

When Q_i corresponds to the event $(C_i > T)$, and $n = 90$, we obtain

$$P\{C_{max} > T\} \leq \min\left(1, 90 \times R(T, 74)\right). \tag{9}$$

This prediction is shown as a dotted line in Figure 1(b).

3.2. Effectiveness

The effectiveness of a watermarking scheme is measured as the probability that the output of the watermark embedder will contain the watermark, subject to constraints on the fidelity of the marked image and the detection threshold or probability of false positive. The effectiveness of the current scheme is measured and reported in Table 1 (row labeled "Eff"). Improvements are possible in the approximate inversion of the log-polar resampling and in the distribution of the difference signal to the log-polar coefficients. Also, the current technique of uniform distribution does not fully exploit the visual properties of the host image.

3.3. Robustness

In a practical setting, RST distortions are usually accompanied by cropping. Figure 2 (f), (g), and (i) show respectively rotation, scaling, and translation with the associated cropping. With the current algorithm, cropping can be viewed as distortion of the extracted signal by additive noise. As such, we expect cropping to degrade the detection value.

In this section seven geometric distortion attacks are examined; rotation with and without cropping, scaling up with and without cropping, translation with and without cropping, and scaling down. Note that scaling down does not imply cropping. In order to isolate the effects of rotation, scaling up, and translation from cropping, the images have been padded with gray as shown in Figure 2 (a). The embedder has been applied to these expanded images and then the gray padding replaced with unwatermarked gray padding prior to detection or attack.

	Threshold				
	0.50	0.55	0.60	0.65	0.70
P_{fp}	10^{-4}	10^{-5}	10^{-7}	10^{-8}	10^{-10}
Eff	95.3	92.1	86.3	71.8	43.2
Rot	97.2	93.6	82.7	63.7	45.8
	95.3	88.2	71.1	49.4	28.4
Scl Up	99.2	97.2	92.8	81.1	66.9
	97.1	93.6	83.9	68.1	49.9
Trans	100	100	100	100	100
	99.0	98.8	97.5	96.0	92.7

(a)

	Threshold				
	0.50	0.55	0.60	0.65	0.70
P_{fp}	10^{-4}	10^{-5}	10^{-7}	10^{-8}	10^{-10}
Eff	97.4	95.7	93.0	88.2	75.6
Rot	94.9	88.6	74.7	51.6	22.7
	85.7	71.1	49.7	26.0	8.7
Scl Up	98.8	97.9	93.9	82.6	58.9
	95.4	88.5	73.6	51.4	26.7
Trans	95.2	89.5	77.3	55.4	29.7
	91.9	83.8	66.3	40.3	18.5
Scl Dn	99.9	100	100	99.5	96.6
	99.6	99.6	99.1	98.6	92.9

(b)

Table 1. Effectiveness and robustness from (a) padded and (b) unpadded tests. Listed are the maximum and minimum robustness values for rotation angles of 4°, 8°, 30°, and 45°; scaling factors of 5%, 10%, 15%, and 20%; and translations factors of 5%, 10%, and 15% of the image size.

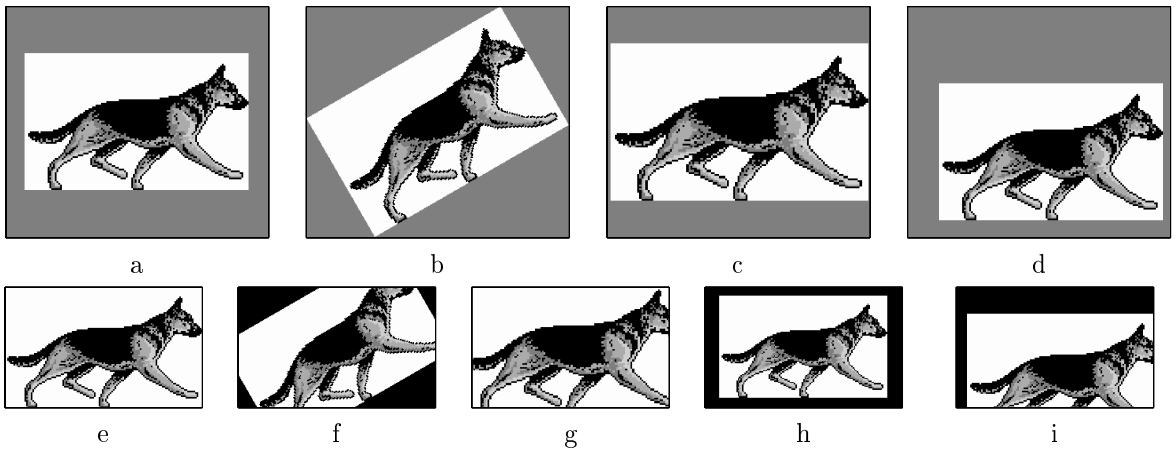


Figure 2. Examples of geometric attacks: (e) and (a) are the original and padded original respectively, (b)-(d) attacks without cropping, and (f)-(i) attacks with cropping

The detection value prior to attack is used to measure the effectiveness of the watermarking scheme. This effectiveness is likely to be reduced in the padded examples since a portion of the watermarked image (the watermarked gray padding) has been replaced with non-watermarked data. However, the purpose of the experiments shown in Figure 2 (b)-(d) and reported in Table 1 (a) is to isolate the effects of geometric distortions from cropping effects. The results of experiments that were based on the original, unpadded images are reported in Table 1 (b). These tests include rotation with cropping, scaling up with cropping, translation with cropping, and scaling down.

For a given threshold, the table lists the percentage of images that withstood each attack, i.e. robustness is the ratio of the number of detections after attack to the number of detections prior to attack. Notice that only images for which the embedding was effective (detection value immediately after embedding was above the detection threshold) contribute to the determination of robustness. For each threshold, the false-positive probabilities are estimated using the model described in Section 3.1.

The results shown in Table 1(a) are for RST distortions without cropping. We see that this algorithm is extremely robust to translation and shows very good robustness to rotation and scale up at moderate false positive rates. The algorithm is also extremely robust to scale down, the results of which are reported in Table 1(b) and are also free from cropping. Table 1(b) also shows the results of RST distortions with cropping. The difference between the translation results of the two tables reveals the sensitivity of this method to cropping. This sensitivity can also be seen in the rotation and scale up experiments.

4. CONCLUSION

Geometric distortions continue to be a major weakness for many watermarking methods. We have described a solution to the common problems of rotation, scale, and translation. This solution is related to earlier proposals in the pattern recognition literature regarding invariants of the Fourier-Mellin transform. However, unlike those proposals, we do not explicitly derive an invariance relationship.

Instead of creating a truly RST invariant signal, we create a signal that can be searched for the effects of RST in a trivial manner. The calculation of this projection is performed by taking the Fourier transform of the image, performing a log-polar resampling and then integrating along the radial dimension. We note that an alternative implementation can be performed using the Radon transform.[24] We have also investigated that implementation but do not report it here.

The one-dimensional watermark has a many-to-one mapping to the two-dimensional image space. This is advantageous, especially when the embedder is based on the principle of communications with side information. Our implementation is a very simple example of this principle and we believe that future work can lead to significant improvements.

Experimental results on a database of over 2,000 images clearly demonstrate that the method is robust to rotations, scale changes, or translations and, without modification, exhibits some resistance to cropping.

Acknowledgements

The authors would like to thank Dr. Harold Stone of NEC Research Institute for helpful discussions on image registration and Fourier-Mellin transform. In addition, the authors would like to thank Dr. J.J.K.O'Ruanaidh at Siemens Research Center, Princeton, NJ, for useful discussions. The test images used in experiments are from the Corel Stock Photo Library 3 image database.[20]

REFERENCES

1. S. Pereira and T. Pun, "Fast robust template matching for affine resistant image watermarks," in *Proc. of the 3rd Int. Information Hiding Workshop*, pp. 207–218, 1999.
2. G. Csurka, F. Deguillaume, J. J. K. O'Ruanaidh, and T. Pun, "A Bayesian approach to affine transformation resistant image and video watermarking," in *Proc. of the 3rd Int. Information Hiding Workshop*, pp. 315–330, 1999.
3. M. Kutter, "Watermarking resistance to translation, rotation, and scaling," in *SPIE Conf. on Multimedia Systems and Applications*, vol. SPIE 3528, pp. 423–31, 1998.

4. M. K. Hu, "Visual pattern recognition by moment invariants," *IEEE Trans. on Info. Theory* **IT-8**, pp. 179–187, 1962.
5. Y. Sheng and H. H. Arsenault, "Experiments on pattern recognition using invariant Fourier-Mellin descriptors," *J. Opt. Soc. Am. A* , pp. 771–776, 1986.
6. D. Casasent and D. Psaltis, "Position, rotation, and scale invariant optical correlation," *Applied Optics* **15**(7), pp. 1795–1799, 1976.
7. D. Casasent and D. Psaltis, "New optical transforms for pattern recognition," *Proc. of the IEEE* **65**(1), pp. 77–84, 1977.
8. J. Altmann and H. J. P. Reitbock, "A fast correlation method for scale- and translation-invariant pattern recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence* **6**(1), pp. 46–57, 1984.
9. J. Altmann, "On the digital implementation of the rotation-invariant Fourier-Mellin transform," *Journal of Information Processing and Cybernetics* , pp. 13–36, 1987.
10. H. Wechsler and G. L. Zimmerman, "2-D invariant object recognition using distributed associative memory," *IEEE Trans. Pattern Analysis and Machine Intelligence* **10**(6), pp. 811–821, 1988.
11. F. Lin and R. D. Brandt, "Towards absolute invariants of images under translation, rotation, and dilation," *Pattern Recognition Letters* **14**(5), pp. 369–379, 1993.
12. M. Ferraro and R. M. Caelli, "Lie transform groups, integral transforms, and invariant pattern recognition," *Spatial Vision* **8**(1), pp. 33–44, 1994.
13. J. J. K. O'Ruanaidh and T. Pun, "Rotation, scale and translation invariant spread spectrum digital image watermarking," *Signal Processing* **66**(3), pp. 303–317, 1998.
14. P. Bas, J.-M. Chassery, and F. Davoine, "A geometrical and frequential watermarking scheme using similarities," in *SPIE Conf. on Security and Watermarking of Multimedia Content*, vol. SPIE 3657, pp. 264–272, 1999.
15. M. J. J. J. B. Maes and C. W. A. M. van Overveld, "Digital watermarking by geometric warping," in *IEEE Int. Conf. on Image Processing*, vol. 2, pp. 424–426, 1998.
16. P. M. J. Rongen, M. J. J. J. B. Maes, and C. W. A. M. van Overveld, "Digital image watermarking by salient point modification practical results," in *SPIE Conf on Security and Watermarking of Multimedia Content*, vol. SPIE 3657, pp. 273–282, 1999.
17. I. J. Cox, M. L. Miller, and A. McKellips, "Watermarking as communications with side information," *Proceedings of the IEEE* **87**(7), pp. 1127–1141, 1999.
18. C.-Y. Lin and S.-F. Chang, "Distortion modeling and invariant extraction for digital image print-and-scan process," *Intl. Symp. on Multimedia Information Processing*, http://www.ctr.columbia.edu/~cylin/publications.html, 1999.
19. G. Depovere, T. Kalker, and J.-P. Linnartz, "Improved watermark detection using filtering before correlation," in *IEEE Int. Conf. on Image Processing*, vol. 1, pp. 430–434, (Chicago), Oct. 1998.
20. Corel Stock Photo Library 3, Corel Corporation, Ontario, Canada.
21. E. De Castro and C. Morandi, "Registration of translated and rotated images using finite Fourier transforms," *IEEE Trans. Pattern Analysis and Machine Intelligence* **9**, pp. 700–703, 1987.
22. M. McGuire, "An image registration technique for recovering rotation, scale and translation parameters," *NEC Research Institute Technical Report* **TR 98-018**, 1998.
23. M. L. Miller and J. A. Bloom, "Computing the probablity of false watermark detection," in *Proceedings of the Thrid International Workshop on Information Hiding*, 1999.
24. R. N. Bracewell, *The Fourier Transform and Its Applications*, McGraw-Hill, 1986.