Towards Overflow/Underflow Free PEE Reversible Watermarking

Ioan-Catalin Dragoi
Electrical Engineering Dept.
Valahia University of Targoviste, Romania
Email: catalin.dragoi@valahia.ro

Dinu Coltuc
Electrical Engineering Dept.
Valahia University of Targoviste, Romania
Email: dinu.coltuc@valahia.ro

***Abstract*—The paper proposes an original prediction error expansion (PEE) reversible watermarking scheme that exploits the sign of the prediction error in order to prevent the overflow/underflow of the pixels close to the bounds of the graylevel range. More precisely, the pixels close to black are directly embedded by PEE when the prediction error is positive. Similarly, the ones close to white are embedded in case of negative errors. When the prediction error sign does not allow error expansion, an original solution is proposed. The pixel is left unchanged and one of its upper diagonal neighbors is embedded by PEE, but with a different prediction error. The diagonal neighbor is selected in order to prevent overflow/underflow. The proposed prediction error strategy ensures detection and reversibility without any additional information. The scheme is of interest for images with large areas of black and white pixels as, for instance, medical images.**

I. Introduction

A very popular approach for image reversible watermarking (RW) is prediction error expansion (PEE) [1]- [9]. The basic principle of PEE is very simple. Image pixels are scanned one by one and, for each pixel, the prediction error (PE) is computed. If no overflow or underflow appears, the PE is expanded two times and one bit of data is embedded into the expanded prediction error. The pixels that are subject to overflow/underflow are left unchanged. The two times expansion is a simple arithmetical shift with one position that clears the least significant bit (LSB) of the PE and thus, creates the free space to hide one bit. Since at most one bit can be embedded into each pixel, in a single stage, the embedding bit-rate is bounded by 1 bpp. The detection stage is very similar to the embedding one. The prediction error is computed, the embedded data is extracted as the LSB of the PE and finally, the original pixel is recovered. Obviously, the procedure works if, at detection, one can differentiate between embedded pixels and those left unchanged because of overflow/underflow. The solution is to provide at detection some additional information. If there is only a small number of such pixels, the solution is to provide a list of coordinates. If the number of problem pixels is not negligible, an overflow/underflow map is needed (with 0 for unchanged pixels and 1 for embedded pixels). The list or, equivalently, the lossless compressed map should be embedded into the image in order to be available at detection.

When the number of overflow/underflow pixels is rather low, and this is the case for most natural images, the PEE schemes are very efficient. In fact, the PEE schemes are the most efficient reversible watermarking schemes for medium and high capacity embedding bit-rates [2]. As the number of overflow/underflow pixels increases, the performance of PEE schemes can dramatically decrease. It should be observed that the risk appears for pixels close to the bounds of the graylevel range. The overflow can appear for "white" pixels and the underflow for "black" ones. Large areas of black and white pixels are usually found in medical images (X-ray, MRI, etc.). In [3], the solution to deal with such images is to estimate the black/white pixels where there is the risk of overflow/underflow and, instead of the standard PEE, to embed these pixels by a different procedure, less efficient than PEE, but ensuring the preservation of the graylevel range.

This paper refines the approach of [3]. The basic idea is that a significant part of the pixels close to the bounds of the graylevel range can be embedded by PEE. For instance, the black pixels can be embedded by PEE if the PE is positive. Similarly, the white pixels can be embedded for negative PEs. After the marking of these close to the bounds pixels, the sign of the corresponding prediction errors is unchanged. When the prediction error sign does not allow expansion, an original solution is proposed. The pixel is left unchanged and one of its upper diagonal neighbors is embedded by PEE, but with a different prediction error. The related work is briefly summarized in Section II. The proposed scheme and the experimental results are presented in Sections III and IV, respectively. The conclusions are drawn in Section V.

II. Related work

In this section, the two staged prediction error expansion scheme of [4] is first described (Section II-A) and then the invariant classification based approach of [3] is discussed (Section II-B).

A. RW based on PEE

The reversible watermarking scheme of [4] starts by splitting the image into two distinct sets, cross and dot, forming a chessboard pattern (Fig. 1.a). The watermark is also split into equal halves, one for each set. The embedding process starts with the pixels from the cross set, the dot set is watermarked only after all the pixels from the cross set were processed. The

978-0-9928-6265-7/16/$31.00 ©2016 IEEE

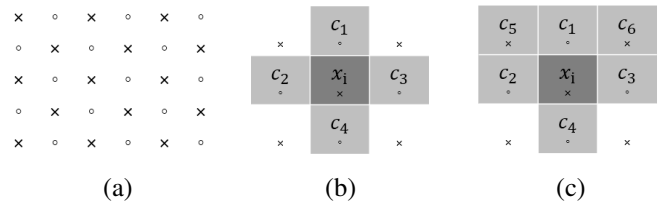

Fig. 1. a) The cross (×) and dot (∘) sets; b) the rhombus context; c) prediction context for the proposed scheme.

decoding is performed in reverse order, starting with the dot set.

Before embedding, the pixels of each set are sorted based on their local variance:

$$\mu_i = \frac{1}{4}\sum_{k=1}^{4}(\Delta\nu_k - \Delta\bar{\nu})^2 \tag{1}$$

where $\Delta\nu_1 = |c_1 - c_2|$, $\Delta\nu_2 = |c_2 - c_4|$, $\Delta\nu_3 = |c_4 - c_3|$, $\Delta\nu_4 = |c_3 - c_1|$, $\Delta\bar{\nu} = (\Delta\nu_1 + \Delta\nu_2 + \Delta\nu_3 + \Delta\nu_4)/4$ and $\{c_1, c_2, c_3, c_4\}$ is the rhombus prediction context around the current pixel (Fig. 1.b). Note that the rhombus context contains only pixels that do not belong to the current set. Thus, μ_i is not altered by the watermarking of the current set.

Only the first n pixels of the sorted set, those with the smallest local variance, are considered for watermarking. The remaining pixels are kept unchanged. The n pixels are then processed one by one, starting with the first value in the sorted set. The current pixel, x_i, is predicted using the rhombus context mentioned above:

$$\hat{x} = \left\lfloor \frac{c_1 + c_2 + c_3 + c_4}{4} + 0.5 \right\rfloor \tag{2}$$

where $\lfloor x \rfloor$ represents the greatest integer less than or equal to x. Based on $\hat{x}$, the prediction error is determined:

$$e_i = x_i - \hat{x} \tag{3}$$

The current pixel becomes:

$$x'_i = x_i + \delta_i \tag{4}$$

The distortion introduced by the watermarking, δ, is:

$$\delta_i = \begin{cases} e_i + b_i, & \text{if } -T \le e_i < T, \\ T, & \text{if } e_i \ge T, \\ -T, & \text{if } e_i < -T. \end{cases} \tag{5}$$

where $b_i \in \{0, 1\}$ is the data bit hidden in x_i and T is the embedding threshold. The threshold T is used together with n to control the available embedding bit-rate. A larger threshold allows for more pixels to contain hidden data, but the distortion introduced by the embedding increases.

At decoding, the pixels in each set are sorted based on (1) and the first n pixels are selected. The last pixel modified by the embedding process, x_n, is the first to be decoded. Each pixel is predicted with equation (2) and its corresponding modified prediction error, e'_i, is determined with equation (3).

Fig. 2. Classic test images: *Lena, Mandrill, Man* and *Kodim20*.

If $-2T \le e'_i < 2T$, x'_i contains a hidden bit, otherwise its value was shifted with T ($e'_i \ge 2T$) or $-T$ ($e'_i < -2T$). The hidden data, $b_i = e'_i \bmod 2$, and the original error, $e_i = \lfloor e'_i/2 \rfloor$, are recovered. The current pixel is then restored, $x_i = x'_i - \delta_i$.

B. RW based on invariant image classification

The graylevel range is limited to an integer interval: $x_i \in [0, L-1]$, where L is the number of possible values ($L = 256$ for 8 bit images). x'_i can potentially exceed the graylevel range. This is known as the overflow/underflow problem and is usually solved with an overflow map, [1]. If $x'_i \notin [0, L-1]$, the current pixel remains unchanged and its position is stored in the map. The map is compressed and inserted into the host image by LSB substitution together with n and T. The original LSBs should be embedded as side information by watermarking.

Only the pixels $x_i \in [0, T) \cup (L-1-T, L-1]$ can produce overflow/underflow because of their proximity to the bounds of the graylevel range. Most natural images (like the first two of Fig. 2) have a small number of values close to pure black (0) or white ($L-1$). But such values are commonly found in medical images (Fig. 3). The bit-rate required to store the overflow map limits the performance of reversible watermarking schemes on these types of images.

The invariant image classification and selective watermarking of [3] reduces significantly the number of overflow/underflow pixels. The n pixels in each sorted set are grouped based on their predicted value:

$$x_i \in \begin{cases} A, & \text{if } T \le \hat{x} \le L-1-T \\ B, & \text{if } \hat{x} < T \\ C, & \text{if } \hat{x} > L-1-T \end{cases} \tag{6}$$

We remind that $\hat{x}$ is an estimated value for x_i. In other words, in the majority of cases, $\hat{x}$ is very close to x_i. Pixels with $\hat{x} \in [T, L-1-T]$ (group A) have a low risk to produce overflow or underflow and are modified by equation (5).

Group B is at risk of underflow. These pixels are watermarked based on their graylevel value (equivalent with the multi-level embedding of [10]):

$$\delta_i = \begin{cases} d_i + b_i, & \text{if } d_i < T, \\ T, & \text{if } d_i \ge T, \end{cases} \tag{7}$$

where $d_i = x_i - x_{\min}$ is the expanded difference and $x_{\min}$ is the smallest graylevel value present in B. Note that $d_i, T \ge 0$, hence $x'_i \ge x_i \ge 0$ and underflow cannot occur. Overflow could be possible (if $\hat{x} < T$ and $x_i > L-1-T$), but it is extremely unlikely.

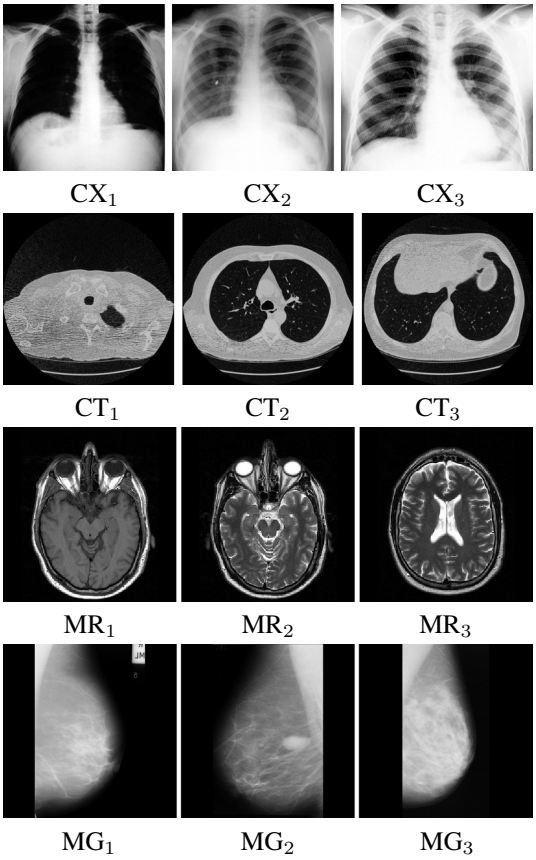

Fig. 3. Medical test images: chest x-rays (CX), computed tomography images (CT), magnetic resonance images (MR) and mammographies (MG).

Group C is at risk of overflow. These pixels are also watermarked based on their graylevel value:

$$\delta_i = \begin{cases} d_i - b_i, & \text{if } d_i > -T, \\ -T, & \text{if } d_i \leq -T, \end{cases} \tag{8}$$

where $d_i = x_i - x_{\max}$ and $x_{\max}$ is the largest graylevel value present in C. Overflow cannot occur: $d_i \leq 0 \Rightarrow x'_i \leq x_i \leq L-1$.

Before going any further, we remind that graylevel difference expansion (equation (7) and (8)) tends to be less efficient than prediction error expansion (equation (5)), even on relatively uniform areas.

III. Proposed scheme

The watermarking scheme of [3] uses only $\hat{x}$ for determining the overflow/underflow risk. Because of this, a number of pixels that cannot exceed the graylevel range were evaluated as such (for example, $\hat{x} < T$, $e_i \geq 0$ and $e_i < T$). An improved classification based on both $\hat{x}$ and e_i is proposed next:

$$x_i \in \begin{cases} A, & \text{if } T \leq \hat{x} \leq L-1-T \text{ or } (\hat{x} < T \text{ and } e_i \geq 0) \\ B, & \text{if } \hat{x} > L-1-T \text{ and } e_i \leq 0 \\ C, & \text{if } \hat{x} > L-1-T \text{ and } e_i > 0 \\ D, & \text{if } \hat{x} < T \text{ and } e_i < 0 \end{cases} \tag{9}$$

In this equation, at detection e_i should be replaced by e'_i.

Pixels in group A are modified based on equation (5). $\hat{x} < T$ normally present an underflow risk, but those pixels in A also have $e_i \geq 0$ and are modified towards $+\infty$, underflow can not occur.

The values in B (overflow risk) are watermarked by using:

$$\delta_i = \begin{cases} e_i - b_i, & \text{if } e_i \geq -T, \\ -T, & \text{if } e_i < -T. \end{cases} \tag{10}$$

This equation ensures that a pixel with zero prediction error ($e_i = 0$, the most common error) can be watermarked by using PEE, i.e., the most efficient watermarking approach. Equation (5) modifies these pixels towards $+\infty$, equation (10) modifies them towards $-\infty$, an important aspect when $x_i = L-1$. It is important to notice that for pixels in A or B, the sign of the prediction error is maintained after embedding.

For the pixels in the remaining two groups (C and D), a secondary predicted value is determined using an extended context (Fig. 1.c):

$$\hat{x}_2 = \left\lfloor \frac{c_1 + c_2 + c_3 + c_4}{4} + 0.5 \right\rfloor + c_5 - c_6 \tag{11}$$

A secondary prediction error is then computed based on $\hat{x}_2$:

$$e_2 = x_i - \hat{x}_2 \tag{12}$$

Next, c_5 and c_6 are used for context embedding. As opposed to [8] and [9], the entire distortion will be inserted in c_5 or c_6, x_i remains unchanged. Any chance to c_5 or c_6 also affects $\hat{x}_2$: $c'_5 = c_5 - \delta \Rightarrow \hat{x}'_2 = \hat{x}_2 - \delta$; $c'_6 = c_6 - \delta \Rightarrow \hat{x}'_2 = \hat{x}_2 + \delta$. In other word, the e_2 value is watermarked by modifying $\hat{x}_2$ instead of x_i. Since the values of c_5 or c_6 have no influence on the primary prediction error, e_i, the reversibility of equation (9) is ensured.

The neighbors of the pixels in group C are modified based on the secondary prediction error. Both c_5 or c_6 are located close to x_i and they are likely to have rather similar graylevel values, and consequently, overflow is much more likely than underflow. c_5 or c_6 must be modified towards $-\infty$. If $e_2 > 0$ then c_5 is modified:

$$c'_5 = c_5 - \delta_i \tag{13}$$

Otherwise ($e_2 \leq 0$), c_6 is modified

$$c'_6 = c_6 + \delta_i \tag{14}$$

These equations use:

$$\delta_i = \begin{cases} e_2 - b_i, & \text{if } -T < e_2 \leq T, \\ T, & \text{if } e_2 > T, \\ -T, & \text{if } e_2 \leq -T. \end{cases} \tag{15}$$

Note that for group C, both (13) and (14) modify their corresponding pixel towards $-\infty$.

Finally, the pixels of group D are similarly marked. One has:

$$\delta_i = \begin{cases} e_2 - b_i, & \text{if } -T \leq e_2 < T, \\ T, & \text{if } e_2 \geq T, \\ -T, & \text{if } e_2 < -T. \end{cases} \tag{16}$$

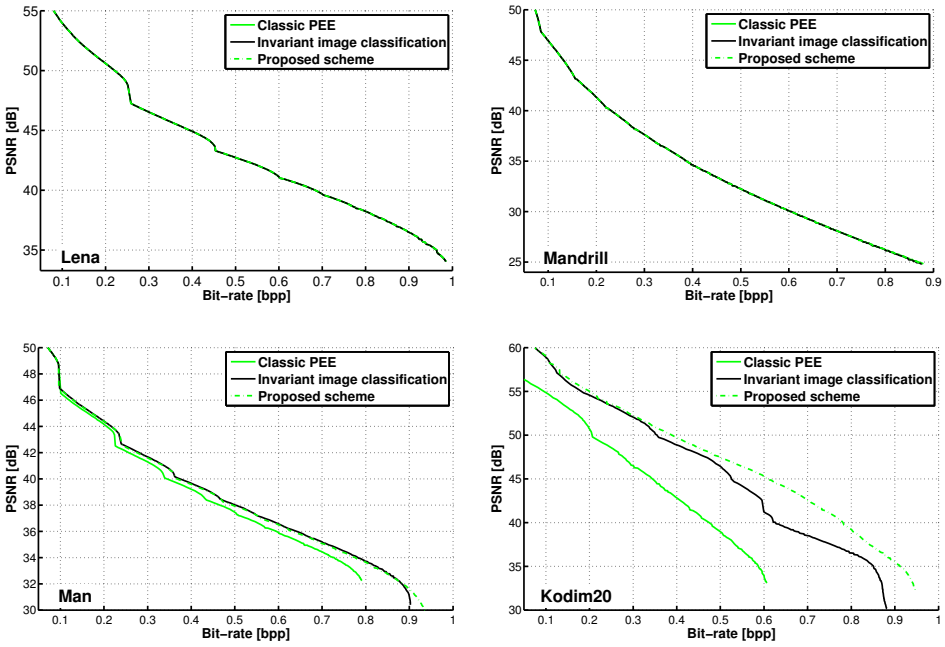

Fig. 4. Experimental results on the classic test images.

If $e_2 < 0$ then c_5 is modified with equation (13). Otherwise equation (14) is used to modify c_6. For group D, the neighboring pixel is always moved towards $+\infty$. Note that a pixel can be watermarked more than once in the same embedding stage (for instance, as current pixel and then as neighbor of lower diagonal current pixels).

IV. Experimental results

The experimental results of the proposed scheme are compared with the ones provided by the classification approach of [3] and the well-known PEE scheme of [4] on four natural test images (Fig. 2) and twelve medical images (Fig. 3). The test set has 8 bit graylevel images of varying size. Thus, *Lena, Mandrill*, the chest x-rays and the computer tomography images have around 512×512 pixels, the magnetic resonance images have 350×350 pixels, *Man* and the mammographies 1024×1024 pixels and finally, *Kodim20* has 512×768 pixels.

Before going any further, note that all three tested schemes use equation (5) to watermark the pixels with $T \leq \hat{x} \leq L-1-T$. The difference in performance comes from the way each scheme processes the pixels with $\hat{x} \in [0, T) \cup (L-1-T, L-1]$.

The results on natural images are presented in Fig. 4. On *Lena* and *Mandrill*, where there is only a small risk of overflow/underflow, all three schemes provide equivalent results. On the remaining two images, (*Man* and *Kodim20*), where there are large regions of black (white) pixels, the proposed scheme outperforms both [3] and [4].

The watermarking results on the twelve medical images are shown in Fig. 5. On the chest x-rays, the classification based approach of [3] is outperformed by the classic PEE scheme of [4]. This is because of marking with (7) or (8) (in order to prevent overflow/underflow) pixels that can be still marked with (5). The proposed scheme is entirely based on PEE and does not have this disadvantage. It also outperforms [4] at higher bit-rates, starting with 0.68 bpp on CX_2 and approximately 0.9 bpp on CX_1 and CX_3.

On the computer tomography images, the poor prediction together with a large number of black pixels considerably limits the performance of [4], while [3] and the proposed scheme still offer bit-rates above 0.3 bpp. Compared to [3], the proposed scheme provides an average gain in PSNR of 0.75 dB on CT_2, 0.71 dB on CT_2 and similar results on CT_1. The proposed scheme also obtains on CT_2 and CT_3 an increase in bit-rate over [3] of 0.05 bpp (an extra 13000 pixel are available for data hiding). An increase in bit-rate is also observed on the other test images.

On magnetic resonance images, the proposed scheme clearly outperforms the other two approaches. Compared to [3], the proposed scheme offers an average gain in PSNR of 1.78 dB, 1.35 dB and 1.86 dB on MR_1, MR_2 and MR_3, respectively. Finally, on mammographies, [4] has superior results for low bit-rates (< 0.2 bpp), but the proposed scheme clearly outperforms [4] when the entire bit-rate domain is considered. The proposed scheme also has superior results compared to [3] for bit-rates above 0.95 bpp.

V. Conclusions

A novel PEE reversible watermarking scheme that combines the expansion of the prediction errors in the directions where there is no risk of overflow/underflow with a special designed context marking procedure has been proposed. The scheme is of interest for images with large areas of black and white pixels as, for instance, medical images. The experimental results obtained so far are promising.

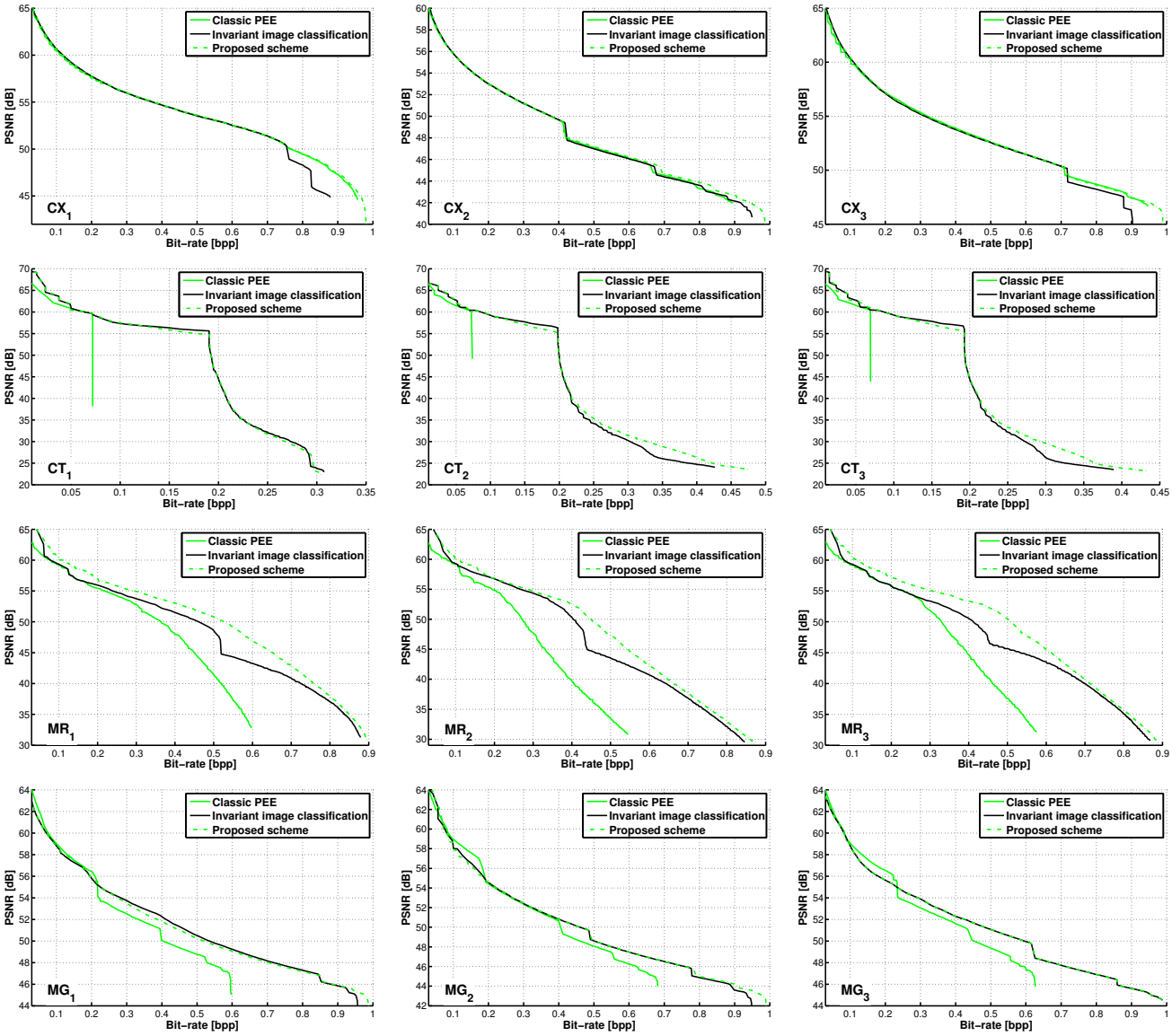

Fig. 5. Experimental results on the medical test images.

ACKNOWLEDGMENT

This work was supported by UEFISCDI Romania, PN-II-PTPCCA-2013-4-1762 and PN-II-PTPCCA-2011-3.2-1162 Grants.

REFERENCES

[1] D. M. Thodi and J. J. Rodriguez, "Expansion Embedding Techniques for Reversible Watermarking", *IEEE Trans. Image Process.*, vol. 15, pp. 721–729, 2007.

[2] I.-C. Dragoi and D. Coltuc, "Local Prediction Based Difference Expansion Reversible Watermarking", *IEEE Trans. on Image Processing*, vol. 23, no. 4, pp. 1779–1790, 2014.

[3] G. Coatrieux, W. Pan, N. Cuppens-Boulahia, F. Cuppens and C. Roux, "Reversible Watermarking Based on Invariant Image Classification and Dynamic Histogram Shifting", *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 111–120, 2013.

[4] V. Sachnev, H. J. Kim, J. Nam, S. Suresh and Y. Q. Shi, "Reversible Watermarking Algorithm Using Sorting and Prediction", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, pp. 989–999, 2009.

[5] X. Li, B. Yang and T. Zeng, "Efficient Reversible Watermarking Based on Adaptive Prediction-Error Expansion and Pixel Selection", *IEEE Trans. on Image Process.*, vol. 20, no. 12, pp. 3524–3533, 2011.

[6] X. Li, B. Li, B. Yang and T. Zeng, "General Framework to Histogram-Shifting-Based Reversible Data Hiding", *IEEE Trans. on Image Process.*, vol. 22, no. 6, pp. 2181–2191, 2013.

[7] W.-J. Yang, K.-L. Chung, H.-Y. M. Liao and W.-K. Yu, "Efficient reversible data hiding algorithm based on gradient-based edge direction prediction", *Journal of Systems and Software*, vol. 86, no. 2, pp. 567–580, 2013.

[8] D. Coltuc, "Improved Embedding for Prediction Based Reversible Watermarking", *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 3, pp. 873–882, 2011.

[9] D. Coltuc, I.-C. Dragoi, "Context embedding for raster-scan rhombus based reversible watermarking", *roceedings of the first ACM workshop on Information hiding and multimedia security (IH&MMSec 13)*, 215–220, 2013.

[10] Z. Ni, Y. Q. Shi, N. Ansari, and W. Su, "Reversible data hiding", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 16, no. 3, pp. 354–362, 2006.