Multimed Tools Appl (2017) 76:10119–10131
DOI 10.1007/s11042-016-3601-1

Blind image counterwatermarking – hidden data filter

Zbigniew Piotrowski[1] · Piotr Lenarczyk[1]

Received: 31 July 2015 / Revised: 6 March 2016 / Accepted: 9 May 2016 /
Published online: 9 June 2016

Abstract Watermarking is a dynamically developing method of copyright protection used for media (sounds, images, films, or 3D objects) that employs signal processing in order to hide additional, invisible information about the owner or author. However, studies until now have not widely considered the problem of attempting to blind removal of hidden data in a manner that allows the watermarked signal to be returned to the original signal. Current considerations of authors of leading articles in the watemarking field focus on the robustness of the method – security system against intentional attacks of removal of the additional information, without taking into account the aspect of simultaneous degradation of the quality and form of the watermarked picture. As has been shown in the article, it is possible to design a perfect filter (same as reversible operations) that allows the removal of additional information from the watermarked picture in a way that makes it possible to return to the form of the host image. This paper describes an ideal filter used for removal of additional, invisible information in forms of an eliminating function and a signal masking function. Their effectiveness has been demonstrated for practical implementations of this type of eliminating and masking filters for watermarking methods in the cepstrum domain.

Keywords Hidden data filter · Watermarking · Image watermarking · Counterwatermarking

1 Introduction

The problem of a hidden data filter has been merely demonstrated [5] in the form of problems of an attack aimed at elimination of the watermark and the masking. In [4, 17] such considerations are not found, while in [2, 16] the problem of hidden data filter is estimated in a manner similar to the masking described in [5]. In the book [1] the authors provide examples of elimination of the watermark from watermarked pictures on the basis of a

✉ Zbigniew Piotrowski
zpiotrowski@wat.edu.pl

[1] Military University of Technology, gen. S. Kaliskiego 2 st., 00-908 Warsaw, Poland

collusion or oracle attack. However, in each of these cases the examples are those of a masking attack, not the perfect removal of a watermark from the watermarked picture, making it possible to return to the host image. An additional element that absolutely must be taken into account is the fact that the perfect hidden data filter must be separated from the robustness of the method which, according to [9, 16], cannot be based on the properties of the algorithm, or access to the watermarking application, but on the general possibilities of transformation of the watermarked picture, both within the space and frequency domains.

Elements of professionally prepared attacks, testing the robustness of watermarking methods have been included in the Stir Mark [8, 14] application. It includes a wide range of methods for potentially preventing the detection of a watermark in a picture containing additional, invisible information. Testing of the possibilities of the watermarking method is performed automatically using a prepared protocol.

However, in literature it is impossible to find algorithms that eliminate the watermark signal from the watermarked signal under the condition of returning to the original signal with high signal quality maintained. The article presents in its first part a theoretical model of the perfect hidden communication filter, an accurate description of both functions – the one eliminating and the one masking the watermark signal, the practical implementation of the hidden communication filter and efficiency tests results.

2 Perfect hidden data filter

2.1 Perfect hidden transmission filter

The description of the perfect hidden communication filter begins with a specification of the communication channel used in the watermarking. Let us consider O as one of the set of all original signals (images, videos, sounds, texts, 3D objects, etc.), W – as one of the set of all watermarks containing information i, K – as one of the set of all keys used in watermarking (not all watermarking applications require K keys). Using this notation, it is possible to describe the coding process E_{wmf} which produces all possible watermarked signals O_{wm} and the watermark decoding D_{wmf} as two functions:

$$E_{wmf} \ : O \times K \times W \rightarrow O_{wm} \tag{1}$$

$$O_{all} \ni O \ , O_{wm} \ , O'_{wm} \tag{2}$$

$$D_{wmf} : \ O_{all} \times K \rightarrow empty \tag{3}$$

$$D_{wmf} : \ O'_{wm} \times K \rightarrow W \tag{4}$$

O_{all} designates one from all possible decoder inputs (in this case tested signals are rejected). It is described in third equation,

O'_{wm} designates one from all signals watermarked after passing through the communication channel, taking into account possible attacks on the signal – both intentional and unintentional.

Within the decoding function, the supplementary information added to the original signal is found through a comparator function $C_{c\tau}$ that compares the recovered form of the watermark from the signal marked with a particular watermark in relation to the a'priori designated decision threshold τ, which gives a response whether additional information is to be found in the signal or not. Only after a positive comparison in the comparator is information i retrieved from the watermark signal W.

$$C_{c\tau}:\ O'_{wm}\rightarrow\{0,1\} \tag{5}$$

If we take into account that in the communication channel the form of the watermarked signal O_{wm} can be changed intentionally or not. However, in both cases, conversion of it into signal O'_{wm} is performed.

The ideal watermark signal elimination function F_{c-wm} is able to recover original signals O' deformed in the communication channel to their approximate original form despite deformation in the communication channel of the watermarked signals O'_{wm}. In the general case, the purpose of the Filtering Block B_{Fc-wm} is to eliminate the watermark signal W from the deformed watermarked signal O'_{wm}.

$$F_{c-wm}:O'_{wm}\times B_{Fc-wm}\rightarrow O' \tag{6}$$

One must take into account that the problem of eliminating the watermark signal through function F_{c-wm} is a different approach to the removal of the watermark signal W, from masking the watermark signal using a masking function M_{c-wm}. Briefly F_{c-wm} can be descripted as an ideal masking function M_{c-wm}. This function removes part of the watermark W, in the ideal case:

$$F_{c-wm}\ni M_{c-wm} \tag{7}$$

However, the function M_{c-wm} is characterised by a partial removal of watermark W from the watermarked signal O'_{wm}, while simultaneously causing degradation O'_{deg} of the target form of signal O'. This stems from the fact that M_{c-wm} interferes with part of the signal O'_{wm} in the process of masking part W. An example of such a type of function and its practical implementation is presented later in the paper.

$$M_{c-wm}:O'_{wm}\times B_{Mc-wm}\rightarrow O'_{deg} \tag{8}$$

2.2 Elimination vs. masking

In the case of deformation of the form of signal O_{wm} into form O'_{wm}, the decoding function D_{wmf} must have the ability to perform watermark detection W but the crucial element is to recover information i, which, de facto, when used in watermarking, usually contains the copyright data for a particular piece of media. In the case of watermark masking a deliberate attack is most likely, aimed at preventing decoding of information i

or watermark detection W. A similar dependency occurs in the case of the eliminating function, but both i and W must be removed, while image $O_{wm}^{'}$ must return to the form $O^{'}$, a condition that is not critical for function M_{c-wm}. Existing considerations in literature [1, 2, 4, 16, 17] have not concerned function F_{c-wm} but only the masking function. In the latter case we find these types of algorithms based on the following boundary conditions for the person conducting the masking:

- access to the signal,

- access only to the watermarked signal – most likely case,
- access to watermarked signals and corresponding information i,
- access to watermarked signals and corresponding original signals (with the goal being to recover information i),

- access to the encoder,
- access to the decoder,
- access to the watermarking algorithm – both E_{wmf} and D_{wmf} functions.

In literature it is possible to find examples of masking filtering, such as [12], where the authors propose a masking method for decoding watermarking algorithms based on spectral dispersion using non-linear filtering, estimating the watermark in the watermarked signal. In the first part, they filter the watermarked signal using a 3x3-sized median filter. Then they subtract from the watermarked signal the difference between the watermarked signal and the median-filtered signal; the difference has been once again high-pass filtered and empirically scaled on the basis of a determined coefficient. Thanks to this method they estimate the watermark signal and can successfully mask it.

As an example of eliminating filtration it is possible to give an example of collusion processing, where the attacker has only the watermarked signals but in this case it is necessary to satisfy the condition of partial uniformity (masking), or total uniformity (elimination) of the watermark signal W. For this type of filtration only watermarked signals are required, which means that it is a blind method. The process will be based on averaging the watermarked signals – the watermark signal will become clear from among the noise of random values of the remaining averaged samples of watermarked signals. For a watermark signal processed in such a way there is nothing else to do other than to remove the recurring samples of the signal of the watermark W by means of subtraction. This attack applies to watermarking methods in which the watermark signal W added to the original signal O is not its function. The effectiveness of this type of elimination filtration has been shown in [7] for algorithms for watermarking films.

In the case of the person conducting the elimination or masking the watermark signal having access to the watermark encoder, it is possible to perform effective masking of the watermark signal (it should be noted that the person conducting the attack does not have to have the physical encoder, just temporary access to it will enable that person to watermark their own signal, or a couple of original signals), and in a special case – to eliminate the watermark signal. This applies especially to algorithms that use the entire space of the original signal, or, for example, in solid blocks, as acquisition of the watermarked signal makes it possible to determine the spatial or spectral range in which the encoder is operating and establish a solid filtering matrix. Adaptive methods are more resistant to such attacks, where a larger collection of original signals and their corresponding watermarked signals are needed for generalized determinations.

Another particular example of a masking algorithm is described in [11], where the authors prove that in the case of access to a single watermarked signal and a decoder, it is possible to recover a part of the original signal. In this case they use pseudo-linear dependencies used in the detector and it is possible to recover the original signal for a watermark signal without the DC component and within the range of values $\{-1, 1\}$. However, it should be noted that the attack in this case is against the watermarking algorithm for broadcast applications, where each user has access to the watermark decoder.

In the case of [10] the authors described a masking algorithm that removes the additional information from the watermarked signal, while maintaining the quality of the original signal, in this case a picture. For algorithm [6] it obtained a PSNR = 36.65 dB, with NC = 0.12, while for [13] a PSNR = 32.95 and NC = 0.28, where these and other attacked watermarking algorithms are of the non-blind type (which greatly limits their use in practical watermarking applications). It should be noted that the masking algorithm quite precisely removes the watermark signal from the watermarked signal, while maintaining good quality of the reconstructed original signal.

3 Implementation of the filter

In the case of article [15] a filter for hidden picture transmission has been implemented and tests of its effectiveness have been conducted. The function of the elimination of the watermark F_{c-wm} begins with the conversion of the picture marked O'_{wm} from RGB to YCbCr representation $O'_{wm\,YCbCr}$. Analysis is then carried out in the cepstrum domain in order to determine the translation value for the added, luminance matrix with reduced energy, translation in space. For this purpose a 2-dimensional Discrete Fourier Transform is performed on the watermarked luminance matrix Y'_{wm}:

$$\begin{aligned}
Y'_{wm\,DFT}(k,l) &= \sum_{x=0}^{X-1}\left[\sum_{y=0}^{Y-1} Y'_{wm}(x,y)\,b^{*}_{YDFT}(l,y)\right] b^{*}_{XDFT}(k,x) \\
Y'_{wm\,XYDFT} &= B^{*}_{XDFT} Y'_{wm\,XY} B^{*T}_{YDFT} \\
b_{DFT}(k,x) &= \sqrt{\frac{1}{X}}\exp\left(j\frac{2\pi k}{X}x\right) \\
b_{DFT}(l,y) &= \sqrt{\frac{1}{Y}}\exp\left(j\frac{2\pi l}{Y}y\right)
\end{aligned} \tag{9}$$

x,y – indexes of discrete spatial positions of pixels,
X,Y – spatial resolution of images,
k,l – indexes of discrete, 2D signal frequencies of the spectrum.

Then on the matrix Y'_{wmDFT} a cube of the 2-dimensional autocepstrum function is calculated:

$$Y'_{wm\ cepst}(m,n) = \left(IDFT\left(\ln\left(\left|Y'_{wm\ DFT}(k,l) =\right|\right)\right)\right)^{3} \tag{10}$$

$$Y'_{wm\ IDFT}(x,y) = \sum_{x=0}^{X-1}\left[\sum_{y=0}^{Y-1} Y'_{wm\ DFT}(k,l) b_{YDFT}(l,y)\right] b_{XDFT}(k,x) \tag{11}$$

m,n – indexes of discrete coefficients of the two-dimensional autocepstral matrix.

In the degraded watermarked picture the translation coordinates of luminance copies will correspond to the coordinates of the cepstral coefficient for which the cube of the two-dimensional autocepstral function, due to the copy of its own signal, reaches a much higher value, in accordance with [3]. Then, after crossing the decision threshold τ, the coordinates of the cepstral coefficient $Y'_{wm\,cepst}(m,n)$ will be responsible for the inverse translation values of the copy of the luminance matrix p_x, p_y, while the subtraction or addition sign will be determined by the phase $Y'_{wm\,cepst}(m,n)$:

$$Y'_{c-wm}(x,y) = Y'_{wm}(m,n) \mp Y'_{wm}\left(m+p_x, n+p_y\right)\delta \tag{12}$$

δ – watermark power coefficient, calculated empirically in [15].

Then, for the luminance matrix of the disturbed watermarked picture Y'_{wm} processed in such a way, a luminance matrix is obtained with the eliminated watermark $Y'_{c-wm}(x,y)$ which is the same as matrix $Y'(x,y)$. The last step is transforming the matrix from YCbCr to RGB, which results in the output signal O'.

A masking function M_{c-wm} has also been implemented for recovering the form of the original signal O'_{deg} degraded in the communication channel, based on the degraded watermarked signal O'_{wm} for cepstral watermarking described in [15] and algorithms that take advantage of functions modulating the added copies of the whole, or component parts of the original signal. The developed masking function uses Wiener blind deconvolution consisting of deconvolution – separation of the watermark signal (which in the general case is de facto imperceptible noise added to the original picture) from the original signal. The diagram of an ideal deconvolution is shown below Fig. 1:

In practice, it is impossible to perfectly separate two convoluted signals with a system of this nature, so a homomorphic filter is used, eliminating in two cases one of the estimated convoluted signals Fig. 2:

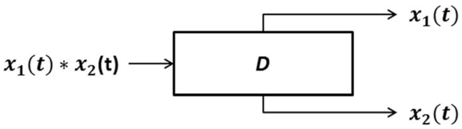


Fig. 1 Diagram of an ideal deconvolution system

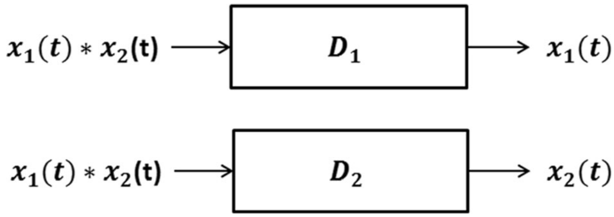


Fig. 2 Example of 2 homomorphic deconvolution filter

A diagram of the masking algorithm is shown below Fig. 3:

A blind deconvolution filter for the luminance matrix of the watermarked signal uses a likelihood maximization algorithm, as a result of which we obtain a filtered luminance matrix Y'_{wiener} and a restored estimation of the PSF (Point Spread Function – response from the image system that processes the host image into a source in the form of a spot image) used by the E_{wmf} function. In order to initiate a blind deconvolution function in the most optimal manner (sample of worst case adaptation has been shown at Fig. 11), an adaptive spatial movement filter has been used, with coefficients h calculated as follows:

- we create an empty matrix for the movement model,
- we supplement it with a vector with a length of l (f.e. 2 pixels) and angle θ $(\overset{\mathring{A}}{135})$, centered on the middle coefficient of the h filter matrix,
- for each coordinate (i,j), calculate the closest ND distance between this location of the ND and the segment of the model,
- $h = \max\ (1 - ND, 0)$,
- we then normalize the coefficient h: $h = h\ (sum\ (h\ (:)))$ Figs. 4, 5, 6, 7, 8, 9, 10, and 11.

The spatial averaging filter is a matrix with dimensions 4x4 with a value of the coefficients 0,0625. At the output of the averaging spatial filter the filtered matrix Y'_{median} is obtained which is subtracted from the degraded watermarked signal, resulting in matrix Y'_{diff} being obtained. It is an estimated watermark matrix used in the function E_{wmf} which we again subtract from the degraded watermarked signal, obtaining as a result matrix Y'_{c-wm} that is the approximate luminance matrix of the degraded original signal O'_{deg}.

4 Efficiency test results

The above-described filter for eliminating hidden communication has been implemented in practice in the Matlab programming environment and its effectiveness has been tested. The number of pictures used as original signals was 99. The experiment consisted in watermarking pictures with a hybrid watermarking algorithm in the section pertaining to the cepstrum domain

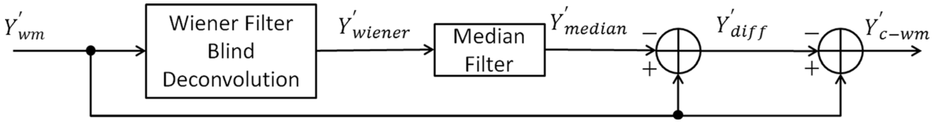
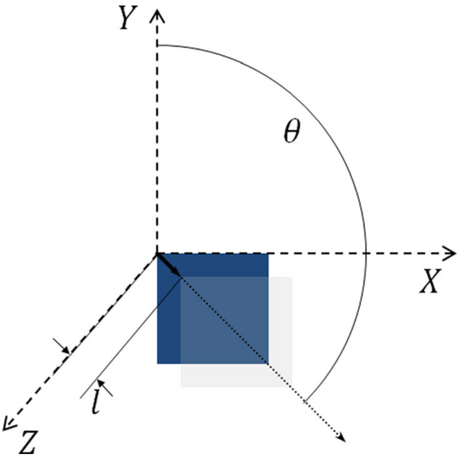


Fig. 3 The diagram of the masking algorithm

Fig. 4 Addition of translated luminance copy with reduced values of coefficients



[15], its detection based on function D_{wmf}, which is tantamount to determining the coordinates of the coefficient of the two-dimensional autocepstral function and translation values p_x, p_y of the copy of the luminance matrix of the original signal O. Then the PSNR value was measured:

$$RMS = \frac{1}{XY}\sum_{i=1}^{N}\sum_{j=1}^{M}\left[O(i,j)-O'_{wm}(i,j)\right]^2 \quad (13)$$

$$PSNR = 20\log_{10}\left(\frac{O_p}{RMS}\right)[dB] \quad (14)$$

O_p – peak value (number of quantization levels for colors).

Fig. 5 Original image

Fig. 6 Watermarked image

For the database of 99 pictures, after use of elimination function F_{c-wm} the average value of the PSNR coefficient between the original and watermarked pictures was $39,19\,dB$, at $BER = 0\%$, while between original pictures and the ones filtered using the elimination function the $PSNR = 67,88\,dB$, while detection of all watermark signals was not possible $(Y^{'}_{wm\,cepst}(m,n) < \tau)$. The result confirm the effectiveness of the hidden communication filter in the form of function F_{c-wm}, in addition in Fig. 7 confirms the return of the form of signal $O^{'}_{wm}$ to $O^{'}$.

The masking function M_{c-wm} proposed by the authors has been implemented in practice, its effectiveness has been tested using the same database of original pictures, like for the function F_{c-wm}. Result of use the masking function is shown at Fig. 8. The efficiency of the watermark masking signal in the form of a percentage of removed information (information was impossible to detect) from the watermarked pictures (the number of degraded original images $O^{'}_{deg}$ in relation to the number of photos). Information i inserted into the watermarked signal $O^{'}_{wm}$ was generated at random. Table 1.

Fig. 7 Image without watermark after use of elimination function F_{c-wm}

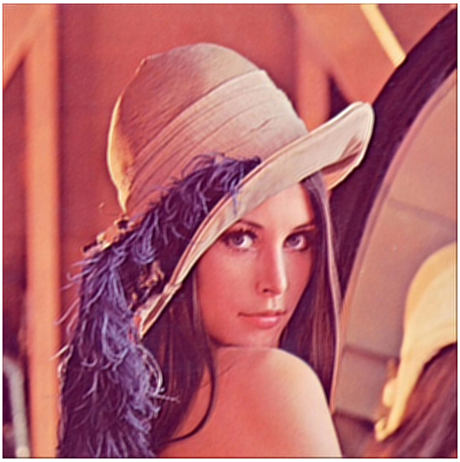


Fig. 8 Image without watermark after use of masking function M_{c-wm}

Effect –the efficiency of the masking of the watermarking signal, measured as the ratio between the watermark signals removed from watermarked pictures and the total number of watermarked pictures, maintaining the condition of returning the watermarked picture to the form of the host image.

M_{size} – sizes of the matrix of the spatial median filter,

l – translation coefficient for the matrix initiating the search for the PSF of the encoding function E_{wmf} of the Wiener blind deconvolution filter,

$PSNR_{oryg-Wm}$ – PSNR calculated between the original and the watermarked picture,

$PSNR_{oryg-c-wm}$ – PSNR calculated between the original picture and the one recovered as a result of the use of the masking function M_{c-wm}.

Taking into account the quality of the recovered host image, the algorithm used $l = 4$ and $M_{size} = [4, 4]$ as the most optimal values for the method of masking the watermark signal.

In addition, the method contained in [15] was tested using a popular masking algorithm removing embedded content described in [12] and it demonstrated high resistance to this type of

Fig. 9 Original image

Fig. 10 Watermarked image

processing: BER was only $2,04\%$. After executing masking function PSNR between the host images and the masked watermark signal was $39,27\,dB$. Tests were performed with the values of the coefficient for scaling non-linear filtering $A = 2$ recommended by the authors at [12].

5 Conclusions

The article pays particular attention to the problem of elimination and masking of watermarked signals, when it is possible to return the signal containing additional, imperceptible information to its initial, i.e. original, form. In the case of precise removal of the watermark signal, the process is called elimination, while when the signal of the watermark is removed in an approximate manner, it is called masking. Special attention should be paid to the fact that articles concerning watermarking written so far disregard these considerations (there are some, not many, masking algorithms, but they have very limited use). The article presents a developed theoretical model of a function aimed at eliminating the additional, imperceptible

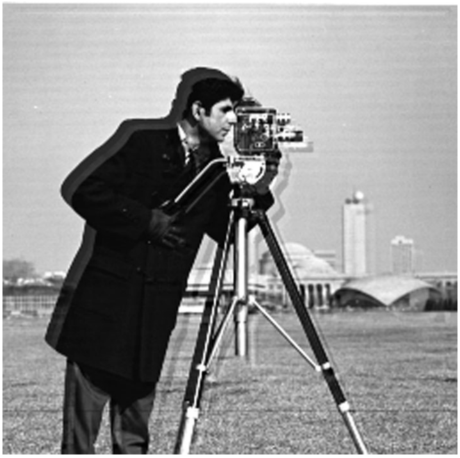


Fig. 11 Image without watermark after desynchronized masking function M_{c-wm}

Table 1 Effectiveness of implemented masking function M_{c-wm}

N	99	99	99	99
Effect [%]	81.82	99.7	98.90	97.98
M_{size}	2 × 2	4 × 4	4 × 4	5 × 5
l	4	6	4	4
$PSNR_{Oryg-Wm}$ [dB]	39.31	37.18	39.36	39.27
$PSNR_{Oryg-c-wm}$ [dB]	32.06	34.93	35.72	36.54

information inserted using a spatial algorithm operating in the cepstrum domain. A masking function was also designed. Both functions have been implemented in practice through efficiency testing which confirmed their high effectiveness. The significant robustness of the cepstral algorithm against a popular masking method has also been demonstrated.

References

1. Arnold M, Wolthusen SD, Schmucker M (2003) Techniques and applications of digital watermarking and content protection. Artech House. ISBN: 1580531113
2. Bartolini F, Barni M (2004) Watermarking systems engineering enabling digital assets security and other applications, CRC Press. ISBN: 978-0-8247-4806-7
3. Bogert BP, Healy MJR, Tukey JW (1963) The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo Autocovariance, Cross-Cepstrum and Saphe Cracking. In: Rosenblatt M (ed) Proceedings of the Symposium on Time Series Analysis, vol 15. Wiley, New York, pp 209–243
4. Cole E (2003) Hiding in Plain Sight: Steganography and the Art of Covert Communication. Wiley Publishing, Inc., Indianapolis
5. Cox IJ, Miller ML Digital watermarking and steganography 2nd ed. MK; ISBN 978-012-372585-1
6. Cox J, Kilian J, Leighton FT, Shamoon T (1997) Secure spread spectrum watermarking for multimedia. IEEE Trans Image Process 6(12):1673–1687
7. Cox IJ, Linnartz J-PMG (1998) Some general methods for tampering with watermarks. IEEE J Select Areas Commun 16(4):587–593
8. Fabien A, Petitcolas P, Anderson RJ, Kuhn MG (1988) Attacks on copyright marking systems, in David Aucsmith (Ed), Information Hiding, Second International Workshop, IH'98, Portland, Oregon, U.S.A., April 15–17, 1998, Proceedings, LNCS 1525, Springer-Verlag, 219–239. ISBN 3-540-65386-4
9. Fridrich J (1998) Applications of data hiding in digital images. Tutorial of the ISPACS '98 Conference, Melbourne
10. Hsu T-C, Hsieh W-S, Su T-S (2008) A new watermark attacking method based on eigen-image energy, intelligent information hiding and multimedia signal processing, 2008. IIHMSP '08 International Conference on 15–17 Aug. 2008, 29–32. doi: 10.1109/IIH-MSP.2008.349
11. Kalker T, Linnartz JPMG, van Dijk M (1998) Watermark estimation through detector analysis, Proc Int Conf Image Proc 425–429
12. Langelaar GC, Lagendijk RL, Biemond J (1998) Removing Spatial Spread SpectrumWatermarks by Non-Linear Filtering. Ninth European Signal ProcessingConference, Island of Rhodos, Greece, pp 2281–2284
13. Lee CH, Lee YK (1999) An adaptive digital image watermarking technique for copyright protection. IEEE Trans Consum Electron 45(4):1005–1015
14. Petitcolas FAP (2000) Watermarking schemes evaluation. IEEE Sign Proc 17(5):58–64

15. Piotr Lenarczyk, Zbigniew Piotrowski (2010) Novel hybrid blind digital image watermarking in cepstrum and DCT domain, mines, 2010 International conference on multimedia information networking and security, 356–361. ISBN: 978-0-7695-4258-4
16. Seitz J Digital watermarking for digital media, In- Information science publishing. ISBN 1-59140-518-1
17. Sencar HT, Ramkumar M, Akansu AN Data hiding fundamentals and applications, 1st edition content security in digital multimedia, ISBN: 9780120471447

Piotr Lenarczyk received the M.S. degree in telecommunication from Military University of Technology, Warsaw, Poland in 2011. His research interests are focused on digital image and real-time video watermarking.

Zbigniew Piotrowski received the M.Sc., Ph.D. degrees in Telecommunications from the Military University of Technology (MUT), Warsaw, in 1996, and 2005 (with honours), respectively. He graduated from the Stanford Center for Professional Development, Stanford University CA, USA (2013). He holds D.Sc. (a habilitation, 2013) in Telecommunications from EF MUT. He holds also title the Professor of MUT from MUT. At present he is DSP engineer in the Telecommunication Institute, Electronics Faculty (TI EF MUT). His main areas of interest are: speech and audio processing, telecommunication systems engineering and data hiding technology.