SPIE

Connecting minds. Advancing light.
SPIE is the international society for optics and photonics

SEARCH

HOME   CONFERENCES + EXHIBITIONS   PUBLICATIONS   EDUCATION   MEMBERSHIP   INDUSTRY RESOURCES   CAREER CENTER   NEWS + VIDEOS

Conference Proceedings

Journals

SPIE Digital Library

Books

Collections

Open Access

Contact SPIE Publications

Progressive ROI coding and diagnostic quality for medical image compression (Proceedings Paper)

Author(s): Alberto Signoroni; Riccardo Leonardi

Date: **9 January 1998**

ISBN: **9780819427496**

○ **PDF**
  Member: **$18.00** | Non-member: **$18.00**

○ **Hard Copy**
  Member: **$24.00** | Non-member: **$24.00**

Add to Cart

Paper Abstract

This work addresses the delicate problem of lossy compression of medical images. More specifically, a selective allocation of coding resources is introduced based on the concept of 'diagnostic interest' and an interactive methodology based on a new measure of 'diagnostic quality'. The selective allocation of resources is made possible by a selection a priori of regions of specific interest for diagnostic purpose. The idea is to change the precision of representation in a transformed domain of region of particular interest, through a weighting procedure by an on- line user-defined quantization matrix. The overall compression method is multi-resolution, provides for an embedded generation of the bit-stream and guarantees for a good rate-distortion trade-off, at various bit-rates, with spatially varying reconstruction quality. This work also analyzes the delicate issue of a professional usage of lossy compression in a PACS environment. The proposed compression methodology gives interesting insights in favor of using lossy compression in a controlled fashion by the expert radiologist. Most of the ideas presented in this work have been confirmed by extensive experimental simulations involving medical expertise.

Current SPIE Digital Library subscribers click here to download this paper.

New Titles Update

Sign up for monthly alerts of new titles released.

Subscribe

Progressive ROI Coding and Diagnostic Quality for Medical Image Compression

A. Signoroni & R. Leonardi

Signals & Communications Lab., Dept. of Electronics for Automation, *
University of Brescia, Italy.
E-mail: leon@ing.unibs.it, alby@imago.ing.unibs.it

ABSTRACT

This work addresses the delicate problem of lossy compression of medical images. More specifically, a selective allocation of coding resources is introduced based on the concept of "Diagnostic Interest" and an interactive methodology based on a new measure of "Diagnostic Quality". The selective allocation of resources is made possible by a selection a priori of regions of specific interest for diagnostic purpose. The idea is to change the precision of representation in a transformed domain of regions of particular interest, through a weighting procedure by an on-line user-defined quantization matrix. The overall compression method is multi-resolution (wavelet-based), provides for an embedded generation of the bit-stream and guarantees for a good rate-distortion trade-off, at various bit-rates, with spatially varying (ROI dependent) reconstruction quality. This work also analyzes the delicate issue of a professional usage of lossy compression in a PACS environment. The proposed compression methodology gives interesting insights in favor of using lossy compression in a controlled fashion by the expert radiologist. Most of the ideas presented in this work have been confirmed by extensive experimental simulations involving medical expertise.

Keywords: medical image, selective coding, diagnostic quality, PACS, wavelet transform, EZW algorithm

1. INTRODUCTION

In the perspective of a digitally equipped Radiology Department, where all images are likely to be archived in digital format, or in a medical center where all supporting information for patients is digital, it is essential to limit as much as possible the memory requirements[1,2] to store the huge quantity of data. This would not only reduce the storage costs but also it would keep reasonable transmission delays for any telemedicine application. Just consider that a single MR, DSA, CT, CR, ... examination can produce images for tens of Mbytes.

In this work, a novel methodology to compress diagnostic pictures is proposed. The approach seems quite attractive as it appears to be fully integrated into the diagnostic activity of a digitalized Radiology department supplied with a PACS (Picture Archiving and Communication System).

The coding algorithm consists of three parts: a unit that performs a linear uncorrelation of the source data by using a Discrete Wavelet Transform (DWT),[3] a unit that quantizes information in the transformed domain using a very efficient non-linear prediction scheme with a locally adjusted weighting of the quantization matrix, and, finally, a context based arithmetic coder (see Fig.1, section 2). The overall coding strategy is a variation of Shapiro's EZW algorithm.[4]

The relatively low perceptual incidence of the coding artifacts, the high efficiency in the rate-distortion sense and the progressivity of the generated bit-stream give this technique high performance in a complexity vs efficiency sense, as indicated in the most recent literature.[5,3]

For the same reasons, it is here considered as a preferable candidate to be used in the biomedical field for storing images with a predefined compression strategy by simply using an iterative truncation policy. It is also very well suited for transmission purposes with scale and/or resolution increasing as more bits become available at the decoder end.

The innovative kernel of our proposal lies in the introduction of a selective allocation of compression resources

according to the presence of one or more Region(s) Of Diagnostic Interest ($RODI$). Such an approach is theoretically justified in a inter–disciplinary framework including the theories of visual attention, the diagnostic practice and the field of human-machine interaction.[6,7]

2. EZW CODING OF WAVELET IMAGES

The introduction of the discrete wavelet representation has arisen much interest to the image coding community. This fact is due to the capability of the wavelet transform to achieve both a frequency compaction and a spatial localization of the transformed coefficients. Following the initial image subband coding scheme,[8] the first examples of image coders based on wavelet transform used only the energy compaction capability, achieving compression through an adequate scanning and a proper bit allocation of the different subband coefficients.[9] Second generation wavelet coders take into account the second feature as well, i.e. the spatial localization of the transformed coefficients, and the consistent occurrence of high energy coefficients across the multiresolution representation of detail information that is provided by the wavelet transform. This allows to efficiently predict information across scales due to the co–localization of high energy information. The Embedded Zerotree Wavelet coding algorithm[4] appears to be one of the most efficient scheme for image compression. It deals in a systematic fashion with the quantization and the prediction of discrete wavelet coefficients using the inherent structure of patterns present in the image.

The discrete wavelet transform of a sample image is shown in Fig.2. The DWT performs a multiresolution dyadic subband decomposition of the original image. The figure indicates as well the filterbank structure that implements the analysis and synthesis stages of the transformation (direct and inverse DWT).[3] Also it is shown the parent–child relationship that exists among coefficients at different resolution levels along the various spatial directions (horizontal, vertical and diagonal).
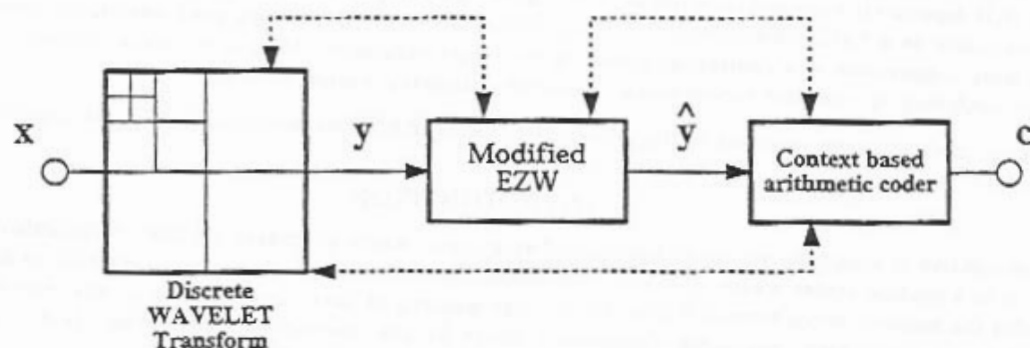

Figure 1. *Image coding scheme.*

It clearly appears in the transformed image that even if the information is uncorrelated after the DWT, there still are residual dependencies among the d_k^j coefficients along the parent–children relationships that exist among the different resolution levels forming a tree structured graph. In particular a key aspect is represented by the ability of predicting non significant information (zero level coefficient after quantization). For images with $1/f$ spectrum characteristics (to which biomedical images or signals belong), if a coefficient appears to be insignificant with respect to a certain threshold at a given resolution level (the root level), it is very likely that none of its direct descendents (which are at higher resolution levels) will be significant. When this non–significance pattern path occurs, a zerotree is created, and this graph structure can be represented by a simple codeword.

Other important characteristics of the EZW algorithm are the recursive scanning strategy of the wavelet coefficients proposed by Shapiro so as to halve the quantization level at each step. This procedure together with the order of the scanning allows to generate an embedded bit–stream which will provide a progressive increase of the reconstruction quality. Finally, the ability to represent all the quantized information using a very small size alphabet (4 symbols) increases the performance of the last stage of the compression scheme, i.e. the adaptive arithmetic coding stage. In particular the reduced number of symbols allows for a better adaptation of the statistics needed by the arithmetic coder, when context is taken into account.
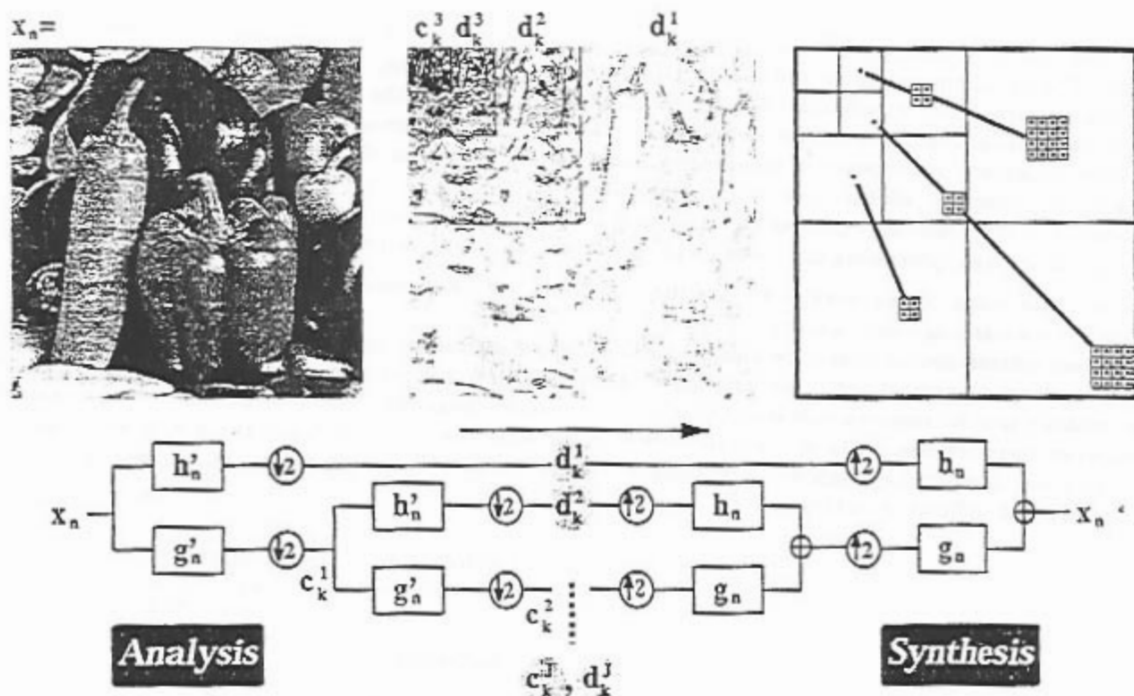
Figure 2. *Visual example, co-spatiality relations and filter bank structure of the Discrete Wavelet Transform.*

3. USING EZW FOR MEDICAL IMAGE COMPRESSION

As far as medical image compression is concerned, the EZW algorithm seems quite adequate for the following reasons:

- It represents an optimum trade-off between coding efficiency and algorithmic complexity of the compression scheme.
- The embedding of the bit-stream guarantees the progressivity of the reconstruction.
- Wavelet based coding lower the perceptual incidence of the coding artifacts with respect to the DCT based coding (e.g., the JPEG standard).
- Thanks to the bit-plane quantization procedure (adaptive threshold being halved at each iteration of the scanning of the wavelet coefficients), reconstruction artifacts should be marginally visible at the spatial location in rough correspondence with the truncation of the bit-stream.
- The EZW compression is adequate for a wide range of compression ratios, which is particularly attractive in the context of medical image compression, given the legal implication of coding artifacts.[1] It well adapts itself to the different compression policies that may be needed for storage or archiving with respect to a telemedecine environment (lossless,[5] visually lossless[7] and lossy).

The progressive reconstruction property of the bit-stream permits to design interactive protocols of communication, that allow an optimal multi-scale/multi-resolution presentation strategy for telediagnosis. For archiving purposes, different temporally well defined compression ratios could be used according to the specific needs of the medical community. This procedure could be directly implemented on the original compressed bit-stream without having to recover the original data, rather by a simple truncation of the bit-stream, thanks to its embedding.

In the specific context of lossy compression, there are clearly medical and legal implications. Accordingly, it is suggested that any scheme that might result in a loss of information, should be fully under control of the medical experts, bringing the decision task regarding the required reconstruction quality into the skills of the medical profession.

For these reasons a lossy compression paradigm is introduced (see Fig.3) so as to incorporate the compression algorithm. The paradigm can be viewed jointly with the perceptually–lossless one proposed by Jayant.[7]

In our case the first block is semantic, because the diagnosis is made by the physician, so that the paradigm can be called "semantically–lossless". The second unit is a "diagnostic parameter estimator", which translate some measurables diagnostic parameters in compression functionality by means of a knowledge based policy[10] and an inhomogeneous (selective) allocation of coding resources.

The validation process can range from an objective level to a high level that includes semantic implications, depending on the type of analysis/processing that need to be carried out on the reconstructed material (after compression).

The ultimate scope of our work is to give the radiologist the total control of the compression result without impairing his normal diagnostic activity.

The proposed scheme should clearly be linked to man–machine interaction issues to allow for high level perceptual tasks, that cannot be carried out in an automated fashion (except possibly in some application specific contexts). This is certainly true for complex operations that lead to a medical diagnosis. The final validation of the compression procedure lies therefore essentially at a semantic level, in the sense that it will be based at the level where semantics comes into place to reach the diagnosis. Accordingly, a diagnostic measure of quality will be proposed later into the paper, to face this difficult objective.
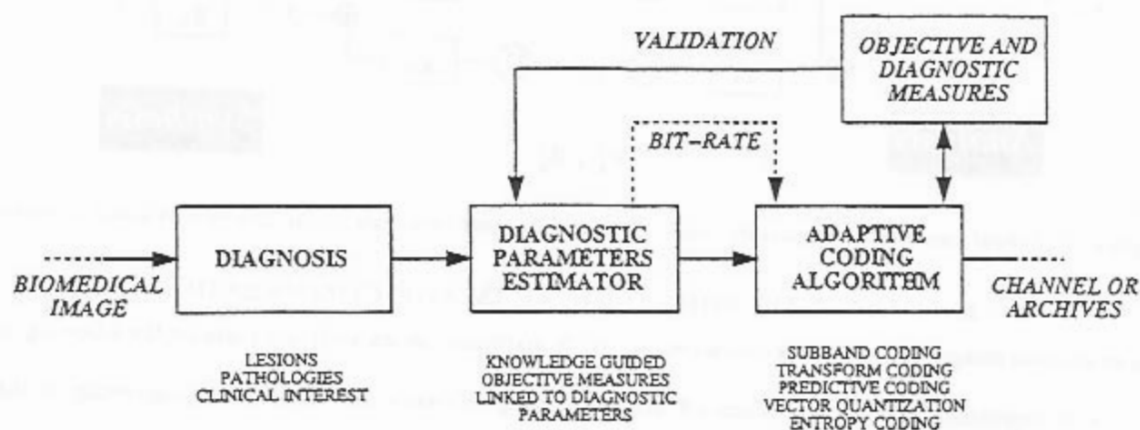
Figure 3. *A lossy compression paradigm for medical image coding.*

4. REGIONS OF DIAGNOSTIC INTEREST

4.1. Selective Coding

Selective image compression can be carried out in different ways, depending on the application context. In fact, the region of interest (in the original image domain, or in a transformed representation) can be selected only if the coder is provided with additional information. This has been named by Nguyen *compression a priori*.[6] The reason to define a "compression a priori" comes from a particular motivation. This motivation can be more or less measured in objective terms depending upon the amount of low or high level perceptual tasks that takes place. The more objective the motivation, the more it may be possible to measure some parametric values upon a variety of regions that have been extracted by some simple segmentation procedure. In the opposite case, it becomes necessary to ask the human observer to provide for a *selection a priori*.

In Nguyen's work[6] remarkable theoretical considerations are provided with respect to the selective data compression strategy. In this framework instead,[10] a high level perceptual approach is taken into consideration, as it involves a medical diagnosis task.

The medical image data are hierarchically organized by applying the approach suggested by Nguyen:

- the selection a priori defines

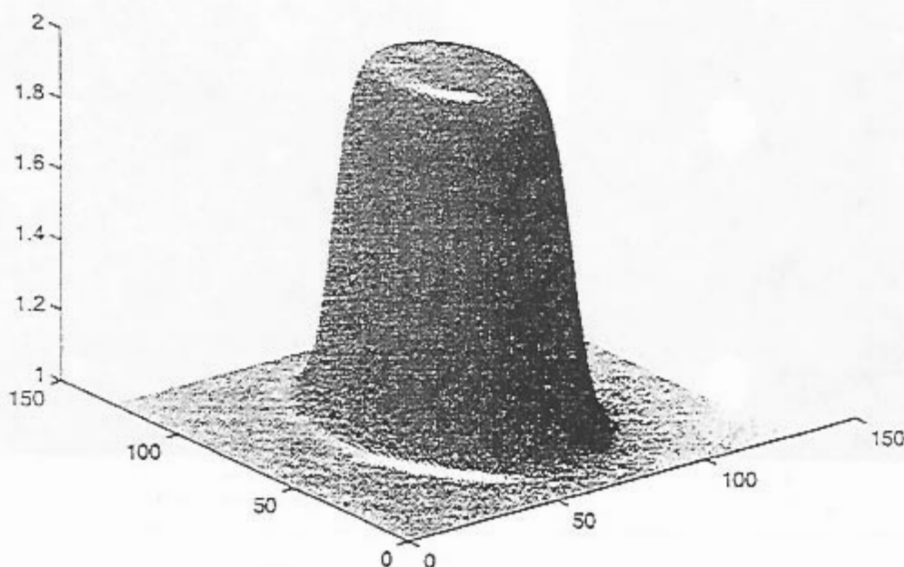 1. the localization of salient information within the picture;

677

Figure 4. *Masking function in the image domain.*

    2. the relative importance of such information in the global image context;

- the compression a priori defines the resource allocation policy in relation to the selection a priori.

To legitimate the use of lossy compression, the final coding result will have to satisfy some kind of "Diagnostic Conservation Criterion".

4.2. Quantization matrices

Perhaps the most intuitive way to achieve selective compression is to transform an existing compression scheme into a selective one. As mentioned previously our starting point is the EZW algorithm.

It will be shown that the proposed technique does not modify the essence of the original Shapiro's algorithm; on the contrary some properties will extend their significance or functionality, as in the case of embedding.

The allocation policy of the compression algorithm is made selective by using some weighting masks (or quantization matrices), which emphasize the image coefficients within the $RODI$. The weighting mask are not applied in the image domain, rather in the Wavelet Transform domain using a set of subband windows; this first avoids the introduction of high frequency coefficients which would cause a drop in coding gain, and also makes the mask framework more versatile. The subband masking exploits the co-spatiality information trees in the wavelet coefficient domain. The masking functions are mirrored–sigmoidal shaped with elliptical symmetry, so as to form a set of bell shaped windows which are maximally flat within the $RODI$ and within the out–of–$RODI$ ($\overline{RODI}$), with a smooth transition region (see Fig.4), the decaying speed of which can be easily modulated.

A too sharp transition between $RODI$ and $\overline{RODI}$ has to be considered a potentially objectionable artifact to be avoided in biomedical field; moreover, the elliptical selection match well the anatomical and pathological morphology (it does not reflect, for example, polygonal contours). Mask informations can be coded using very few bits.

A set of parameters is defined so as to make the subband masking a flexible and versatile tool for the selective coding in a variety of contexts: for example, subbands weighting coefficients were adjusted according to perceptual criterions; they also take into account the length of the wavelet filters and diagnostic considerations about the absolute extension of the $RODI$ with respect to the related scaled versions into the subband domain.

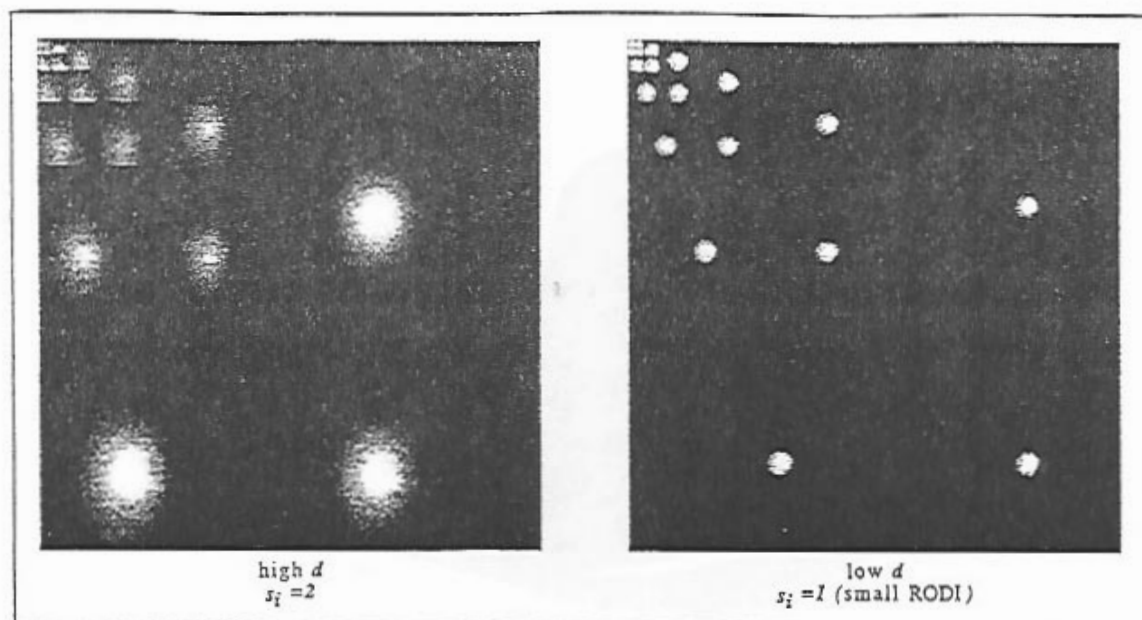The analytical expression of the subband set of masks is the following:

Figure 5. *Examples of subband-mask functions.*

$$m(x,y) = (t^{(ij)} - b^{(ij)}) + b^{(ij)} \cdot \left[\frac{1}{1 + \exp\left(\sqrt{(x - x_c^{(ij)})^2 \cdot e^2 + (y - y_c^{(ij)})^2} - r/s_i\right) \cdot d} \right] \quad (1)$$

- (x_c, y_c) represents the center of the mask;
- e represents the ellipticity degree $\in (0, \infty)$; $e = 1$ for a circle;
- s represents the scaling factor;
- r represents the main radius;
- d defines smoothing of the decay;
- t defines the $RODI$ amplitude (top);
- b defines the $\overline{RODI}$ amplitude (bottom).

where $i = 1 \ldots p$ is the decomposition level, p is the decomposition depth, and $j = 1 \ldots 3$ is the considered subband (h,v,d) which is set to 0 only for $x_c^{(p0)}, y_c^{(p0)}$, the low-pass subband.
Examples of such set are shown in Fig.5; the decision of which type of mask is to be used is briefly discussed in the next section.

It appears that the mask does not disturb the progressivity property of the compression scheme; on the contrary it is shown that the above property evolves in a so-called hierarchical "multi-progressive" fashion through the use of the weighting parameters.

This concept is shown in a very intuitive fashion in Fig.6, whereas in a more detailed presentation work,[10] a more precise justification based on the embedding Shapiro's hierarchy has been proposed.

5. AN INTERACTIVE METHODOLOGY

Another issue addressed in this work is the simplification of the selection of the $RODI$ parameters by a non technical compression expert, such as a radiologist at the time of diagnosis. A set of three diagnostic parameters, which are well adapted to the natural training of the physician, have been identified:
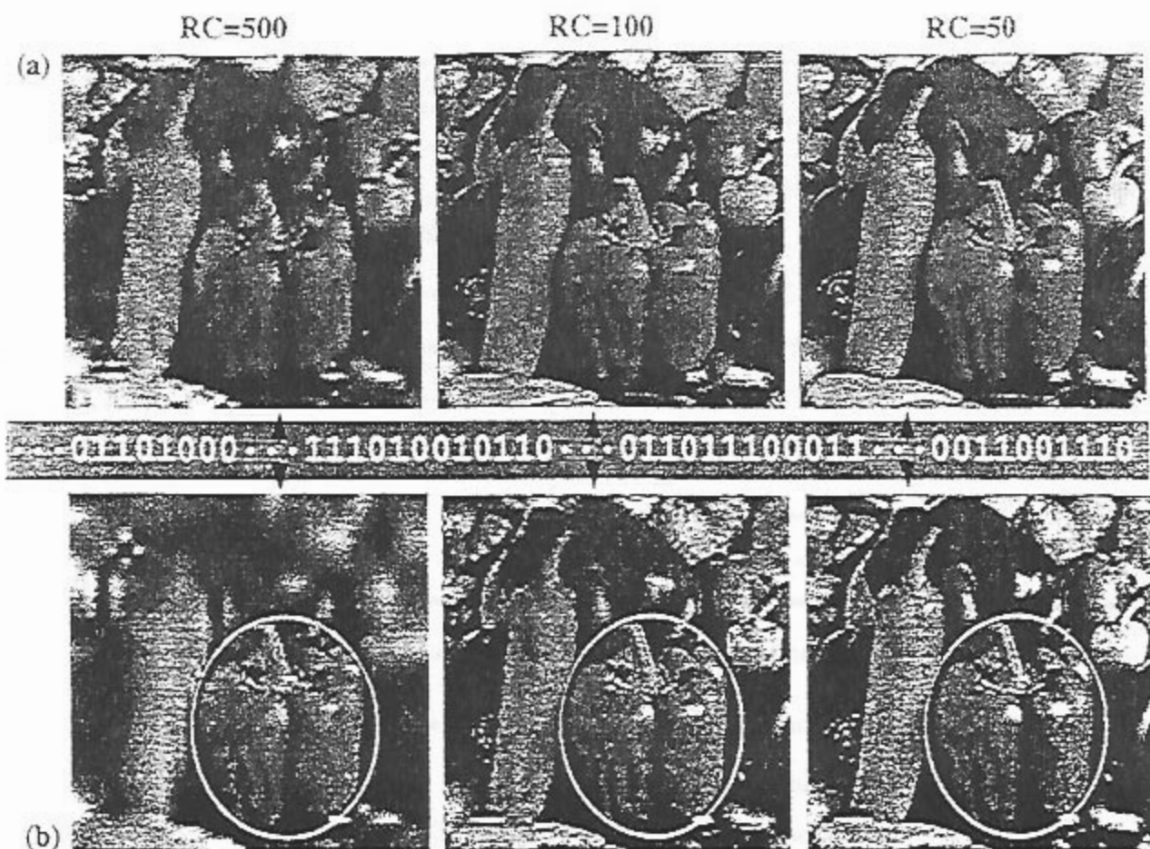
Figure 6. *The concept of (a) progressivity; (b) hierarchical progressivity.*

- the relative importance I_r among $RODI$ and $\overline{RODI}$ (the main parameter which regulate the mask amplitude);

- the absolute reference I_a of the $RODI$ (which is related to the truncation of the coding–stream);

- a selectivity factor S of the $RODI$-$\overline{RODI}$ transition (which regulates, visually and in an energy related fashion, the $RODI$-$\overline{RODI}$ transition on the resulting coded image).

A set of relationships which translates the diagnostic parameters in weighting ones, has been introduced. This allows to perfectly integrate the proposed methodology in the usual diagnostic activity of the radiologist. He is able to control the compression at will, and the technique evolves naturally as part of establishing a diagnosis.

The use of the quantization matrices allows for a selective allocation of coding resources (of help to the diagnostic task), but it introduces also additional information which is semantically very relevant (with a marginal increase of time costs) to specify the $RODI$ shape and the weighting parameters. This information may help in establishing a possible diagnosis and can help for the subsequent archive consultations or teleradiology applications.

In order to measure the subjective quality, a new measure of *Diagnostic Quality* is proposed. It defines the ability to perform the same diagnosis on the reconstructed coded image with respect to the original one. In particular, the "diagnostic quality" creates the right balance between theoretical framework and experimental practice. The following operative definition of diagnostic quality can be directly exploited for experimentations and for professional use:

a lossy reconstructed radiological image can be accepted from a diagnostic quality point of view, if a physician with same qualification level can establish the same diagnosis on the reconstructed image (with all coding information enclosed) with respect to the original one.
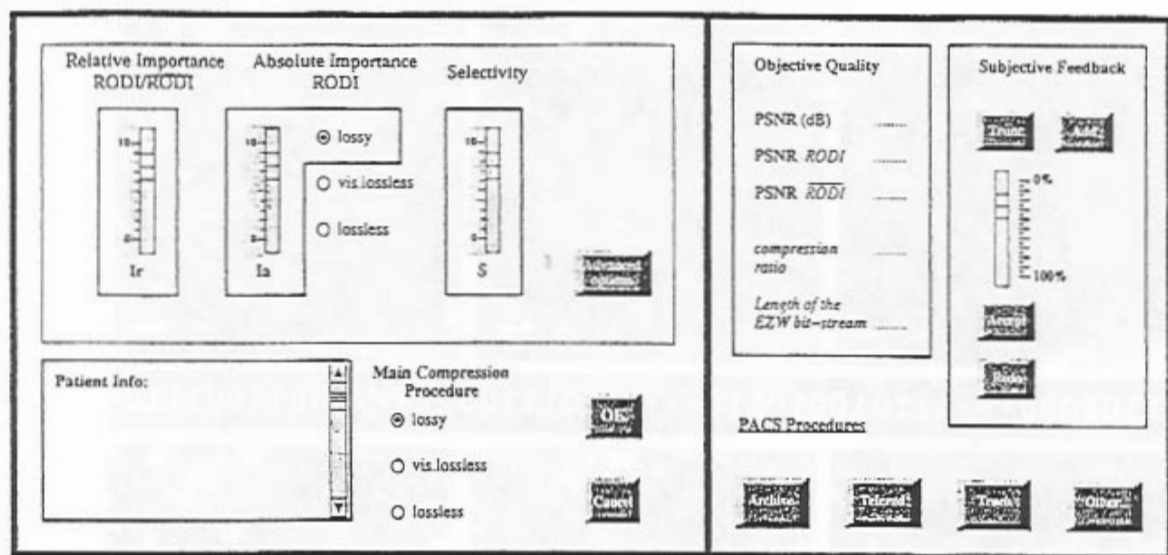
Figure 7. *Man–machine interface for managing the compression.*

The proposed compression methodology is conceptually subdivided in three stages:

- diagnosis with diagnostic parameter and $RODI$ selection;

- check and validation of the compression result in order to start a PACS procedure (archiving, teleradiology, teaching, ...);

- remote reception or archive reconstruction.

The first two steps are consecutive whereas the third can take place with a certain temporal delay (in order of minutes up to some years).

In Fig.7 a simple graphical user interface is provided. It allows to manage the coding application. This facilitates the *selection a priori* task (the $RODI$ is interactively determined with the mouse) and provides for an efficient validation window.

This approach gives the physician complete control of the lossy compression in little time (as it was experimentally tested), especially when the radiologist is confident with the methodology.

6. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of the methodology, a series of objective and subjective experimentations have been carried out with the support and critical contribution of the medical counterpart. These experiments aimed at describing the coding system and its performance, as well as testing the possibility of a professional usage of this interactive compression methodology.

We have obtained plots of the operational R–D curve and estimated the relationships that may exist between objective distortion measures and the mask features (size, diagnostic parameters,...). This allowed to obtain a complete characterization of the algorithm and specifically a set of quantitative considerations about the use of weighting masks.

In Fig.9 the operational Rate–Distortion curve is plotted for a particular masking profile, shown in Fig.8a. The double alternating slope of the curves is a characteristic of the EZW algorithm (which alternates a sequence of dominant and subordinate passes[4] of quantization).

The subjective experimentations were carried out on a set of significant pathological cases so as to test the system from a critical diagnostic quality point of view; for example single wide critical lesion or multiple small centers. Tests were essentially carried out on MR images 512x512 and 8bpp of resolution (after 12 bit gray–level windowing and/or contrast selection), as shown in Fig.10 and Fig.8a.
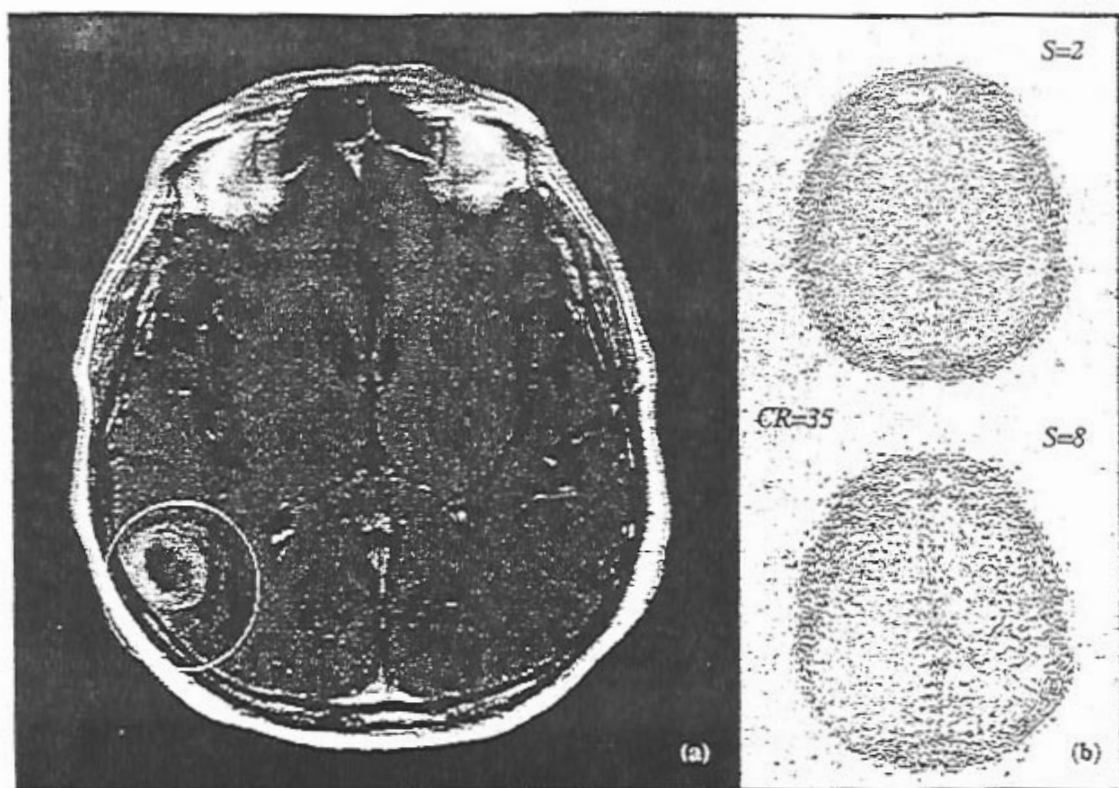
Figure 8. *Brain image of a bruise: (a) the RODI, (b) different error images for two choice of the selectivity coefficient.*

Experiments were performed to establish:

- the benefit of the use of the weighting masks;
- the relationship between diagnostic quality and presence of visible artifacts;
- the optimal strategy to define perceptive weighting options;
- the inhomogeneous distribution of quality due to the selectivity of the used mask.

In Fig.11 the advantage of using masks is demonstrated in terms of visibility of a small center in a low contrasted MR image. This advantage is obvious when the radiologist would like a nearly lossless quality in the $RODI$ to allow for a correct diagnosis.

It has been possible to establish a certain correlation between diagnostic quality and objectionability of the artifacts. However in most cases, up to a specific compression ratio (usually $30 - 40$, at the tested resolution level), the radiologist appears not to be disturbed by the presence of some visible artifacts.

Furthermore, selectivity regulates not only the passage from $RODI$ and $\overline{RODI}$ quality but, as a consequence, addressed the coding resource allocation in the different regions of the image with respect to a target compression ratio. As an example this is indicated in Fig.8b.

Selectivity is sometimes discriminant to preserve diagnostic information outside the $RODI$. This is the case of Fig.12 where a bad choice for S and a high compression ratio may cause the loss of a center. We notice, however, that this selection a priori by an expert radiologist is very unlikely given the small size of critical regions represented by the centers and the potentially distributed nature of the pathology. A more adequate selection a priori indicate a small value for S and a moderate compression ratio (or a multiple RODI).

The achieved results demonstrate the possibility for a professional usage of the proposed methodology for archiving
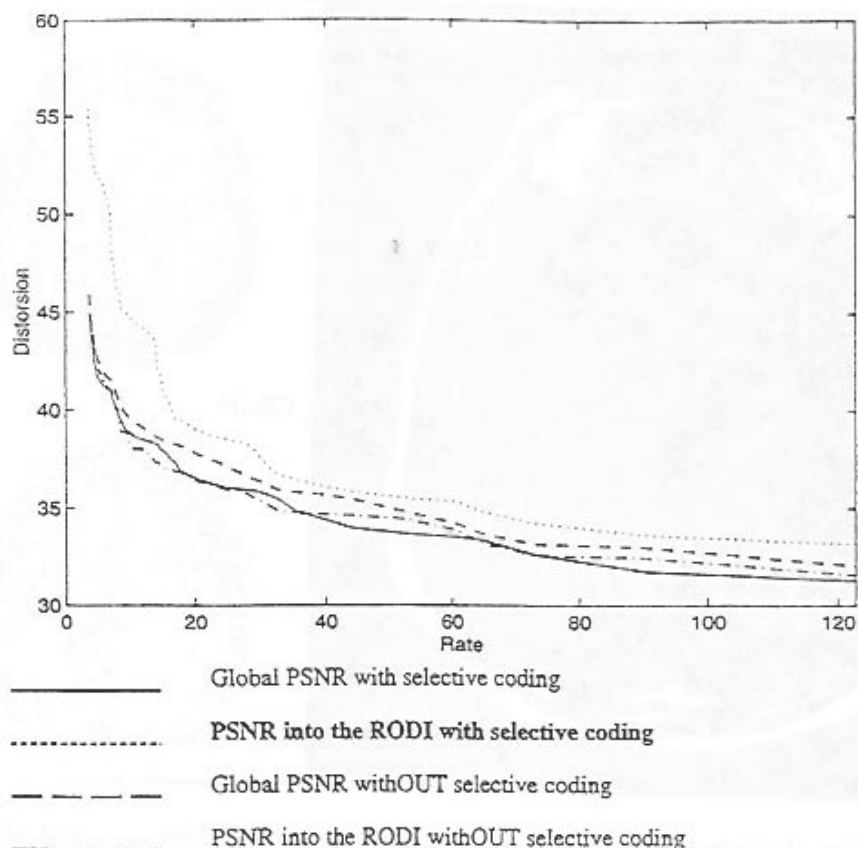
Figure 9. *Rate–Distortion plot for a coding instance, with and without application of the mask.*

and for teleradiology, with relevant economic benefits considering the high compression ratios that could be obtained ($\sim 30 - 50 : 1$) while preserving the diagnostic quality.

7. CONCLUSIONS

This work has proposed a semantical methodological approach, based on Regions of Diagnostic Interest, to selectively allocate compression resources, allowing to reach various quality degrees at different locations in the reconstructed image. Such quality can clearly range from perfect (lossless) to visually lossless or even lossy. In this context, it has been possible to evaluate some medical and legal implications of lossy compression for biomedical images.

A fundamental rule was played by the **Diagnostic Quality** parameter in order to define the boundary and the applicability degree of the weighting masks. This parameter has been extensively used during the subjective experimentations and can also be used in a "natural" fashion within the normal diagnostic practice. This kind of man–machine interaction is needed as the high level diagnostic task cannot be reached with an automatic algorithm. The compression degree and the resource allocation can fully reflect the medical and legal responsabilities of physicians that perform the diagnosis and validate the quality of the compressed image. Furthermore, thanks to the progressivity of the bit–stream, higher compression can be reached at a later stage, depending on the memory needs within the PACS. At high compression ratio (around 50, starting from 8bpp image representation) no experimental simulations lead to a false lesion.

Future research perspective are focused on the extension of the proposed methodology on a frame sequence of anatomic sections and the introduction of volumetric *RODIs*, using semi–automatic selection algorithms for medical suggestion and anatomical deformation field pursuit.
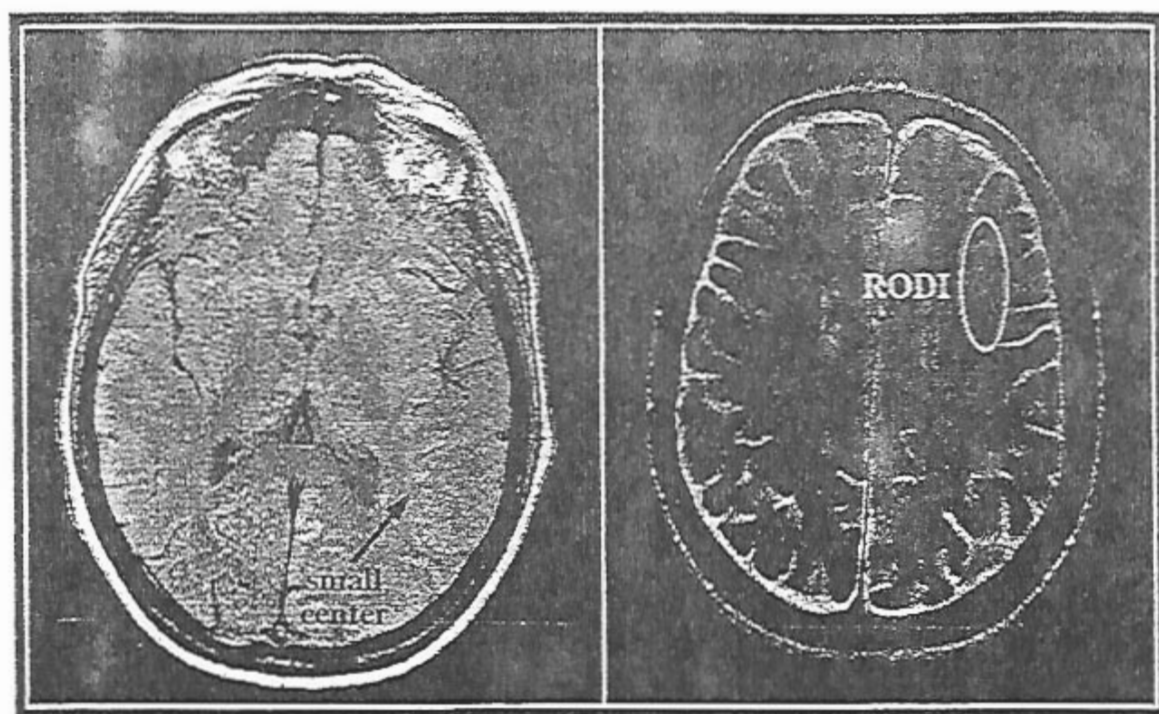
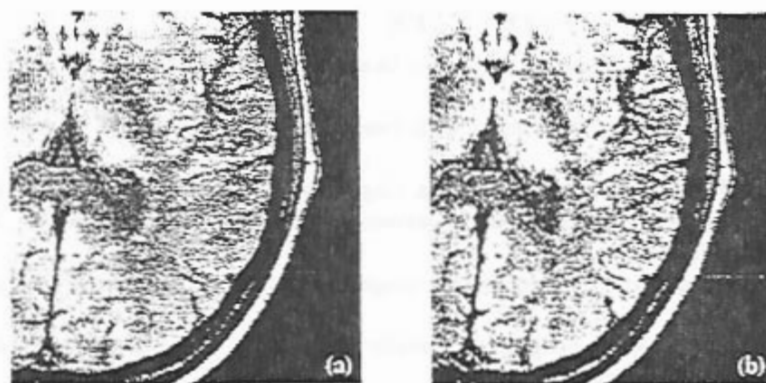Figure 10. *Two brain images concerning single and multiple centers.*

Figure 11. *Particular of Fig.10a: (a)without and (b)with masking. In the first case the center can't be easily distinguished from the adjacent anatomical structure (C.R.=20).*
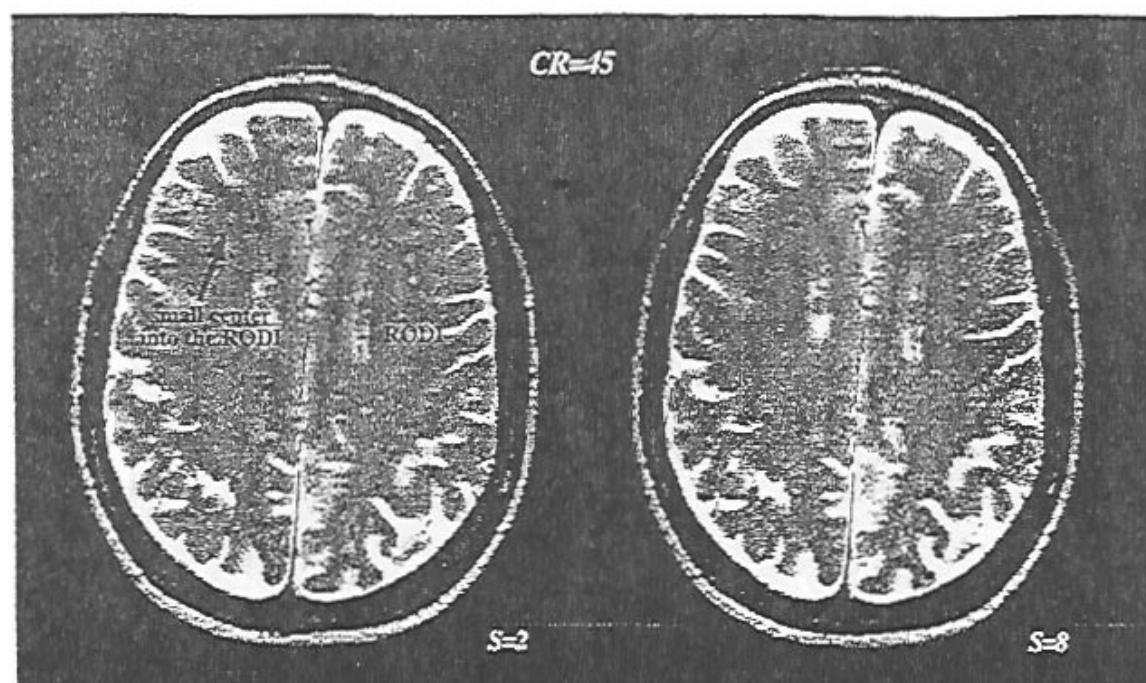
Figure 12. *Example about selectivity: S=2 is correct; S=8 is bad.*

ACKNOWLEDGMENTS

REFERENCES

1. S. Wong, L. Zaremba, D. Gooden, and H. Huang, "Radiologic image compression – a review," *Proc.of the IEEE* **83(2)**, pp. 194–219, Feb. 1995.

2. J. Ström and P. Cosman, "Medical image compression with loss-less regions of interest," *Signal Proc.* **59(2)**, pp. 155–172, Jun. 1997.

3. M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Englewood Cliffs, NJ: Prentice-Hall, 1995.

4. J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. on Signal Proc.* **SP-41(12)**, pp. 3445–3462, Dec. 1993.

5. O. Egger and M. Kunt, "Embedded zerotree based lossless image coding," *Proc. of the ICIP-95* , pp. 616–619, 1995.

6. E. Nguyen Phuc, *Compression sélective et focalisation visuelle: application au codage hybride de séquences d'images*, Ph.D. thesis, Université de Rennes 1 (F), 1995.

7. N. Jayant, J. Johnston, and B. Safranek, "Signal compression based on models of human perception," *Proc. IEEE* **81(10)**, pp. 1385–1422, Oct. 1993.

8. J. Woods and S. O'Neil, "Sub-band coding of images," *IEEE Trans. on Acoust., Speech, and Signal Proc.* **ASSP-34(5)**, pp. 1278–1288, May 1986.

9. M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. on Image Proc.* **IP-1(2)**, pp. 205–220, Apr. 1992.

10. A. Signoroni, *Codifica basata su Regioni di Interesse mediante Trasformata Wavelet: applicazioni alle Immagini Biomediche.*, Thesis of Laurea, University of Brescia (Italy), Jun. 1997.