A SPATIAL MASKING TECHNIQUE FOR OPTIMAL FREQUENCY-BASED IMAGE WATERMARKING

P. Martins, P. Carvalho
Centre for Informatics and Systems
University of Coimbra
Pólo II, Pinhal de Marrocos,
3030-290 Coimbra, Portugal
{pjmm,carvalho}@dei.uc.pt

Abstract

This paper describes an adaptive spatial masking technique for frequency-based image watermarking schemes. The strategy consists of defining minimum and maximum allowable pixel perturbations. These minimum pixel perturbations, defined as constraints of an optimal watermark embedding problem, are obtained from measuring the effects in terms of inaccuracy of the most damaging interpolation methods over images. By including these perturbations, the robustness of watermarking schemes is improved: the watermark strength in images regions which are less vulnerable to distortions is increased.
The effectiveness of the proposed scheme is illustrated with experiments made on two watermarking algorithms where the described masking technique was applied.

Key Words
Watermarking, spatial masking, minimum pixel perturbations

1 Introduction

Digital watermarking consists of hiding information into digital content such as videos, images, audio or text; and later, extract or simply detect the hidden data. Some of the basic requirements on a digital watermarking scheme are imperceptibility and robustness to attacks [1]. The first requirement means that a human cannot distinguish the existent differences between the data and the watermarked data, while the latter one is described as the ability to maintain or recover the watermark, despite being manipulated by attacks such as geometrical transformations, lossy compression, or common signal processing operations. In order to ensure robustness, the watermark information is usually redundantly distributed over many samples (pixels in image and video) of the host data, however maintaining imperceptibility.

Several image watermarking algorithms based on perceptual models of the Human Visual System (HVS) have been proposed in order to provide the best trade-off between robustness and imperceptibility, i.e., to give the maximum watermark strength without making it perceptible to the human eye. Most of these methods rely on the use of contrast thresholds that are the measure of sensitivity of the HVS for different spatial frequencies [2][3][4][5][6]. For instance, in [6] is defined a new contrast measure adapted to natural scene images, called *isotropic local contrast*, based on the work of Peli [7]. The contrast masking of the HVS is then modeled according to isotropic local contrast by means of visual experiments and the strength of the watermark is adjusted according to the defined contrast masking model, during the embedding stage.

Besides defining the levels of imperceptibility, it is important to establish which image regions are less sensitive to distortions caused by common watermark attacks in order to improve the robustness of the scheme. An interesting video watermarking strategy exploiting this issue was suggested in [8]; the watermark was locally embedded into sub-frames which exhibited lower expected distortions. These regions were obtained by computing the average value of distortion bound over each sub-frame, i.e., the maximum absolute value difference between the intensity of a pixel and the intensity of the pixels in the nearest neighbourhood. Obviously, sub-frames which exhibited lower average values were the selected ones.

The spatial masking strategy described in this paper is a variation of the techniques presented in [9] and [10]. Besides the definition of maximum pixel perturbations, it takes into account which regions are less affected by distortions. Regions where interpolation reveals to be more accurate, it will be imposed a higher watermark strength, however satisfying the imperceptibility constraints defined for those regions. The main idea behind this approach is to provide more robustness to attacks such as interpolation.

The remainder of the paper is organized as follows: Section 2 describes the proposed spatial masking method. In Sections 3 and 4 are described watermarking algorithms where the spatial adaptive masking was applied. Section 5

shows some experimental results and, finally, in Section 6, conclusions and future research directions are given.

2 Proposed spatial masking

Based on a Stationary Generalized Gaussian Image Model with an auto-covariance function $R_x = \sigma_x I$, which takes into account local features of the image, Voloshynovskiy et al. [11] introduced the Noise Visibility Function (NVF), which was the starting point for a spatial masking technique proposed by the same authors. In their method, the NVF of an image $I(k,l)$ was computed at each pixel position:

$$NVF(k,l) = \frac{w(k,l)}{w(k,l) + \sigma_x^2}, \tag{1}$$

where $w(k,l) = \gamma[\eta(\gamma)]^\gamma \frac{1}{\|r(k,l)\|^{2-\gamma}}$, $r(k,l) = \frac{I(k,l)-\overline{I}(k,l)}{\sigma_x}$, $\eta(\gamma) = \sqrt{\frac{\Gamma(\frac{3}{\gamma})}{\Gamma(\frac{1}{\gamma})}}$ and $\Gamma(t) = \int_0^{+\infty} e^{-u}u^{t-1}du$ is the *gamma function*. Once the Noise Visibility Function had been computed, the *maximum pixel perturbation* was obtained for each pixel:

$$\begin{aligned}\Delta_p(k,l) = \quad & (1 + K \cdot CST(I(k,l))) \cdot (1 - NVF(k,l)) \cdot S_0 \\ & + NVF(k,l) \cdot S_1,\end{aligned} \tag{2}$$

where S_0 and S_1 were real positive scalars defining the maximum pixel perturbations in textured and flat regions, respectively, K was a positive scalar and CST was the *contrast sensitivity threshold* at the luminance value $I(k,l)$, given in terms of the change of luminance divided by the luminance as defined in [9]. Since distortions are less visible in textured areas, and the NVF tends to 0 in these regions, while in flat regions it tends to 1; S_0 should be higher than S_1. In [9], S_0 is about 30, while $S1$ is about 3, and K is 5, which means that the allowable distortion is increased by 5 in textured regions where the luminance level is high. Thus, given a watermarked version $I_w(k,l)$ of an original image $I(k,l)$, its watermark is considered imperceptible if inequality (3) holds.

$$|I_w(k,l) - I(k,l)| \leq \Delta_p(k,l) \tag{3}$$

Besides the imperceptibility constraints (3), it is possible to define minimum perturbations that will increase the robustness of the embedded watermark. The choice of these perturbations is based on the fact that image regions with small spatial gradients (where interpolation is more accurate), the amount of distortion available to perform attacks is reduced [8].

Let $I(k,l)$, $1 \leq k \leq m$, $1 \leq l \leq n$, be an 8-bit image or the luminance channel of a RGB image. The *nearest neighbouring* of pixel (i,j) is defined as follows:

$$\mathcal{N}(i,j) = \{(k,l) : |k-i| \leq 1, |l-j| \leq 1, (k,l) \neq (i,j)\} \tag{4}$$

In order to evaluate the effects of interpolation's inaccuracy, like in [8], the maximum distance between values of nearest neighbouring pixels is computed at each pixel position:

$$Y(i,j) = \max_{(k,l)\in\mathcal{N}(i,j)} |I(k,l) - I(i,j)| \tag{5}$$

We decided to measure the effects of interpolation from (5) because when an intensity value $I(i,j)$ is replaced by another value $\hat{I}(i,j)$ through bilinear interpolation or nearest neighbour interpolation – which are the most damaging methods, namely due to aliasing [12] –, we have $\left|\hat{I}(i,j) - I(i,j)\right| \leq Y(i,j)$. Therefore, the regions where $Y(k,l)$ exhibits lower values correspond to the regions where interpolation is more accurate, while the ones where $Y(k,l)$ exhibits larger values, indicate that they are more affected by the errors of interpolation methods and distortion attacks, despite being the regions where the level of the watermark imperceptibility is higher, as depicted in Figure 1. This result suggests that the minimum pixel perturbation should be higher in regions where $Y(k,l)$ is lower. In order to obtain these perturbations, we define (6).

$$\tilde{Y}(i,j) = \frac{Y(i,j)}{255} \tag{6}$$

The minimum pixel perturbation at pixel (i,j) will be

$$\Delta_q(i,j) = (1 - \tilde{Y}(i,j)) \cdot S_2, \tag{7}$$

where S_2 is a real positive scalar, slightly smaller than the factor S_1 defined in (3). By imposing $S_2 \leq S_1$, we will have $\Delta_q(i,j) \leq \Delta_p(i,j)$, since $\Delta_q(i,j) \leq S_2$ and $\Delta_p(i,j) \geq S_1$.

The minimum pixel perturbation is added to the constraint (2), yielding (8).

$$\Delta_q(k,l) \leq |I_w(k,l) - I(k,l)| \leq \Delta_p(k,l) \tag{8}$$

Similarly to the schemes proposed in [9] and [10], the robustness of the watermark can be increased by solving a constrained optimization problem whose optimal solution is the maximum strength value satisfying the constraints defined in (8). Therefore, the constrained optimization problem to solve is:

$$\begin{aligned}&\textit{maximize } \alpha \\ &\textit{subject to} \\ &|I_w(k,l) - I(k,l)| \leq \Delta_p(k,l) \\ &|I_w(k,l) - I(k,l)| \geq \Delta_q(k,l) \\ &0 \leq I_w(k,l) \leq 255 \\ &\alpha_l \leq \alpha \leq \alpha_u,\end{aligned} \tag{9}$$

where α is the strength of the watermark and α_l, α_u are the minimum and maximum values of α, respectively.
Unlike the optimization problems formulated in [9] and [10], this one cannot be solved by the Simplex method,

since the inclusion of the minimum pixel perturbations constraints – which are non-convex –, makes impracticable to treat the problem as a linear programming one. Consequently, the problem is solved as a non-convex non-linear constrained optimization problem.

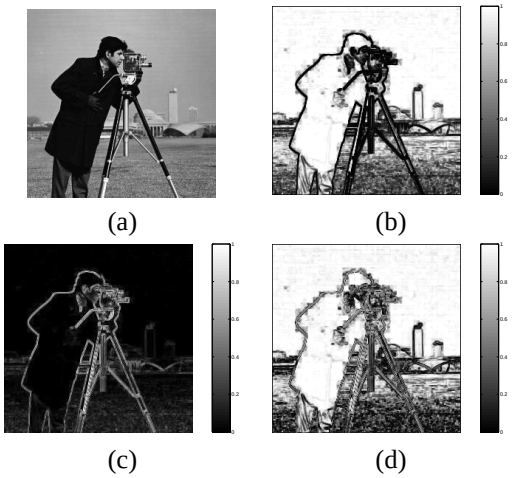


Figure 1. (a) original image; (b) NVF of (a); (c) maximum distance between values of nearest neighbouring pixels in (a); (d) absolute difference between (b) and (c).

3 Algorithm I

The watermarking strategy described in this section is a straightforward additive frequency-domain scheme and it intends to illustrate how the spatial masking suggested in the previous Section can be applied to methods which embed the watermark into a transform domain. A canonical scale at embedding and detection stages will be used in order to exemplify the effects of interpolation over the watermark detection.

3.1 Watermark embedding

Given an image $I(k,l)$, $1 \leq k \leq m$, $1 \leq l \leq n$, the watermark embedding stage includes into the following steps:

1. Resize $I(k,l)$ to the canonical size $p \times p$, yielding a new image $I_{Can}(k,l)$.
2. Compute $\Delta_p(k,l)$ and $\Delta_q(k,l)$, the maximum and minimum pixel perturbation of $I_{Can}(k,l)$, respectively.
3. Compute the Discrete Fourier Transform (DFT) of $I_{Can}(k,l)$. Let $F(u,v)$ be the obtained result.
4. Extract $M(u,v)$, the magnitude of $F(u,v)$.
5. Generate a bimodal pseudo-random sequence $w = \{w_i : w_i \in \{-1,1\}, i = 1, \dots L\}$ from a key k.
6. Pseudo-randomly select L coefficients satisfying (10), based on the key k.

$$f_1 \leq \sqrt{u^2 + v^2} \leq f_2 \tag{10}$$

7. Obtain $M_w(u,v)$, the modified magnitude after the watermark embedding. To embed the watermark, each selected magnitude coefficient is changed such that:

$$M_w(u_i,v_i) = M(u_i,v_i)(1 + \alpha_i w_i), i = 1 \cdots L, \tag{11}$$

$$M_w(-u_i,-v_i) = M_w(u_i,v_i), i = 1 \cdots L, \tag{12}$$

where α_i is the *i-th* component of the optimal solution of the minimization constrained problem (13), which is an adaptation of (9) to this algorithm.

$$\begin{aligned} &\textit{minimize } -\textstyle\sum_{i=1}^{L} \alpha_i \\ &\textit{subject to} \\ &|IDFT(F_w(u,v)) - I_{Can}(k,l)| \leq \Delta_p(k,l) \\ &|IDFT(F_w(u,v)) - I_{Can}(k,l)| \geq \Delta_q(k,l) \\ &0 \leq IDFT(F_w(u,v)) \leq 255 \\ &0 \leq \alpha_i \leq 1 \end{aligned} \tag{13}$$

8. Compute the Inverse Discrete Fourier Transform (IDFT) of $F_w(u,v)$. Let the result be denoted as $I_{Can_w}(k,l)$.
9. Resize $I_{Can_w}(k,l)$ to a $m \times n$ image, yielding $I_w(k,l)$, the watermarked version of $I(k,l)$.

3.2 Watermark detection

Given an image $I'(k,l)$, $1 \leq k \leq m'$, $1 \leq l \leq n'$, the detection process is as follows:

1. Resize $I'(k,l)$ the canonical size $p \times p$. Let $I'_{Can}(k,l)$ be the result image.
2. Compute $F'(u,v)$, the DFT of $I'_{Can}(k,l)$.
3. Extract $M'(u,v)$, the magnitude of $F'(u,v)$.
4. Generate a pseudo-random sequence $w = \{w_i : w_i \in \{-1,1\}, i = 1, \dots L\}$ from a key k.
5. Pseudo-randomly select L coefficients satisfying (10), based on the key k.
6. Compute the similarity coefficient (14) between w and M and compare it to a threshold T in order to minimize both false alarms and false rejections.

$$S(w,M) = \frac{\sum_{i=1}^{L} w_i M(u_i,v_i)}{\sum_{i=1}^{L} M(u_i,v_i)^2} \tag{14}$$

If $S(w,M)$ is greater than T, the detector reports that the watermark is present.

4 Algorithm II

The algorithm described herein is based on the scheme proposed by [13]. It intends to provide robustness to geometrical distortions. The resilience to this kind of attacks is achieved by inserting the watermark into a domain which is rotation, scaling and translation (RST) invariant. To obtain a RST invariant domain, the following steps are performed:

1. Compute the log-polar mapping (LPM) of the image with its centroid as the origin.
2. Apply a 2-D Discrete Fourier Transform to the log-polar version of the image.
3. Extract the magnitude of the Fourier Transform which is invariant to rotation, scaling and translation.

Besides the watermark strength being set adaptively, the main difference between this method and the one proposed by [13], relies on the point which is selected as the origin of the LPM. In [13], it is proposed to perform the log-polar mapping with the centroid of a circular region as the origin, instead of the whole image region. This centroid is obtained by calculating first the image centroid G_0. Then, the centroid of a circular region with radius r and center G_0 is computed. The new point G_1 is used as the origin of a circular region with radius r and the centroid G_2 of this region is computed. G_1 is compared to G_2; if they coincide, the process stops, otherwise, it is repeated until they converge on the same point. The aim of this strategy is to have an invariant point as the origin of log-polar coordinates system, despite the geometrical distortions that the image might have suffered. However, experiments showed that this method is time-consuming and it is unstable when the image suffers attacks like cropping. So, the image centroid is used instead. It exhibits a more than reasonable stability when dealing with some geometric distortions

Due to lack of space, we refer to [13] for a more detailed description of the properties of the Log-Polar mapping and the Discrete Fourier Transform, and how they can be applied to obtain a RST invariant domain.

4.1 Watermark embedding

The embedding stage consists of obtaining a RST invariant domain and embed the watermark into it. Given the image to mark, the embedding process starts by computing the image centroid. To provide some robustness to lossy compression and common image processing distortions, the centroid is computed over a low-pass filtered version of the image. By doing this, the centroid becomes invariant to some distortions created by these operations. The next step is to convert the image into its LPM version such as the image maximum and minimum pixel perturbations, with the image centroid as the origin. At this stage, any rotation on Cartesian coordinates is converted into a cyclic shift. The LPM is then followed by a 2-D Fourier Transform. Its magnitude is taken to make a domain which is also invariant to rotations. Then, a watermark generated from the copyright owner's key k as described in (15) is embedded into the RST invariant domain.

$$w = \{w_i : w_i \in \{-1, 1\}, i = 1, \dots l\} \tag{15}$$

The l magnitude coefficients, where l is the watermark length, satisfying

$$f_1 \leq \sqrt{\rho_i^2 + \theta_i^2} \leq f_2 \tag{16}$$

are selected to embed the watermark. The selection of these coefficients is based on the copyright owner's key k which generates a permutation of the coefficients satisfying (16) and takes the first l coefficients from the permutation vector. To embed the watermark, the selected magnitude coefficients $M(\rho_i, \theta_i)$ are changed such that:

$$M_w(\rho_i, \theta_i) = M_w(-\rho_i, -\theta_i) = M(\rho_i, \theta_i)(1 + \alpha_i w_i), \tag{17}$$

where α_i is the is the *i-th* component of the optimal solution of the minimization constrained problem adapted from (9) to this particular algorithm. Thus, given $F_w(\rho', \theta')$, the DFT of the LPM of the marked image; $LPMI(\rho, \theta)$, the original image LPM; $LPM\Delta_p(\rho, \theta)$ and $LPM\Delta_q(\rho, \theta)$, the LPM version of the maximum and minimum perturbations, respectively; the optimization problem to solve is:

$$\begin{aligned}
&\textit{minimize } -\textstyle\sum_{i=1}^{l} \alpha_i \\
&\textit{subject to} \\
&|IDFT(F_w(\rho', \theta')) - LPMI(\rho, \theta)| \leq LPM\Delta_p(\rho, \theta) \\
&|IDFT(F_w(\rho', \theta')) - LPMI(\rho, \theta)| \geq LPM\Delta_q(\rho, \theta) \\
&0 \leq IDFT(F_w(\rho', \theta')) \leq 255 \\
&0 \leq \alpha_i \leq 1
\end{aligned} \tag{18}$$

Once the watermark has been computed, the inverse operations are performed. However, the inverse LPM (converting log-polar map coordinates into Cartesian coordinates), such as the LPM, induces the inevitable interpolation errors, leading to a loss of image quality. So, to avoid this situation, the ILPM is only applied to the watermark signal and finally added to the given image.

4.2 Watermark detection

The detection scheme has the same initial operations as the embedding scheme. Thus, a low-pass filter is first applied to the image and then, its centroid is computed. The second step consists of computing the LPM of the image. Then, the 2-D DFT is applied to the image LPM and its magnitude is extracted. The watermark sequence is then generated from the copyright owner's key k. A set of selected coefficients satisfying (16), is selected according to the strategy described in the embedding scheme and, finally, the correlation coefficient (19) is computed and compared to a threshold T in order to minimize both false alarms and

false rejections.

$$\rho(w, M) = \frac{\sum_{i=1}^{l}(w_i - \overline{w})(M(\rho_i, \theta_i) - \overline{M})}{\sum_{i=1}^{l}(w_i - \overline{w})^2 \sum_{i=1}^{l}(M(\rho_i, \theta_i) - \overline{M})^2} \quad (19)$$

Thus, if $\rho(w, M)$ is greater than T, the detector reports that the watermark is present.

5 Experimental Results

To test the described watermarking scheme in Section 3 and, consequently, the proposed spatial masking, the image "Cameraman" (a gray-scale image of size 256 by 256) was watermarked and some attacks were performed over it. Two watermarked versions of the image were created: one was the result of an embedding procedure using the proposed spatial masking, while the other one, was the result of embedding the same watermark, but ignoring the minimum pixel perturbations during the perceptual mask analysis. Figure 2 depicts the original image and its watermarked version using algorithm I and the proposed spatial masking technique. On these experiments, the watermark length

(a) (b)

Figure 2. (a) original image; (b) watermarked image using algorithm I.

was set to 200 and the canonical size was set to 200 by 200 pixels. The attacks included scaling (from a scale factor of 0.1 to 2.5, using nearest neighbour, bilinear and bicubic interpolation methods), lossy compression. According to Figures 4 and 3, it is possible to conclude the robustness of the scheme to these attacks. Concerning JPEG compression, the watermark was successfully detected down to a quality factor of 15%, and it was able to survive to a wide range of scale changes. When compared to the results obtained with the same algorithm, but ignoring the minimum pixel perturbations, it is readily seen that the introduction of the aforementioned perturbations increased the robustness of the scheme.

The proposed spatial masking technique was also tested on algorithm II. In this case, the image "Peppers" (a 200 by 200 RBG image) was watermarked. Like in the previous tests, two watermarked versions of the test image were created. Figure 5 depicts the original image and its watermarked version using algorithm II and the proposed spatial masking technique. In these last experiments, the watermark length was set to 500 and the sampling rates on angular and radial direction were both set to 300. The attacks included only rotation. Figure 6 depicts the detection performance of the proposed solution under rotation attacks. The increase of robustness to rotation attacks, by adding minimum pixel perturbations, is not so meaningful. However, the successive loss of information due to the conversion to LPM and its inverse operation, during the embedding stage of algorithm II, explains the observed results.

6 Conclusions and future work

In this paper was proposed the introduction of minimum pixel perturbations constraints into a spatial masking technique based on the Noise Visibility Function. These minimum pixel perturbations are a result of measuring the effects of distortions on image regions: the ones which are more resilient to distortions will have higher minimum pixel perturbations, while regions which are more vulnerable to distortions, will have lower pixel minimum pixel perturbations. By defining these minimum pixel perturbations and deriving the maximum pixel perturbations from the Noise Visibility Function, it is possible to develop an optimal embedding watermarking scheme, where regions which are less affected by distortions can be more perturbed after the watermark insertion, however maintaining the good levels of imperceptibility imposed by the maximum pixel perturbations.

According to the experiments, the inclusion of minimum pixel perturbations improves the robustness of the described watermarking schemes against attacks such as rotation, scaling and lossy compression.

Further research directions include applying the proposed masking strategy to a DCT embedding technique and to exploit some methods related to constrained optimization in order to improve the effectiveness of the optimization problem, since tends to exhibit a high computational cost and some solutions are sub-optimal.

References

[1] S. Voloshynovskiy, S. Pereira, T. Pun, J. J. Eggers, & J. K. Su, Attacks on Digital Watermarks: Classification, Estimation-Based Attacks, and Benchmarks, *IEEE Communications Magazine, 39*(8), 2001, 118–127.

[2] C. I. Podilchuk & W. Zeng, Image-adaptive watermarking using visual models, *IEEE Journal of Selected Areas in Communications, 16*(4), 1998, 525–539.

[3] D. Kundur & D. Hatzinakos, A Robust Digital Image Watermarking Method using Wavelet-Based Fusion, *IEEE Signal Processing Society 1997 International Conference on Image Processing (ICIP'97)*, 1997.

[4] M. D. Swanson, B. Zhu, & A. H. Tewfik, Transparent Robust Image Watermarking, *1996 SPIE Conf. on Visual Communications and Image Proc., 3*, 1996, 211–214.

[5] R. B. Wolfgang, C. I. Podilchuk, & E. J. Delp, Perceptual Watermarks for Image and Video, *Proceedings of the IEEE, 87*(7), 1999, 1108–1126.

[6] M. Kutter & S. Winkler, A vison-based Masking Model for Spread-Spectrum Image Watermarking, *IEEE Transactions on Image Processing, 11*(1), 2002, 16–25.

[7] E. Peli, Contrast in Complex Images, *Journal of Optical Society of America, 7*(10), 1990.

[8] K. Su, D. Kundur, & D. Hatzinakos, A content-dependent spatially localized video watermark for resistance to collusion and interpolation attacks, *Proc. IEEE Int. Conf. on Image Processing, 1*, 2001, 818–821.

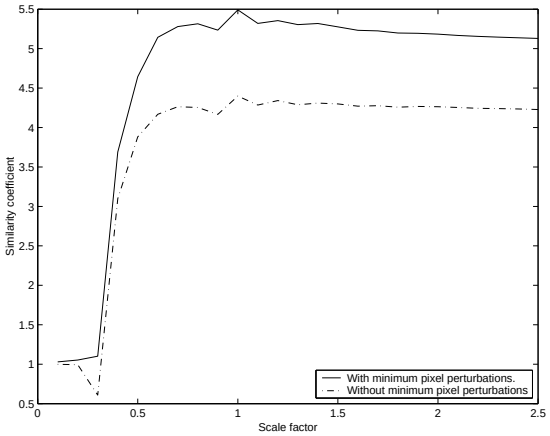
[9] S. Pereira, S. Voloshynovskiy, & T. Pun,Optimal transform domain watermark embedding via linear programming, *Signal Processing, Special Issue: Information Theoretic Issues in Digital Watermarking*, 2001.

[10] A. Santos, L. M. e Silva, P. Martins, & P. Carvalho, Frequency-based Watermarking with Spatial Masking for DRM of MMS Content, *Proc. of the Int. Conf. on Image and Signal Processing*, Honolulu, Hawaii, USA, 2003, 39–43.

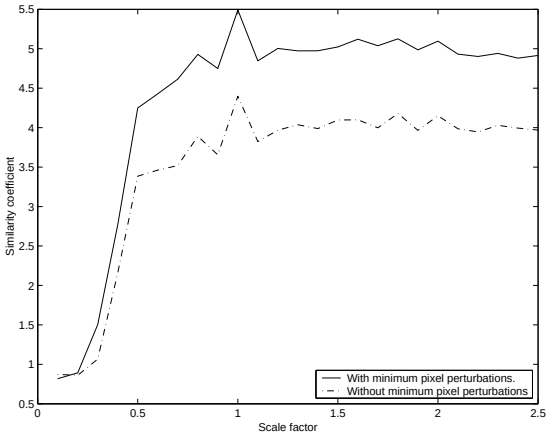
[11] S. Voloshyhovskiy, A. Herrigel, N. Baumgartner, & T. Pun, A stochastic approach to content adaptive digital watermarking, *International Workshop on Information Hiding*, Dresden, Germany, 1999, 212–236.

[12] T. M. Lehmann, C. Gönner, & K. Spitzer, Survey: Interpolation Methods in Medical Image Processing, *IEEE Transactions on Medical Imaging, 18*(11), 1999, 1049–1075.

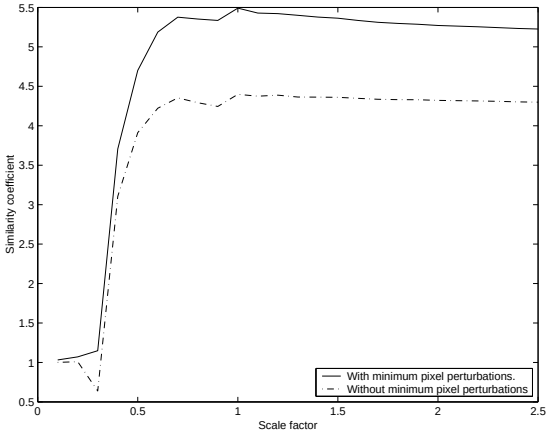
[13] B.-S. Kim, J.-G. Choi, C.-H. Park, J.-U. Won, D.-M. Kwak, S.-K. Oh, C.-R. Koh, & K.-H. Park, Robust Digital Image Watermarking against , Geometrical Attacks, *Real-Time Imaging, 9*, 2003, 139–149.



(a)



(b)



(c)

Figure 3. Similarity coefficients for several scaled versions of the watermarked image, using: (a) bilinear interpolation; (b) nearest neighbour interpolation; (c) bicubic interpolation.

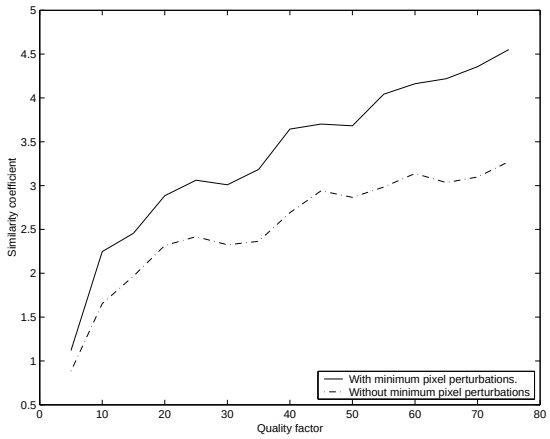


Figure 4. Similarity coefficients for several JPEG quality factors.

(a) (b)

Figure 5. (a) original image; (b) watermarked image using algorithm II.

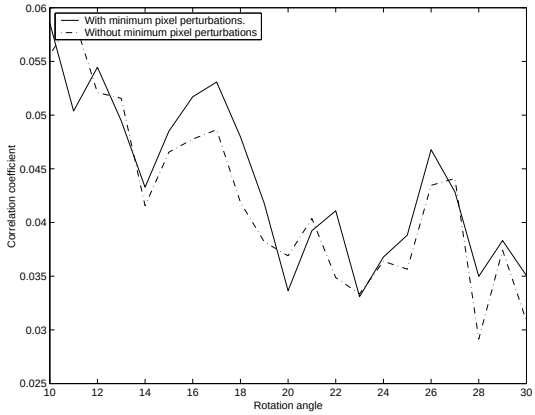


Figure 6. Correlation coefficient for several rotation angles.