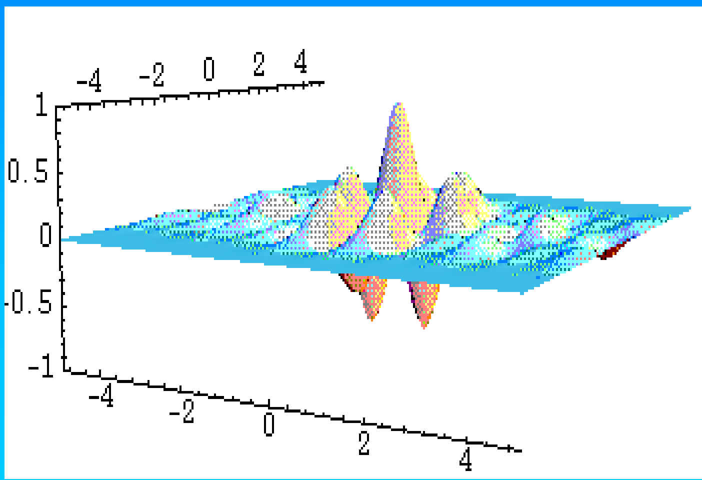
“Wavelet analysis for image fusIon”



A thesis presented

by

Livanos George

To

The Department of Electronics and Computer Engineering

in partial fulfillment of the requirements

for the degree of Diploma

Technical University of Crete

Supervisory Committee:

Supervisory Committee:	**Professor**	**Michalis Zervakis (Supervisor)**
	Professor	**Nikolaos Sidiropoulos**
	Associate Professor	**Athanasios Liavas**

July 2005

“Wavelet analysis for image fusIon”

by

Livanos George

Approved:

Michalis Zervakis
Thesis Director

Nikolaos Sidiropoulos

Athanasios Liavas

D E D I C A T I O N

This thesis is dedicated to my family

Abstract

Image fusion constitutes a basic subset of a more general concept named as combination-unification of information (known by the term information fusion) and has become a basic tool for a large number of applications. Fusion of images (or image fusion) is the process of combining two or more images into a single one retaining important features from each and is an active subject of research with application in medical and satellite images, quality control and machine vision.

In the general area of image processing, wavelet analysis constitutes a powerful mathematical tool. Wavelets are simply a set of functions that satisfy certain conditions and are able to fully represent and reconstruct input signals via a mathematical model. Wavelet analysis was developed independently in the fields of lots of knowledge areas. The interchanges between these fields during the last fifteen years have led to many new applications, including image processing.

In this thesis, an effort is made to combine these two concepts. We present four basic wavelet transforms and discuss recent results on the use of these algorithms for image fusion. The advantages and disadvantages of each method are taken into consideration and a detailed analysis of evaluation results is performed in order to compare these wavelet-based fusion algorithms. We implement and test the Discrete Wavelet Transform (DWT), two improved expanded versions of it, the Shift-Invariant Discrete Wavelet Transform (SIDWT) and the Dual Tree Complex Wavelet Transform (DT-CWT) and finally the Mallat-Zhong Discrete Wavelet Transform (MZ-DWT). Our application of wavelet analysis in image fusion is focused in medical images (Computed Tomography and Magnetic Resonance Images). For objective comparison, we extract qualitative and quantitative results through a well defined performance metric.

Approved:

Michalis Zervakis

Thesis Director

Acknowledgements

I would like to express my appreciation to Dr. Michalis Zervakis for the great amount of time, effort, support and mainly his patient guidance throughout the time that I have been working on this thesis. I also thank Dr. Nikolaos Sidiropoulos and Dr. Athanasios Liavas for reading and correcting my thesis

Contents

P r e f a c e

Image fusion constitutes a basic subset of a more general concept, not very recent but certainly rapidly developing, named as combination-unification of information (known by the term information fusion) and has become a basic tool for a great number of applications. Fusion of images (or image fusion) is an active subject of research with application in medical and satellite images, quality control and machine vision. Although it belongs to the more general area of information fusion, it has particular and unique characteristics that differentiate it because of the complexity of the nature of image understanding, leading to the development of special methods for image fusion.

On the other hand, wavelets constitute a powerful mathematical tool, rapidly developing and applicable to a great number of scientific areas. When introduced, not very recently, it was thought to bring "fresh air", to become a revolution, in mathematics and change the scientists' mindset about data processing and computation. They were developed independently in the fields of lots of knowledge areas. The interchanges between these fields during the last fifteen years have led to many new applications, including image processing, our point of interest.

In this thesis an effort is attempted to introduce some basic wavelet transforms and present recent results on the use of these algorithms for image fusion in particular. The advantages and disadvantages of each method are mentioned, the theoretical background of each algorithm is briefly analyzed and a detailed analysis of evaluation results is presented in order to compare these methods and make things clear to the reader. In this report, the application of wavelet analysis in image fusion takes place in medical images (Computed Tomography and Magnetic Resonance Images) but the general fusion scheme can be expanded to other image formats too. In this case the qualitative and quantitative results and conclusions might be different, as image fusion algorithms are very "objective": they depend very much on the kind of application used.

It 's worth mentioning that most terms in this subject have been established in English in international bibliography and have no precise corresponding in other languages, so any possible translation attempted can be trial.

Terminology

IF	Information fusion
CT	Computed tomography
MR	Magnetic resonance
MRA	Multiresolution Analysis
DWT	Discrete Wavelet Transform
SIDWT	Shift Invariant Discrete Wavelet Transform
DT-CWT	Dual Tree-Complex Wavelet Transform
DDWT	Dyadic Discrete Wavelet Transform
CWT	Continuous Wavelet Transform
MZDWT	Mallat Zhong Discrete Wavelet Transform
HH	High-High
HL	High-low
LH	Low-High
LL	Low-Low
MS	Maximum Selection
WA	Weighted Average
WBV	Window based verification
RMSE	Root Mean Square Error

Introduction

The availability of multi-sensor systems in key regions of image applications today, such as battlefield target recognition, automated inspection, robot automatic navigation e.t.c., has encouraged researchers to work in multi-sensory image fusion..

Each picture is a compressed representation of real world. Fusing images having redundant information we acquire a more complete comprehension of reality. However, due to the imperfections in image sources and in the internal complexity of image understanding, fusion requires the maintenance of redundant information and the removal of inconsistencies in the initial source images, preserving at the same time a valid interpretation of these.

The term image fusion is translated in the bibliography as "the combination of two or more images for the creation of a new image that will contain more information using a given algorithm" [1].

Depending on the level of description in which information is fused, image fusion takes place in the following levels:

- **Pixel-level,** also known as picture or data or signal level. Fusion of images in this level is the process of combining two or more spatially registered pictures in an “enriched” one. Images are described in the spatial or frequency domain. The characteristics in each separate initial source image should be preserved or enriched in the fused image and any artifact should be avoided. Multiresolution techniques have a particular application in this category.
- **feature level** .In this level the initial images are described by edge or region maps, shape feature values, fuzzy measures, probabilities e.t.c. It requires algorithms in order to recognize objects based mostly on the statistical characteristics of dimension, shape, edges, and regions. Segmentation algorithms have been proved useful.
- **Decision level,** also known as symbol level. It separately processes the entry images in order to derive information and applies decision rules to achieve common interpretation and remove the differences.

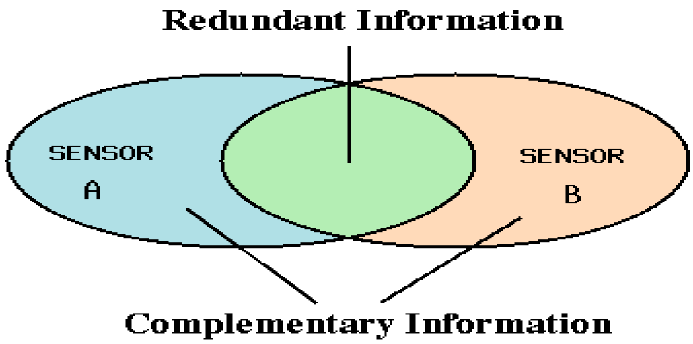
Image fusion algorithms can be categorized according to the above approach. An analytic presentation of this categorization is described in the Appendix, in the first part A:” Categorization of fusion methods”.

Up to day a great effort of research is focused in the development and implementation of various fusion algorithms, most of which are suitable for concrete applications. A more general methodology is required for planning and evaluating image fusion techniques. Even more, an objective criterion-measure is desirable in order to evaluate the various algorithms, in a general scale, independent from the application in which they are used. In this thesis we’ll use the wavelet analysis, which belongs to pixel-level fusion techniques, in order to implement the image fusion procedure.

Definition: image fusion can be defined as the assimilation of information acquired by two or more sensors viewing the same scene. The result of this procedure can be images, which are not necessarily the visual pictures or symbols that describe the scene. The fused image may only be understood on the basis of some specific knowledge of the sensor data [2].

The complementary information in a picture/image tends to suppress the confusions and the ambiguities of other pictures/images. As an example we can mention medical imaging where Computed Tomography images provide details about the structure of bones while Magnetic Resonance images are informative about blood flow and soft tissue density [3].In order to develop both sets of information we should combine them somehow. The redundancy contained in each picture can be consistent or conflicting, a fact that should be avoided. In general terms, the goal of image fusion is to unify dissimilar data to export more accurate, reliable and useful data. Having in mind the image properties above, image fusion should satisfy the following conditions [3]:

1. increase completeness by providing complementary information
2. limit uncertainty and inaccuracy
3. minimize redundant information
4. resolve conflicting, incompatible information
5. acquire a precise and meaningful representation of scene



Schematic definition of redundancy/complementarity

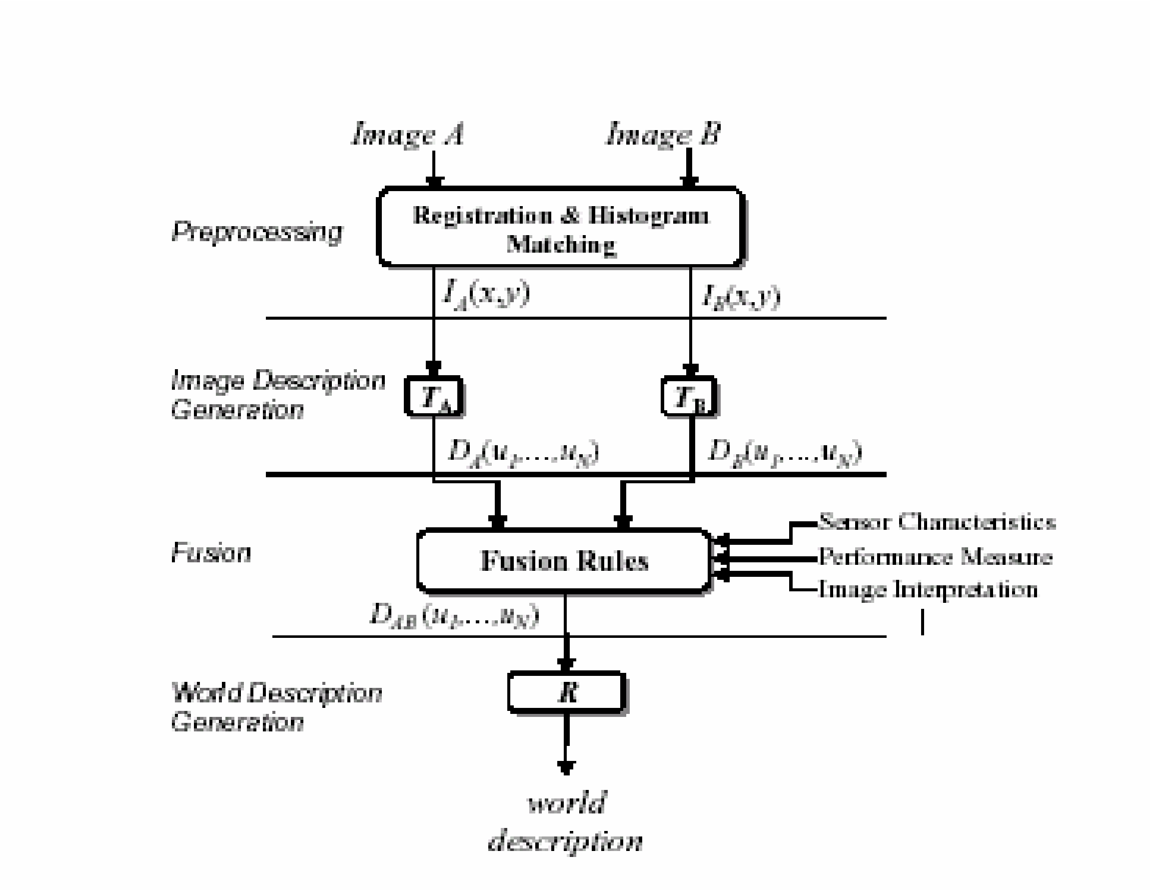
Summarizing the above, image fusion has (or should have) the following attributes-characteristics:

Advantages	Objectives
• Improve the reliability of the results(by using redundant information) • Improve the capabilty of producing different kinds of results according to the needs of applications(by using complementary information) • Reduce the total amount of information without loss of image quality or content data • Focuses on the features we are interested in • Multi-sensor image fusion reduces cost as it combines data from some simple sensors instead of making one complicated and expensive device	• Extract all the useful information from the images • Do not introduce inexistent information(artifacts) and incosistensies • Remain reliable and stable to the imperfections of the images and the sensors(such as mis-registration)

Generally, the problem of image fusion can be summarized in the following processes [3]:

1. preprocessing
2. generation of image description
3. combination of images (fusion)
4. generation of world description

Fusion takes place under a concrete mathematic background, separate and proportional with the technique that is followed. Under this prism, images should be modeled and described in this mathematic model before being processed .After fusion, a moreover treatment is needed to convert the fused result, via some mathematic background again, to a suitable form in order to become analyzable and comprehensible as a description of real world.



Stages of process of image fusion

A concise analysis of the process above follows:

1. Preprocessing

This stage includes the registration of the image and the histogram matching. The scene of each image should aliase each other. If the pictures are not of the same size, interpolation and resampling of the pixels values should be performed in order to acquire the same size. Some techniques do not presuppose the existence of images of equal size provided that the level of analysis of each image is known. Image registration is very important; it constitutes the first basic step for almost all the techniques of image fusion.

However we should take into consideration that errors in the registration process cause errors, even failure in the fusion algorithms, so the stability and robustness of algorithm with respect to mis-registration need to be considered [3]

Since most algorithms in pixel and feature level are numerical, histogram matching is often needed in order to scale the source image pixel values to the same range. On the contrary in the higher level, decision level, the only requirement is an identity association [3].

2. Generation of image description

Depending on the degree of abstraction, an image is described in one of the following levels:

- **pixel-level**, also known as image or data or signal level.

Images can be represented in the spatial or in frequency domain. Raw image data are a 2D or 3D spatial array. This can be converted in the frequency domain, in which the image is decomposed into different levels of analysis. The representation of multiple analysis is named “image pyramid” and the operators for “building” the pyramid include 2D Wavelet transformation, Gaussian, Laplacian and morphological filtering. The two representations are absolutely equivalent. The spatial array can be recovered by inverse transform of the pyramid.

- **segmentation level:**

In this level, the image is described by edge or region maps. The maps are result of segmentation algorithms or edge detectors. The edges can be refined using thinning and linking operations

This level is usually unified with the attaché in order to constitute the more general level, the feature level as it is reported in bibliography.

- **feature level:**

In this level the image is described using a set of feature data in an N-dimensional feature space, where one class of objects is grouped and can be separated from other classes. The feature data are obtained from feature extractor for the regions or the edges in an image.

- **decision level:** It is a symbolic description that receives the forms of symbols, propositions, rules, e.t.c. . The images should be described in the theoretical background in which fusion is executed.

Images have to be described in one of the levels above using the appropriate mathematical background. For example when fuzzy logic approaches are used for fusion, in pixel-level, each pixel is determined by a fuzzy relation depending on its value, while in segmentation level, each edge-point is determined by a relation depending on the gradient at this point.

3. Image fusion

The process of fusion plays the central role in image fusion techniques. The fusion operator is selected according to the physical characteristics of sensors, the performance measure and the interpretation of image [3], generally according to the requirements of application. The same operator does not produce the same quality results in all the applications. Fusion operators can be averaging (weighted), selective (min, max)) or rules from fuzzy logic theory, the probabilistic theory and evidence theory when uncertainty exists.

4. Generation of description in the world

The result of the stage above usually belongs in the same theoretical background and in the same level of abstraction as the generation of description of image in stage 2.The result should be converted or processed farther in order to compose a comprehensible description in the real world.

The following stages of processing are also mentioned in bibliography:

- **Noise removal,** in order to remove any possible inserted noise in the source images.
- **Histogram equalization,** in order to export the details and maximize the content information of the source images.
- **Registration,** in order to make images ideal to be processed.
- **Fusion,** where combination of characteristics takes place.
- **Increase of the output image resolution,** as bigger image size exposes more details.

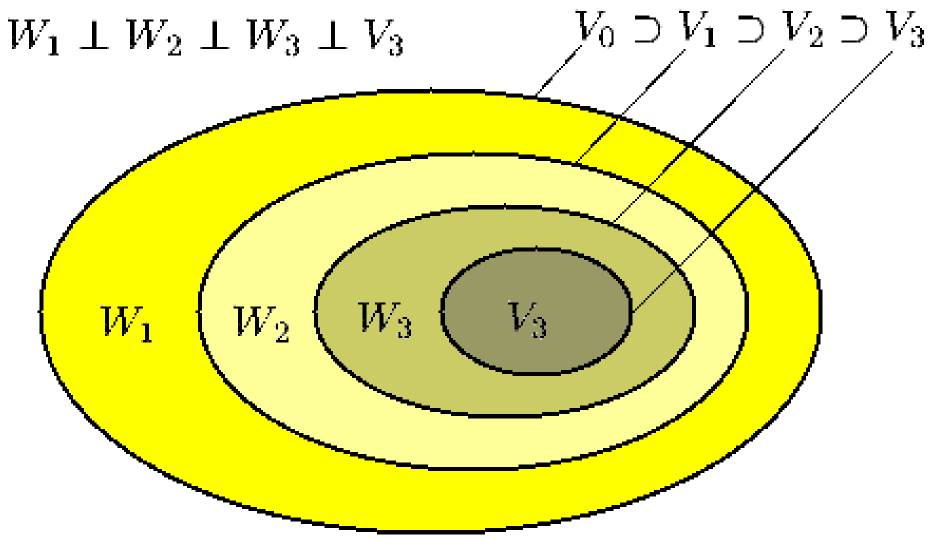
Image fusion techniques are very widely used in the regions of industry, biomedical, in armed forces and in telecommunication services. A synopsis of formal applications is the following [4]:

- **Intelligent robotics:** The applications include intelligent viewing control and automatic target and trajectory recognition. Motion control is required in such applications based on feedback from various types of sensors.
- **Medical imaging:** Images of computed tomography with X-Ray beams and magnetic resonance can be fused to use in computer and image guided surgery.
- **Industry-manufacturing:** Fusion of images is very widely used in the industry for goals of inspection, viewing, assembly and diagnostics.
- **Armed forces and law enforcement:** These applications include detection and recognition of objects, targets or events such as concealed weapons or flight guidance in the night.
- **Telecommunications-remote sensing:** A big part of bibliography is dedicated in this object. The quantity of data to be treated is enormous; therefore the reduction of information is essential and urgent. Fusion techniques are categorized in photographic and numerical methods.

In this thesis, the application of image fusion in medical imaging combining Computed Tomography and Magnetic Resonance images is presented. The technique to be followed is wavelet analysis, a pixel-level based algorithm. For a further study of image fusion techniques refer to the Appendix A

The wavelet transform, wavelet analysis in general, is a powerful tool for multiresolution analysis. During the recent years, with the rapid development of wavelet theory, researchers began to apply wavelet multiresolution decomposition to take the place of pyramid decomposition for the needs of image fusion. Wavelet analysis can be considered as one special type of the pyramid decomposition scheme, retaining most of its advantages but introducing a much more theoretical background.

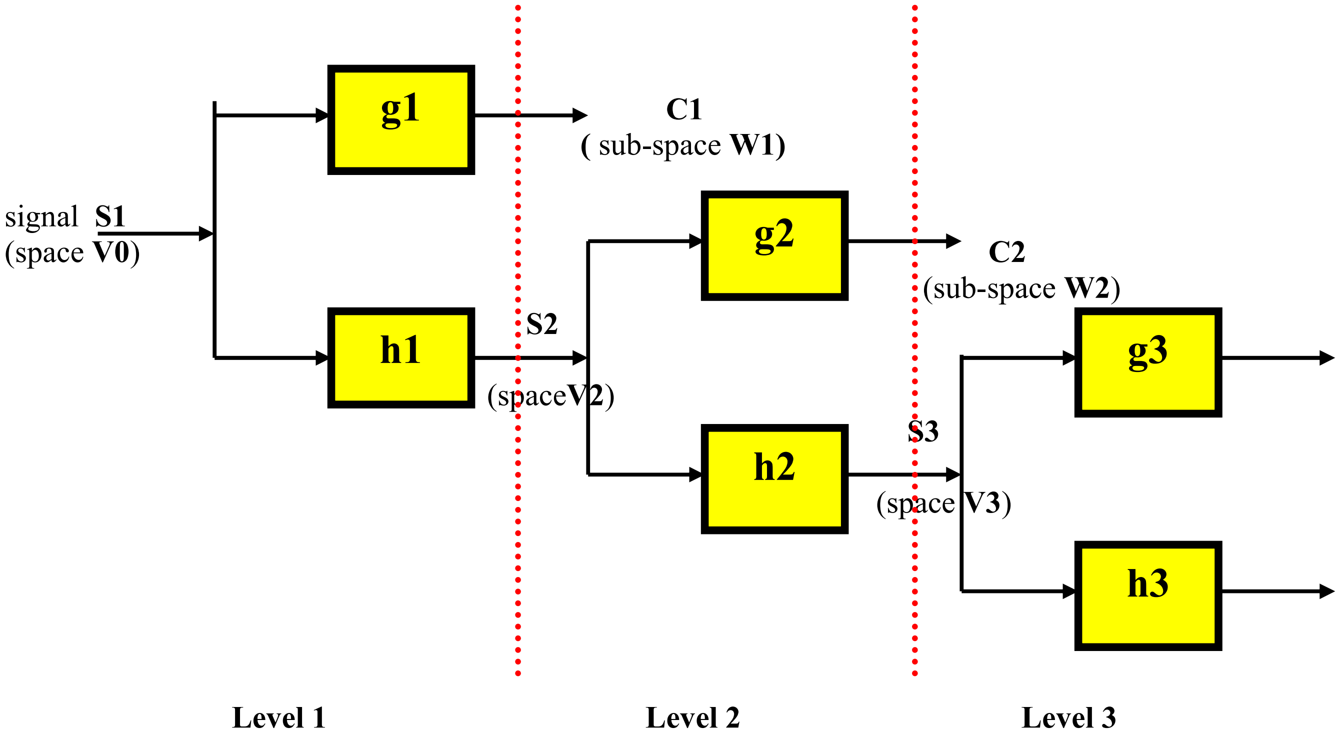
Multiresolution analysis requires a set of cascaded, multiresolution sub-spaces as illustrated below:



General scheme: cascaded multiresolution spaces

The original space V0 can be decomposed into a lower resolution-level sub-space V1, the difference between sub-spaces V0 and V1 can be represented by sub-space W1. Similarly, sub-space V1 can be decomposed into a lower resolution sub-space V2 via sub-space W2 and so on until the last level of decomposition is reached. For an N-level (or stage) decomposition scheme we get N+1 sub-spaces, N spaces Wi and one last, final sub-space VN. This procedure is followed because, in many cases, it's much easier to analyze the decomposed elements than analyze the whole original element itself. [5]

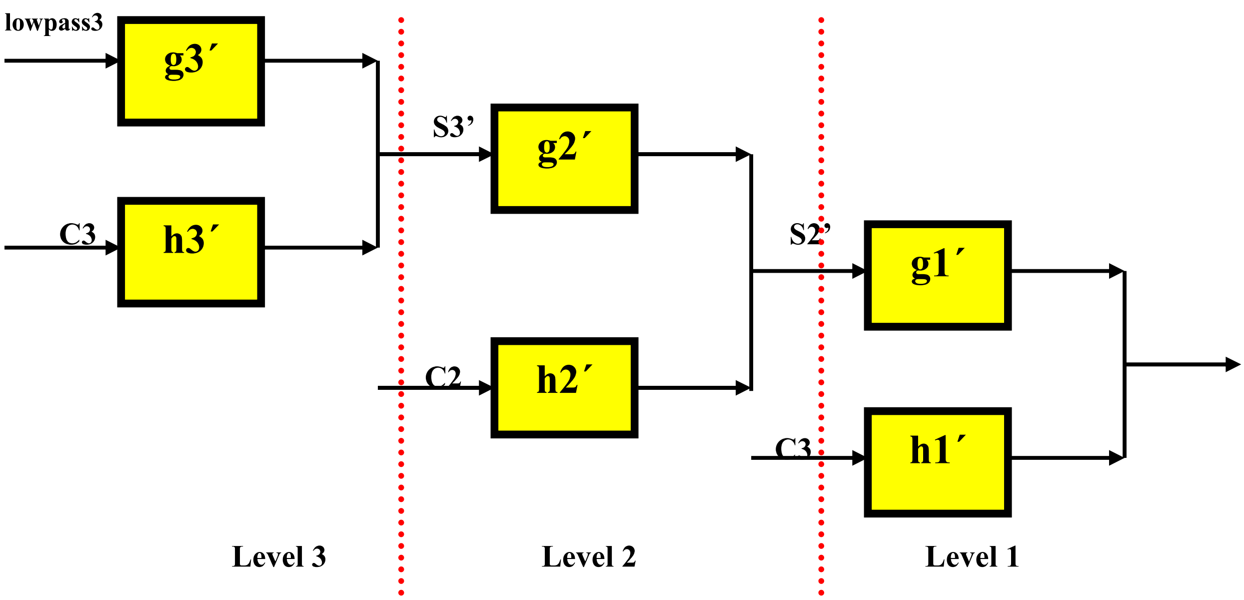
The above general scheme can be applied to signals, in the frequency domain, via a pair of filters, dividing the original signal into two subbands (low frequency and high frequency one) and then, recursively, apply the same procedure to the low frequency subband for the needs of the next level of decomposition according to the following figure:



Signal decomposition via filters

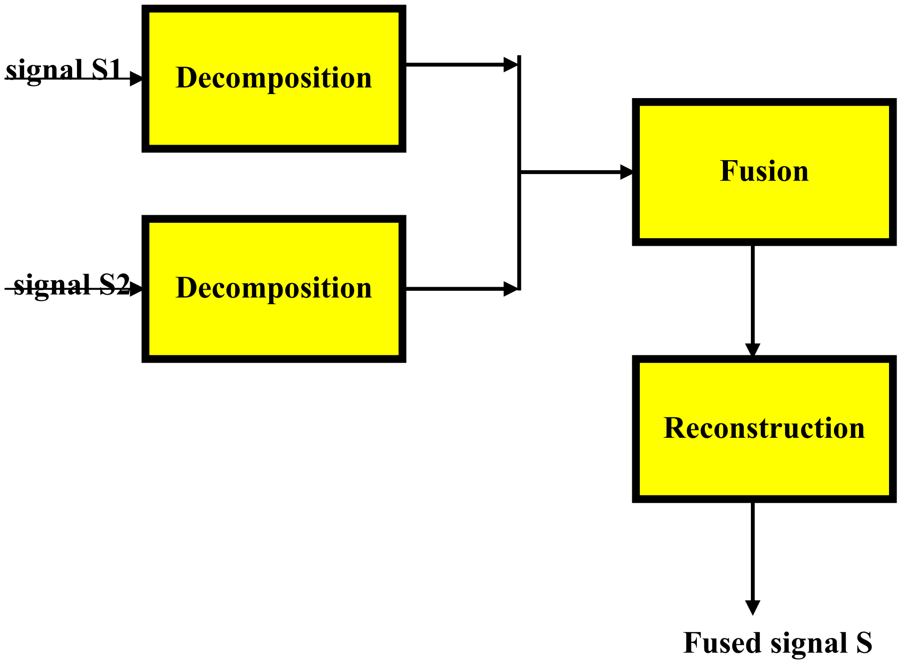
This procedure, in order to be useful and applicable, has to be capable to be reversed, to give the researcher the ability through finite and certain steps to move back and reconstruct the original source signal. The more "similar" the reconstructed signal is the more efficient and preferable the multiresolution algorithm is. That is the goal: to point out and clarify the parameters of the suitable method (.e.g. the filters and the number of levels to be used) to have the best result according to a carefully defined quality metric.

The inverse steps are similar, illustrated in the scheme below:



Signal reconstruction via filters

Having decomposed the input signals, fusion can take place, fusing each corresponding component produced during the decomposition procedure leading to a single, fused element for each subband, at each level. Following the reconstruction procedure the final, fused signal is calculated, containing the relevant, useful information from all the input signals. The fusion implementation is summarized in the following flowchart:



General fusion procedure

Considering the above scheme the implementation is left to the researcher, having in mind the following issues:
-which decomposition algorithm should be used?
-what are the best parameters of the decomposition mechanism?
-which fusion rule should be used?
-how can the performance of the methods be evaluated objectively? Which quality metrics should be followed?

In the following chapters an effort is made to answer the above questions by using wavelet analysis as decomposition/reconstruction algorithms in order to fuse 2-D signals, in particular medical images. The evaluation of the methods, the comparison of the results and the extraction of conclusions are of great importance.

Chapter 1

Wavelets

The wavelet transform creates a summation of elementary functions (= wavelets) from arbitrary functions of finite energy [7]. Wavelets are functions that satisfy certain mathematical requirements and are used to represent data or other functions. They are considered as an alternative, a development of the Fourier transform, which played an essential role in the evolution of the way functions were processed in mathematics. The basic idea that lies behind wavelets is to analyze data according to scale (or resolution). For many years mathematicians were looking for more appropriate functions than sines and cosines (the basis of Fourier transform) to approximate signals. The arrival of wavelets tended to offer two basic advantages against these: a) they don't stretch out to infinity and b) they are suitable in representing data of sharp discontinuities [7].

The wavelet analysis procedure is to define a suitable prototype function, named as **analyzing wavelet** or **mother wavelet** or **basis function**. Two "versions" of the mother function are obtained: one low-frequency (known as **scaling function**), used for temporal analysis and one high-frequency function (known as **wavelet function**) used for frequency analysis. Using coefficients in a linear combination of these functions we can fully represent the source signal. So, data operations can be performed using only the corresponding wavelet coefficients, which, of course, need to be suitably chosen and play an important role in the determination of structure characteristics at a certain scale in a certain location [7]. One of the goals of changing data through this transformation is to bring it in a form that is more readily and easily interpretable than it was before. Once we have determined the coefficients we can manipulate them in place of the original function, since the coefficients uniquely determine this function. If we are clever in selecting the basis function, many of the coefficients will be small, so that we can neglect these terms and represent the original data with just a few wavelet coefficients.

The wavelet transform is a two-parameter expansion of a signal in terms of a particular wavelet basis function. If $\psi(x)$ represents the mother wavelet, all other wavelets are computed by the following equation:

$$\psi_{\alpha,\tau}(t)=(\frac{1}{\sqrt{\alpha}})\psi(\frac{t-\tau}{\alpha})$$

where the scaling factor α, usually $\alpha=2^{-j}$, defines the scale and the factor τ, usually $\tau=kT2^{-j}$, defines the dilation. In this case the wavelet basis functions are obtained by

$$\psi_{j,k}(t)=(2^{j/2})\psi(2^{j}t-kT)$$

The parameterization of time/space by integer k and the frequency/scale by integer j turns out to be very effective. Selecting different values of j, k the different wavelets are computed.

The multiresolution representation of a signal needs two closely related basic functions. In addition to the **wavelet function** *$\psi(t)$,* there is a need for another basic function called the **scaling function**, which is denoted as *$\varphi(t)$* and is derived in a similar way to the wavelet function

$$\varphi_{j,k}(t)=(2^{j/2})\varphi(2^{j}t-kT)$$

As a result the wavelet expansion for a signal x (t) is given by the following decomposition series in which the scaling and wavelet functions are utilized

$$x(t)=\sum_{k} a_k \varphi_{j_0,k}(t)+\sum_{k,j} c_{j,k}\psi_{j,k}(t) \quad ,$$

where coefficients a_k are referred to as *approximation* or *lowpass* coefficients at original scale j_0 and coefficients $c_{j,k}$ are referred to as *detail* or *highpass* coefficients at scale j. As noticed, the scaling function produces the lowpass information and the wavelet function produces the highpass data of the signal. As it will be shown in the next chapter, instead of using mathematical functions the calculation of the coefficients is generally formulated in terms of a particular set of filters, whose structure is obtained from the knowledge of the mother wavelet and the scaling function.

A real or complex-value continuous-time function ψ (t) satisfying the following properties is called a wavelet:

1. $\int_{-\infty}^{\infty} \psi(t)dt = 0$, which means that ψ (t) has to be oscillatory, it must be a wave
2. $\int_{-\infty}^{\infty} |\psi(t)|^2 \, dt \langle \infty$, which means that ψ (t) must have finite energy

Wavelets can be divided in two categories:

- **Orthogonal wavelets**, whose coefficients are given by the equation

  $$c_{j,k} = x(t)\psi_{j,k}(t)dt \quad , \quad a_k = x(t)\varphi_{j,k}(t)dt$$

  and the orthogonality property is satisfied

  $$\psi_{j,k}(t)\psi_{m,n}(t)dt = \begin{cases} 1 & \text{if j=m and k=n} \\ 0 & \text{otherwise} \end{cases} \quad ,$$

  $$\varphi_{j,k}(t)\varphi_{m,n}(t)dt = \begin{cases} 1 & \text{if j=m and k=n} \\ 0 & \text{otherwise} \end{cases} \quad ,$$

  $$\varphi_{j,k}(t)\psi_{j,k}(t)dt = 0$$

  In this case the analysis and synthesis filters are not symmetric, which is very useful in image processing. In addition, the order of the filter is always an even number.
- **Biorthogonal wavelets**, where the wavelet and scaling functions appear in pairs $\psi(t),\psi'(t)$ and $\varphi(t),\varphi'(t)$. In this case the one set $\varphi(t),\psi(t)$ is used for the analysis/decomposition of the signal and the other pair $\varphi'(t),\psi'(t)$ is used for the synthesis/reconstruction of the final signal. Due to more flexibility in this case, the analysis and synthesis filters can be forced to be symmetric. The order of the filters, unlike the orthogonal case, can be an even or an odd number.

Attempting a historical flashback, wavelets came out when mathematicians were led from **frequency analysis** (Fourier transform) to **scale analysis**, that is, analyzing functions by creating mathematical structures that vary in scale. The procedure is simple: define a function, shift it by some amount, change its scale and apply it in order to approximate the source signal. Repeat this procedure: take the structure created, shift it, change its scale and get a new approximation of the input signal. And so on, until the predefined number of scales is reached. The advantage of scale analysis is that it turns out to be less sensitive to noise.

In the 1930s, the term **scale-varying basis functions** was introduced (see Appendix B for more details), a key point in understanding wavelets. It was found that the so called Haar basis function was proved more efficient than Fourier basis function for special applications. At the same time, the definition of the energy of a function and the fact that its computation produced different results depending on the interval the energy was distributed led to the need to discover a function that can vary in scale but be able to conserve energy: wavelets.

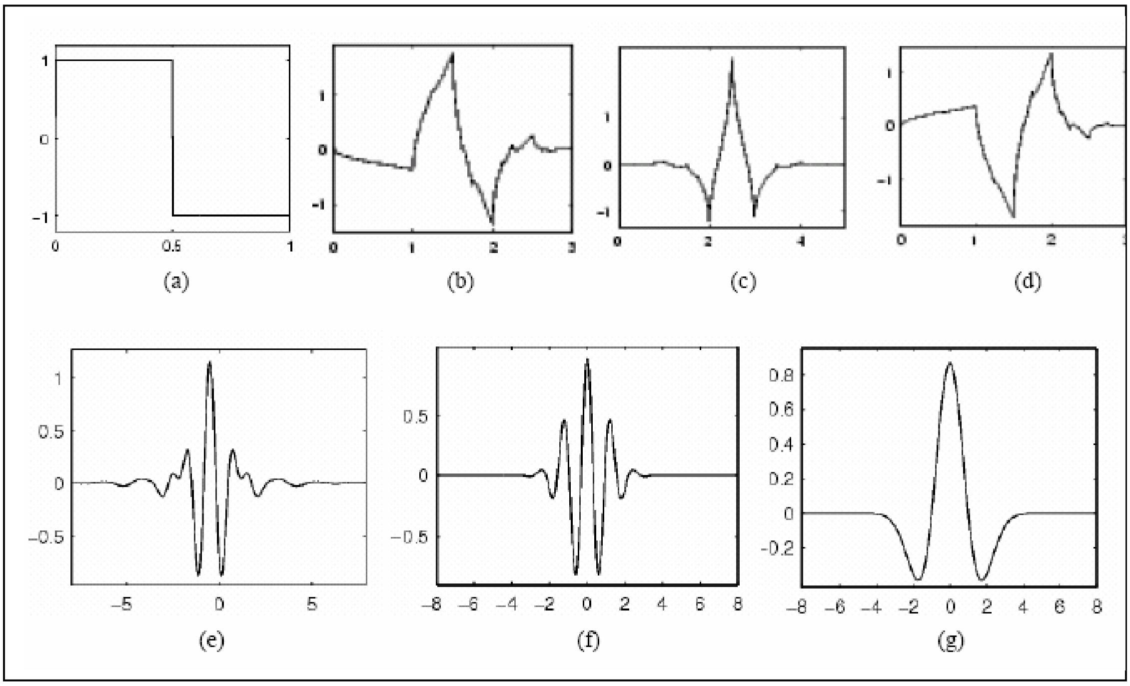
In 1985, Stephane Mallat, the establisher of wavelets, gave them a great impulse through his research in digital signal processing, combing filters, pyramid algorithms and wavelet bases. His work inspired other researchers to define their own wavelet functions, each used in different knowledge areas, ending to Ingrid Daubechies, who constructed a set of wavelet orthonormal basis functions, the cornerstone of wavelet applications today.

Fourier and wavelet analysis have some very strong common links but also some basic dissimilarities. They are both linear operators, they can be viewed as a rotation in function space to different domain: sines and cosines for Fourier transform and mother wavelets for the wavelet transform. Finally, they are both localized in frequency, which is useful in calculating power distributions.

On the other hand, wavelet functions are localized in space while Fourier sine and cosine are not, a property that is proved useful in data compression, feature detection and noise removal. Wavelets are negligible outside some interval, meaning that outside some interval they either vanish or decay exponentially. Exponential functions, vice versa, are not negligible in the whole real-area, so Fourier Transform cannot be used for local data analysis. In this way, wavelets provide the ability to obtain frequency representation in a small interval of our data independent of the data outside of that interval. Due to this "locality" wavelet basis functions offer a much more efficient way of approximating discontinuities. Another advantage of wavelets is that we can adjust the wavelet transform parameters by combining different basis functions. For example, in order to obtain detailed frequency analysis we need long basis functions while in order to obtain signal discontinuities we would like short basis functions. The above can be achieved by having short high-frequency mother wavelets and long low-frequency ones. This is the most powerful property of wavelets. Finally, the efficiency of calculating the coefficients in the wavelet expansion is greater than the one of the Discrete Fourier Transform [14].

Concluding, it is worth mentioning that wavelet transforms do not have certain, predetermined basis functions, such as sines and cosines. There is an infinite set of basis functions to be used. And this is the difficulty and the goal at the same time: which basis should be used in order to have the best result for a particular application? The different wavelet families make different trade-offs between how compactly the basis functions are localized and how smooth they are. Some common wavelet families are illustrated below (see Appendix C for more details on how wavelets look like).

Wavelets are used in a great number of applications. Some of them include computer and human vision, fingerprint compression, denoising noisy data, musical tones e.t.c. In this report we will focus on the use of wavelet analysis in image fusion, presenting the most well-known and efficient wavelet transforms met in bibliography.



Wavelet families (a) Haar (b) Daubechies4 (c) Coiflet1 (d) Symlet2 (e) Meyer (f) Morlet (g) Mexican Hat.

Any discussion of wavelets starts with the Haar wavelet while Daubechies wavelets are the most popular. In most wavelet transform applications it is required that the original signal be nearly ideally synthesized from the wavelet coefficients. This condition is referred to as perfect reconstruction. The first four are capable of perfect signal reconstruction, the last three are symmetric. As mentioned, the choice of the suitable wavelet family is application dependent. For example, in image processing it is very desirable to use symmetric wavelets as they make it easier to deal with the boundaries of the image and, in addition, human vision is more tolerant to symmetric error than asymmetric one. By comparison, Coiflets wavelets are closer to symmetry, though not perfectly symmetric. Perfect symmetry is possible only for complex wavelet filters and biorthogonal wavelets [15]. For most applications it is desirable to have real filter coefficients, so the only choice for the class of symmetric wavelets would be biorthogonal wavelets.

For a more detailed presentation of wavelet families it is recommended to refer to the Appendix C:"Wavelet families".

Chapter 2

Perfect reconstruction transforms

Several wavelet based techniques for image fusion have been met in bibliography. Some of them require an advanced mathematical background, others have poor performance, others are very specialized. After a careful research, this report focuses on four basic wavelet transforms, fully implementable and applicable to medical images.

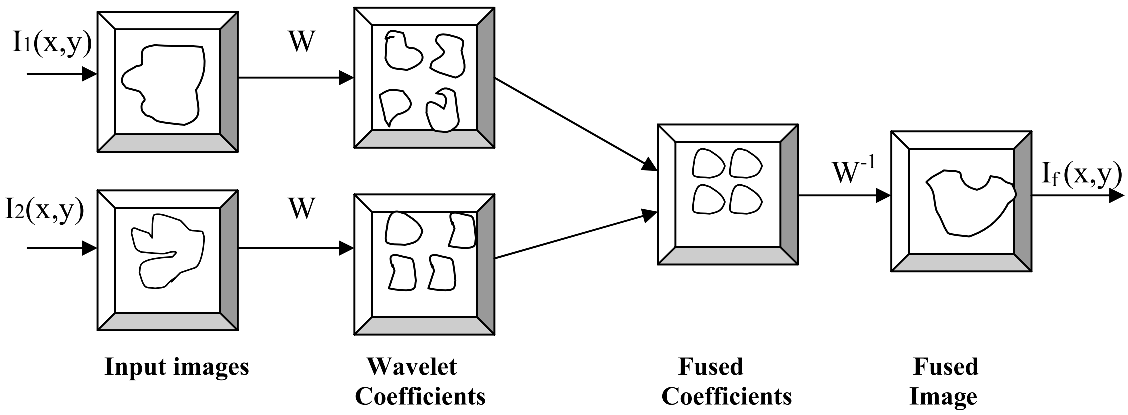
The general idea that lies behind all wavelet based fusion schemes is simple:

-the wavelet transforms W of two fully registered images $I_1(x,y)$ and $I_2(x,y)$ are computed

-these transforms are fused using a suitable fusion rule φ

-the inverse wavelet transform W^{-1} is computed and the fused image $I_f(x,y)$ is reconstructed

$$\mathbf{I_f(x,y) = W^{-1}\Big(\varphi \Big(W\,(I_1(x,y))\,,\, W\,(I_2(x,y))\Big)\Big)}$$



The advantages of wavelet based fusion schemes against similar pyramid fusion schemes are summarized in the following:

- They provide directional information, as data are processed in both vertical and horizontal orientation and in some versions in the diagonal orientation.
- In pyramid based image fusion, the fused images often contain artifacts where the input images are very different, while in wavelet based fusion such imperfections are not present.
- Images fused using wavelet transforms are less sensitive to noise [9].

2.1 Discrete Wavelet Transform

The discrete wavelet transform (DWT) is a spatial-frequency decomposition that provides a flexible multiresolution analysis of an image [9]. In one dimension it is simply a linear combination of wavelets in order to represent the source signal f(x)

$$f(x) = \sum_{m,n} c_{m,n} \psi_{m,n}(x)$$

where $\boldsymbol{\psi_{m,n}(x)}$ is the scaled version of the mother wavelet $\boldsymbol{\psi}$

$$\psi_{m,n}(x) = 2^{-m/2}\psi(2^{-m}x - n) \qquad \textbf{(1)}$$

m, n integers and $\mathbf{c_{m,n}}$ suitably selected coefficients.

In the general case, if ψ(t) represents the mother wavelet, all other wavelets are computed by the following equation:

$$\psi_{\alpha,\tau}(t)=(\frac{1}{\sqrt{\alpha}})\psi(\frac{t-\tau}{\alpha}) \qquad \textbf{(2)}$$

where the scaling factor α, usually $\alpha=2^j$, defines the scale and the factor τ, usually $\tau=kT2^j$, defines the dilation. In this case the wavelet basis functions are obtained by $\psi_{j,k}(t)=(2^{-j/2})\psi(2^{-j}t-kT)$. Setting $\alpha=2^m$ and n=kT in (2) we obtain the transform of equation (1)

$$\psi_{m,n}(t)=2^{-\frac{m}{2}}\psi(2^{-m}t-n)$$

and the source signal f(x) is represented by the equation

$$f(x) = \sum_{m,n} c_{m,n} \psi_{m,n}(x)$$

which is known as the **Dyadic Discrete Wavelet Transform (DDWT)** because the scale factor α is now assigned a value that is a power of 2. In the general case, not the absolutely correct, when referring to the Discrete Wavelet Transform we mean the Dyadic Discrete Wavelet Transform. This notation is used, unfortunately, in international bibliography causing a small compatibility problem.

The variables m and n are integers that scale and dilate the mother function ψ to generate wavelets. The scale index m indicates the width of the wavelet and the location index n indicates its position. Notice that the mother functions are rescaled, or "dilated" by powers of two, and translated by integers. What makes wavelet bases (=the mother functions used to generate wavelets) especially interesting is the self-similarity caused by the scales and dilations. Once we know about the mother functions, we know everything about the basis. A further explanation of basis is presented in Appendix B.

For an iterated wavelet transform extra coefficients $\mathbf{a_{m,n}}$ are required at each scale, where $\mathbf{a_{m,n}}$, $\mathbf{a_{m-1,n}}$ approximate the signal at resolution 2^m and 2^{m-1} respectively, while coefficients $\mathbf{c_{m,n}}$ represent the difference between one approximation and the other. The coefficients $\mathbf{a_{m,n}}$ at each scale and position are computed by using a scaling function applied repetitively to the signal through the convolution procedure. This function is usually implemented through a low pass filter h_n according to the relation

$$a_{m,n} = \sum_k h_{2n-k} \cdot a_{m-1,k}$$

while coefficients $\mathbf{c_{m,n}}$ are calculated by repetitively applying a high pass filter g_n to the signal according to the relation

$$c_{m,n} = \sum_k g_{2n-k} \cdot a_{m-1,k}$$

where the lowest level coefficient $\mathbf{a_{0,n}}$ is the original signal.

It is helpful to think of the coefficients $\mathbf{a_{m,n}}$, $\mathbf{c_{m,n}}$ as a filter. The filter or coefficients are placed in a transformation matrix, which is applied to a raw data vector. The coefficients are ordered using two dominant patterns, one that works as a smoothing filter (lowpass filter), and one pattern that works to bring out the "detail" information (high-pass filter). Coefficients $\mathbf{a_{m,n}}$ represent the smoothed data version and the coefficients $\mathbf{c_{m,n}}$ the detailed one. The transformation matrix is applied in a hierarchical algorithm, sometimes called a pyramidal algorithm. The coefficients are arranged so that odd rows contain an ordering of wavelet coefficients that act as the smoothing filter, and the even rows contain an ordering of coefficients with different signs that act to bring out the data's detail. The matrix is firstly applied to the original, full-length vector. Then the vector is smoothed (the lowpass filter is applied) and decimated by half and the matrix is applied again. Then the smoothed, halved vector is smoothed, and halved again, and the matrix applied once more. This process continues until the final number of index m, that was user defined, is reached (number of stages or scales or levels). That is, each matrix application brings out a higher resolution of the data while at the same time smoothing the remaining data. The same matrix is applied at each resolution.

Having applied the above transform, the reconstruction of the estimated signal is required. The goal is simple: the reconstructed signal to be as similar as possible to the source signal. The procedure to be followed is the inverse of the above, the inverse wavelet transform, which is simply an iterative application of filters to the calculated coefficients $a_{m,n}$, $c_{m,n}$ according to the equation

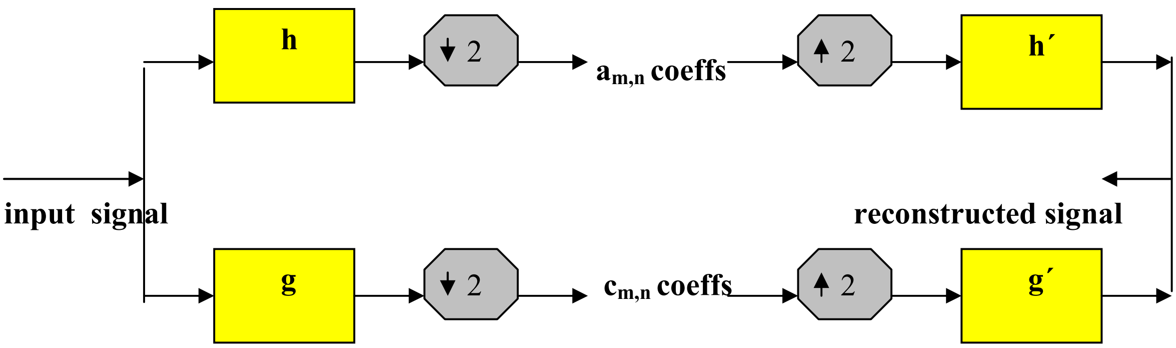
$$a_{m-1,k} = \sum_n \left(g'_{2n-k} \cdot c_{m,n} + h'_{2n-k} \cdot a_{m,n} \right)$$

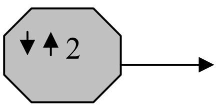
where **g′, h′** are the high pass and low pass synthesis (or reconstruction) filters respectively, capable for perfect reconstruction. **The condition for perfect reconstruction is that the z transform of the output signal is identical to that of the input signal**. Ideally we would like

$$\mathbf{W^{-1}\left(W\left(I(x,y) \right) \right) = I(x,y)}$$

where W^{-1} , W denote the inverse and the forward wavelet transform of the signal I(x,y) respectively.

The following figure summarizes the above:





up/down sampling of the corresponding signal

For perfect reconstruction, filters h, g, h′, g′ should satisfy the following property:

$$H'(z)\cdot H(-z)+G'(z)\cdot G(-z)=0$$

$$H'(z)\cdot H(z)+G'(z)\cdot G(z)=2z^{-d}$$

where $H'(z), G'(z)$ are the z-transforms of the filters h′, g′ respectively and $H(z), G(z)$ are the z-transforms of the filters h, g respectively. The first condition implies that the reconstruction is aliasing-free and the second that the amplitude distortion has amplitude of one. It can be observed that the perfect reconstruction condition does not change if we switch the analysis and synthesis filters.

A key point is the wavelet family to be chosen. As mentioned above, this is a very subjective matter; the selection is application dependent and let to the researcher's ability and "imagination". Any attempt to uniquely determine and suggest the optimum wavelet family (which will also determine the filter structure to be applied) can be misleading. However, it is useful to make a basic categorization of the wavelet approaches that can be followed and set a bread-and-butter question: **Real or Complex wavelets?**

Complex and real wavelets: After a very careful research in international bibliography, it was attempted to compare these two basic wavelet approaches, to get familiar with their characteristics and to make conclusions about their effectiveness in image fusion application. Unfortunately, in most sources found, a complicated mathematical background was introduced and required, which will not (and cannot) be analyzed in this report. A simple approach to real and complex wavelet issues is shortly mentioned in the Appendix C.

Complex wavelets can provide both shift invariance and good directional selectivity, with only modest increases in signal redundancy and computational load. However, development of a complex wavelet transform with perfect reconstruction and good filter characteristics has been proved difficult until recently [16]. Complex wavelets were created to overcome the two key problems of the typical wavelets:

-*Lack of shift invariance*, which means that small shifts in the input signal leads to major variations in the distribution of energy between wavelet coefficients at different scales. The energy distribution between the various levels depends critically on the position of the key features of the signal and not on the features themselves. This is caused by aliasing due to subsampling at each level.

- *Poor directional selectivity*, which means that the diagonal features are not appropriately represented because only three spatial orientations are revealed (vertical, horizontal and all the diagonal directions included in one diagonal orientation)

However, a significant problem arises because the final signal cannot be perfectly reconstructed beyond the first level, when the input becomes complicated.

Complex wavelets have the following properties:

a)the complex wavelet transform has the same structure as described in the previous figure with the difference that the filters have complex coefficients and generate complex outputs. In this case each component produced contains two parts: the real and the imaginary one, introducing a 2:1 redundancy (becomes 4:1 in two dimensions).

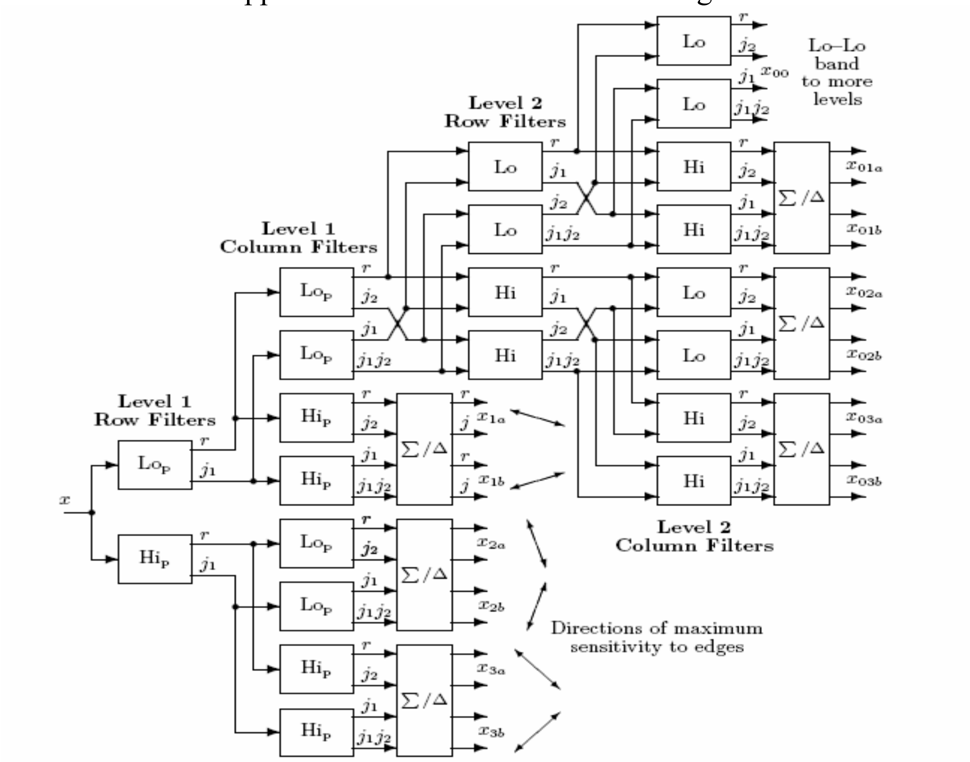
b) the phases of the complex wavelets vary approximately linearly with input shift, which makes the interpolation between consecutive complex samples simple and accurate [16].

Extending complex wavelets to two dimensions is achieved by separable filtering along columns and rows. Additional filtering with complex conjugate filters of the row and column filters is needed, as two adjacent quadrants of the spectrum are required to fully represent a 2-D signal. As a result separate imaginary operators j_1 and j_2, for row and column processing are maintained. This produces four-element complex vectors {a, b, c, d} =a + b j_1 +c j_2 + d j_1 j_2. If j =j_1 = j_2 once and –j = j_1 = -j_2 in the other case then the 4-element vector is divided into a pair of conventional complex 2-element vectors;

vector1 = (a-d) + (b+c)j and vector2 = (a+d) + (c-b)j,

corresponding to the sum and difference operations.

The two-dimensional approach is illustrated in the following flowchart



**The downsampling of the output of each filter is not illustrated in the scheme, but is a key part of the process, in order to make the figure more clearly visible

Complex filters in multiple dimensions provide true directional selectivity, despite being implemented separably, because they are still able to separate all parts of the m-dimensional frequency space. To provide shift invariance and directional selectivity, all of the complex filters should emphasize positive frequencies and reject the negative ones, or vice versa. Unfortunately, it is very difficult to design an inverse transform, based on complex filters as illustrated above. Although such filters can be designed to give perfect reconstruction easily at level 1 by applying the constraint that the reconstructed signal must be real, the same constraint cannot be applied at further levels where the inputs and the outputs are complex. **As an alternative complex wavelet transforms can be modeled, in a “pseudo-way”, by using real wavelet transforms with some changes and additions**. The basic complex wavelet transform is the Dual-Tree Complex Wavelet Transform that will be introduced in section 2.1.3.

To complete the presentation of the general characteristics of the general wavelet transform scheme a short categorization of it is presented.

At first, wavelet transform can be divided into two forms: a) the Continuous Wavelet Transform and b) the Discrete Wavelet Transform.

The Continuous Wavelet Transform is defined as

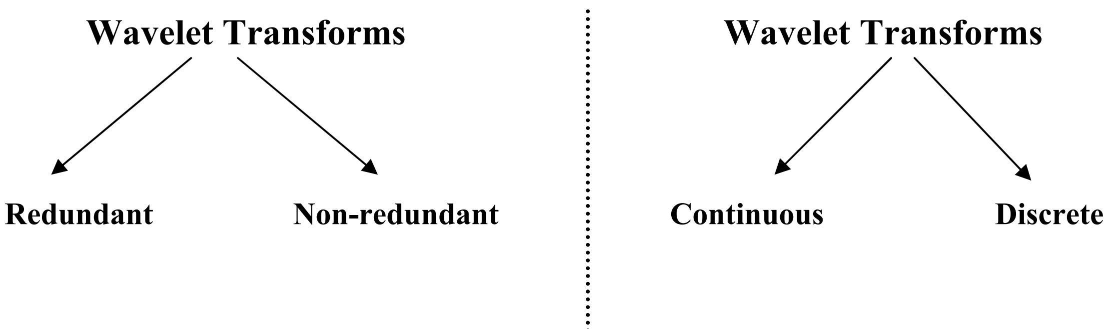
$$X_{WT}(a,\tau) = \frac{1}{\sqrt{|a|}} \int x(t) \cdot \psi^{*}\left(\frac{t-\tau}{s}\right) dt$$

where x(t) is the signal to be analyzed, ψ(t) is the mother wavelet, α is the scale factor, usually $\alpha=2^{j}$ in the case of the Dyadic Wavelet Transform and τ is the dilation factor, which relates to the location of the wavelet function as it is shifted through the signal and corresponds to the time information in the Wavelet Transform. The scale parameter α is defined as 1/frequency and corresponds to the frequency information.

The Discrete Wavelet Transform evolved as an attempt to limit the computational cost of the Continuous Wavelet Transform (CWT). It is easy to implement and reduces the computation time and resources required. It comes from the CWT by discretizing the scale and location parameters. In CWT , the signal are analyzed using a set of basis functions which relate to each other by simple scaling and translation while in the case of the DWT, a time-scale representation of the signal is obtained using digital filtering techniques. The signal to be analyzed is passed through filters with different cut-off frequencies at different scales.

When the energy of the input signal is finite, not all values of decomposition are needed to reconstruct the original signal, provided that we are using a wavelet that satisfies some admissibility condition. In such cases, discrete analysis is sufficient and continuous analysis is redundant [Matlab Documentation, Wavelet Toolbox].

The Wavelet Transforms could also be categorized in **redundant** and **non-redundant**. A wavelet transform is redundant if it carries out the data to be processed in its initial form, without eliminating its size, while non-redundant transforms process input data converting it to a simpler and more abstract form containing less information (e.g. by subsampling the original signal). Redundant representation uses much more scale and position values to achieve accuracy while non-redundant representation offers less complexity with the cost of data loss.



Categorization 1 **Categorization 2**

The main characteristics of CWT and DWT are summarized in the following table:

Continuous Wavelet Transform (CWT)	Discrete Wavelet Transform (DWT)
$C(\alpha,\tau)=\int_R s(t)\frac{1}{\sqrt{\alpha}}\psi(\frac{t-\tau}{\alpha})dt$ s(t)=original signal, C(α,τ)=wavelet coefficients $a \in R^+ - \{0\}, \tau \in R$	$C(\alpha,\tau)=\int_R s(t)\frac{1}{\sqrt{\alpha}}\psi(\frac{t-\tau}{\alpha})dt$ s(t)=original signal, C(α,τ)=wavelet coefficients $a = 2^j, \tau = k2^j, (jk) \in Z^2$
Advantages	**Advantages**
• **Calculations are performed in Fourier space, so frequency is known exactly** • **Offers great accuracy** • **No loss of information** • **Good directionality** • **No limitations on wavelet families selection**	• **Uses discrete values of scale and location** • **Orthogonality removes redundant representations** • **The amplitude of the wavelet coefficients is associated with sharp signal variations** • **Fast computation, simple implementation**
Disadvantages	**Disadvantages**
• **Redundant information** • **Edge effects introduced by FFT** • **Large computational load**	• **Decimation of data** • **Difficult to discern what frequency each level corresponded with** • **Shift-invariance due to missing elements** • **Aliasing of information between the levels** • **Poor directionality**

In the sections that follow the Discrete Wavelet Transform (the Dyadic Discrete Wavelet Transform to be absolutely correct in expression!) is introduced. Three versions of it are analyzed.

- the **regular Discrete Wavelet Transform (R-DWT)**, which is the general wavelet transform met in bibliography
- the **Shift Invariant Discrete Wavelet Transform (SIDWT)**, an improved version of the DWT in order to overcome one of its basic limitations: shift invariance
- the **Dual-Tree Complex Wavelet Transform (DT-CWT)**, an improved and overcomplete, expanded version of the DWT that overcomes all of its limitations

Discrete Wavelet Transform

Regular DWT **Dual-Tree Complex WT** **Shift-invariant DWT**

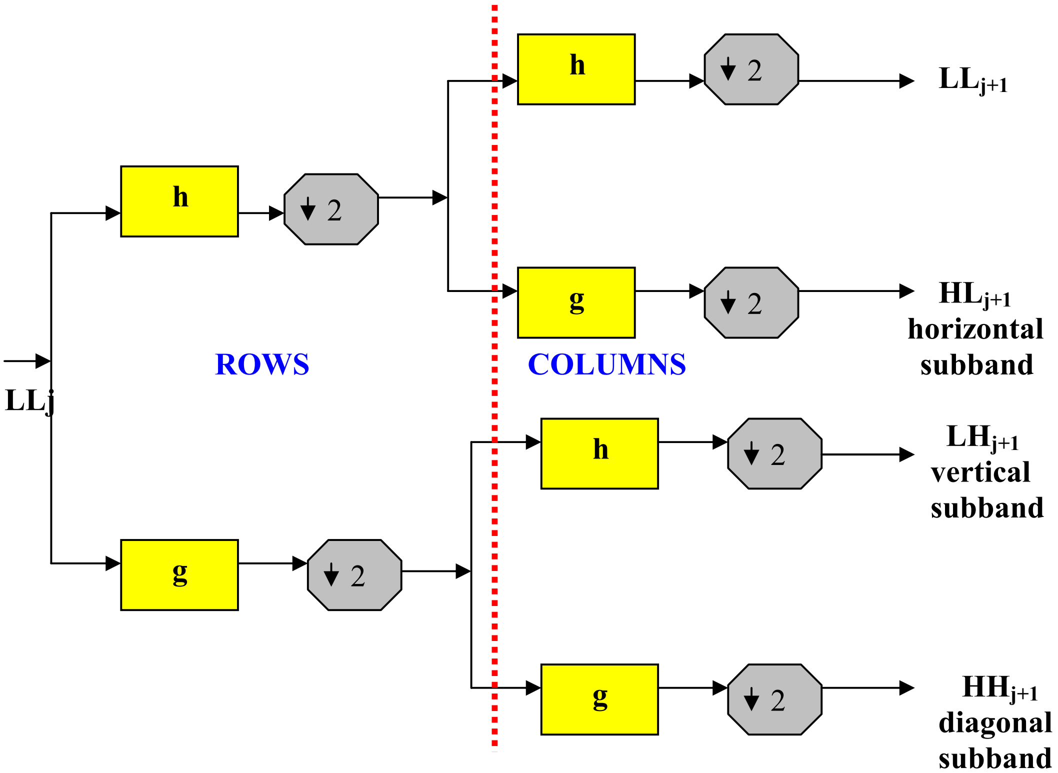
Real DT-CWT **Complex DT-CWT**

In the final section (2.1.4) an attempt to use complex wavelets in the Discrete Wavelet transform is presented, considering all of the limitations described and the difficulty in implementation.

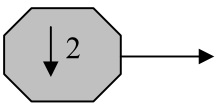
2.1.1 Regular Discrete Wavelet Transform

2.1.1.1 Decomposition of images

The DWT can, of course, be extended to two dimensions, which means to images too, as an image can be considered as a 2-D signal. The procedure is very similar to the above: we use two 1-D wavelet transforms, one for each dimension, horizontal and vertical. As a result, four sets of coefficients (subbands), two from the transform in each direction, are created at each scale/level. Starting from the horizontal frequency first (rows of the image data matrix) and ending to the vertical frequency (columns of the image data matrix) the subbands are obtained; the high-high (HH), the high-low (HL), the low-high (LH) and the low-low (LL) one. The LH, HL, HH subbands contain the vertical, the horizontal and the diagonal frequencies respectively and the LL subband is the source image for the next filter. This counts for each level. By recursively applying filters to the low-low subband, for N steps, called levels, a multiresolution scheme is constructed. Each subband is an image too, with half of the size of its corresponding "ancestor" lowpass image, a downsampled version of it, preserving a special part of its information. The values of the pixels of each subband image are the values computed by the application of the wavelet transform to the corresponding position. The above steps are summarized in the following scheme:

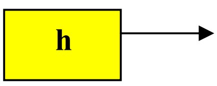


Decomposition procedure at level j

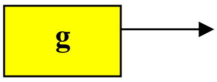


downsampling of the corresponding subimage

j=current decomposition level, **j+1**=next decomposition level



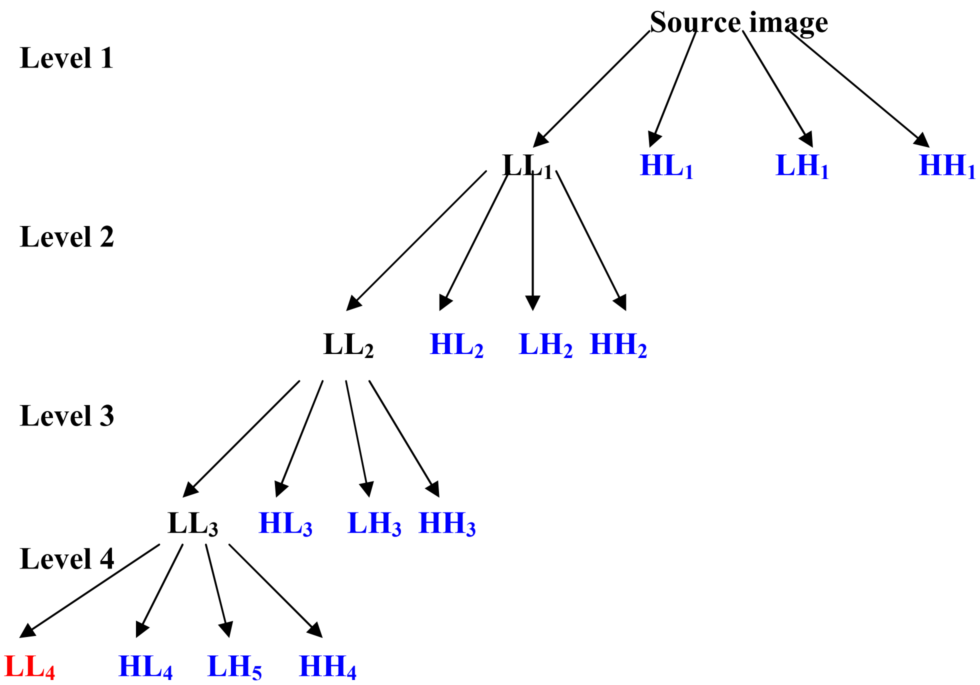
lowpass analysis/decomposition filter, convolution with the input image



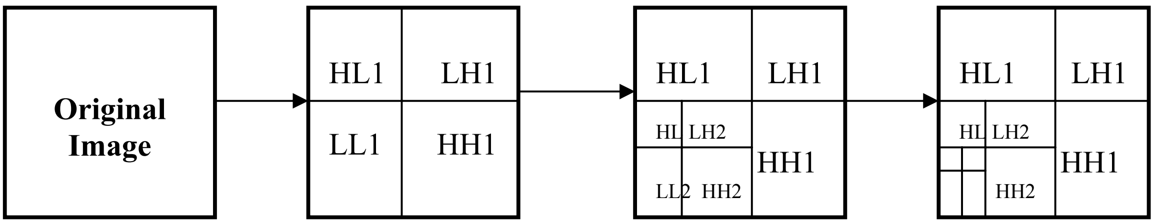
highpass analysis/decomposition filter, convolution with the input image

The filters *h* and *g* give the wavelet coefficients a_j and c_j as described in pages 19, 20.

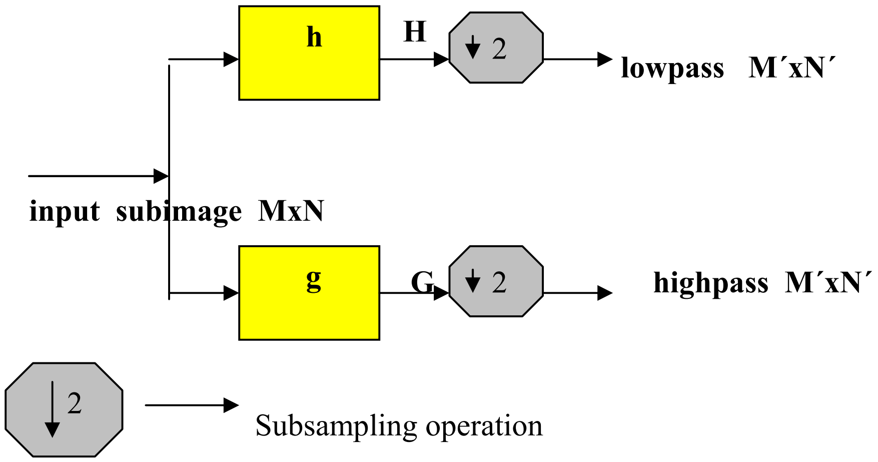
The procedure above decomposes the original image to downsampled subimages, each of which inherits certain characteristics from its "ancestor" and the source image in general. Given the number N of levels, 3 highpass images and a lowpass one are generated at each level. At the end, we have 3*N highpass images and one remaining lowpass image, the one of the last level. At each level the size of the subimages is subdoubled. As a result, the maximum number of levels is limited by the size of the source image; cannot overcome the value $\log_2$ (image size). This decomposition process can be illustrated by the following tree-structure



In this way the image structure, level by level, gets the following sequence:



Remark: It's worth making a notice on the filtering-sampling procedure



According to what was mentioned above, the size of the highpass and lowpass images should be M/2xN/2, which means M'=M/2, N'=N/2.

In fact, according to theory, the convolution of the subimage with the corresponding filter increases the size of the produced image by a certain factor.

If the length of each filter is 2W and n=length(s), where s is a signal in general, the size of the signals H and G is n+2W-1, so the size of the highpass and lowpass outputs is

$$\mathrm{Floor}\left(\frac{n-1}{2}\right) + W$$

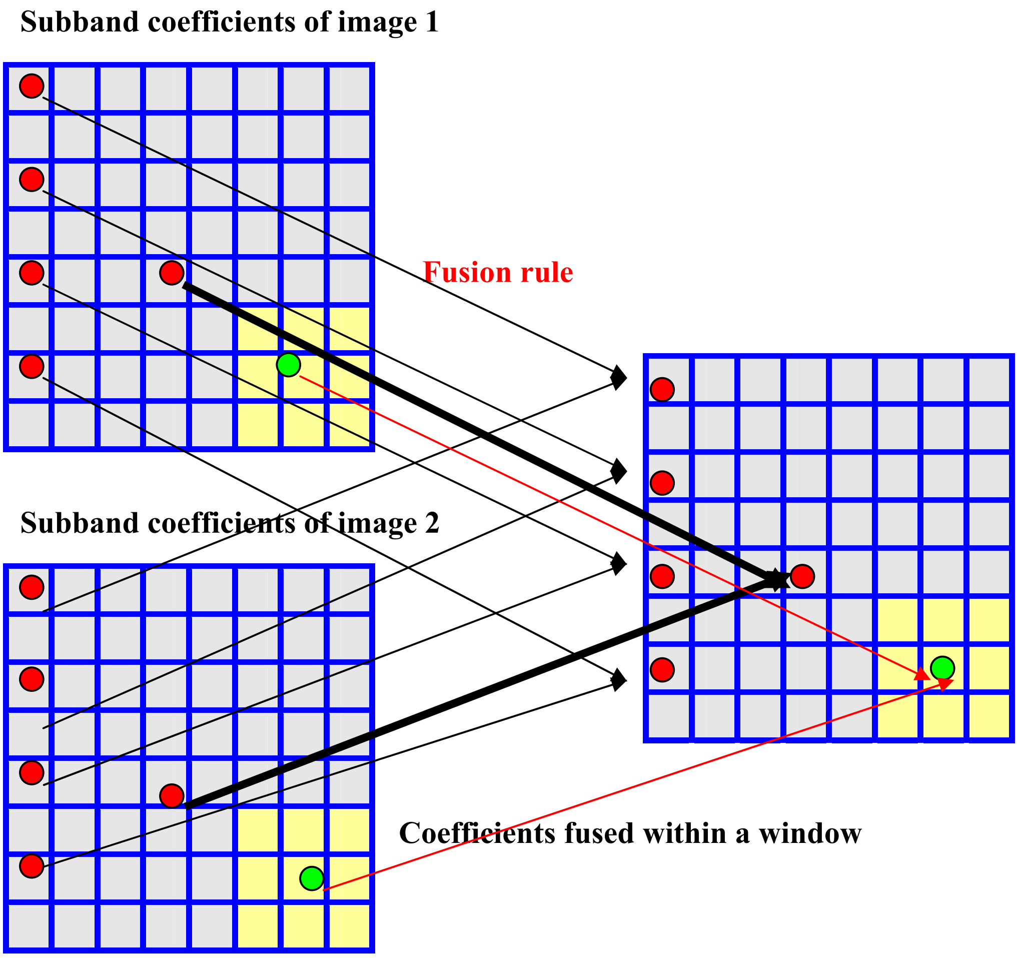
In our case, applying the above conclusion in each direction, the size of the lowpass and highpass subimages is

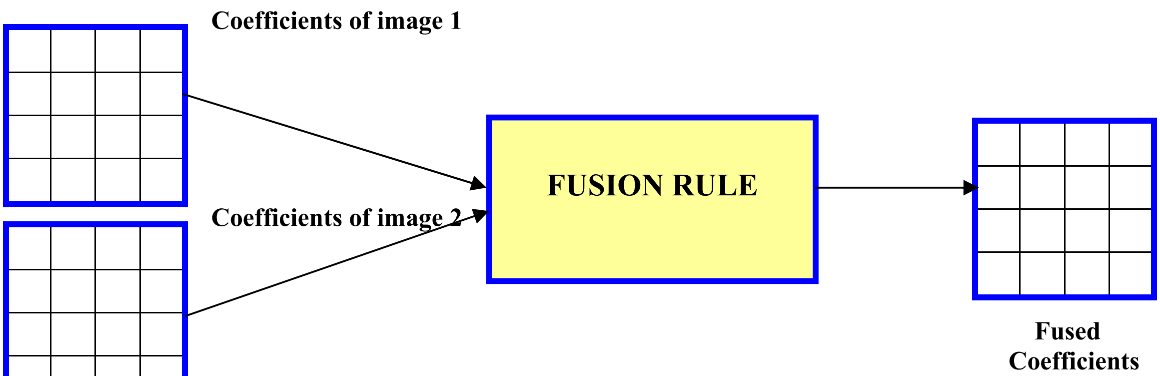
$$M' \times N' = \left(\mathrm{Floor}\left(\frac{M-1}{2}\right) + W\right) \bullet \left(\mathrm{Floor}\left(\frac{N-1}{2}\right) + W\right)$$

***This fact has to be taken into consideration in the reconstruction procedure, which means to keep the appropriate part of the subimage, in order the size of the reconstructed image to be the same as the original**. In other case reconstruction is not absolutely correct.**

2.1.1.2 Fusion process

Having decomposed the input images the fusion process takes place in order to fuse the coefficients of each image, filter the most “useful” information from them and combine it. Fusion takes place at pixel-level, processing each pair of coefficients in each subband created (HL, LH, HH and the one LL) independently through a well-defined selection criterion, called fusion rule.





Schematic illustration of fusion process

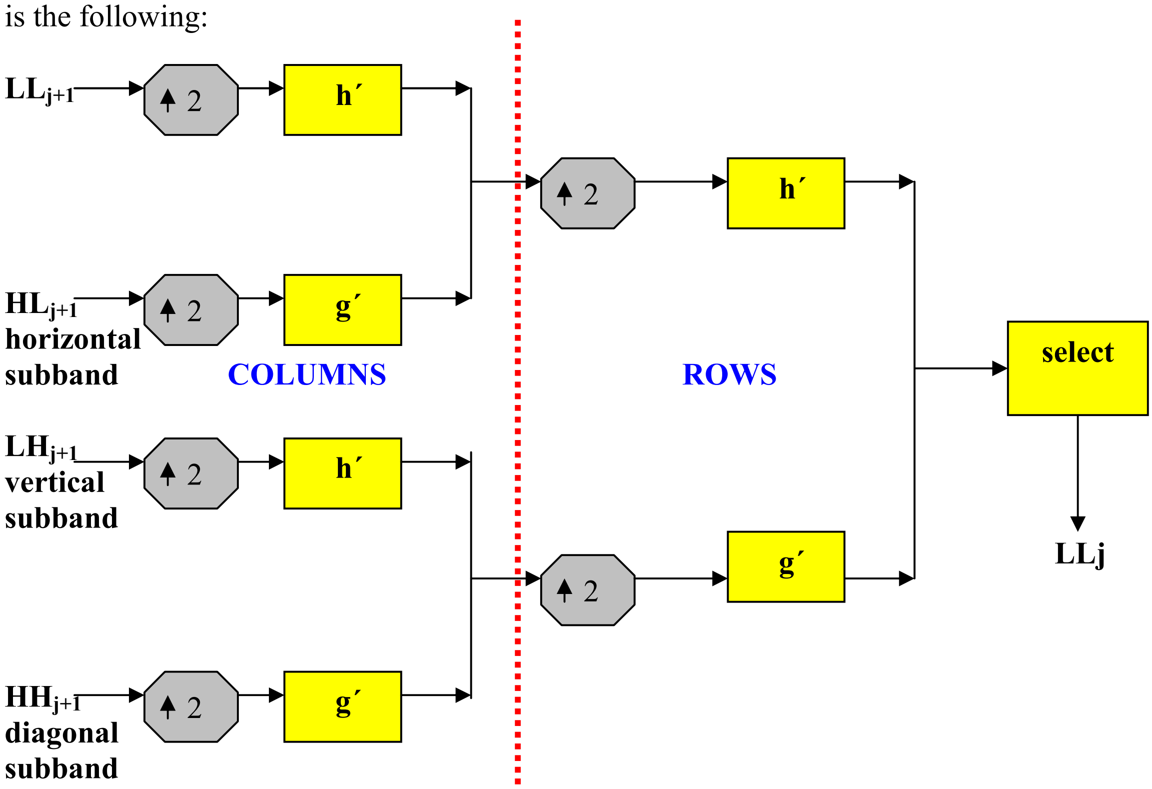
Three of the most common fusion rules are described below [9]:

- Maximum selection (MS) scheme: we simply select the coefficient in each subband with largest value, either taking the magnitude or just the maximum of the two.
- Weighted average (WA) scheme: also called the Burt's fusion rule. This scheme uses a normalized correlation between the two image subbands over a small area (window of neighbor coefficients-pixels). The fused coefficient is calculated from this measure via a weighted average of the two subband coefficients [Burt and Kolcynski, 1993]. The weighted average is defined by a mathematical equation, the parameters of which will determine which of the two images will contribute to the fused image.
- Window based verification (WBV) scheme: also known as Li's method. This scheme creates a binary decision map to choose between each pair of coefficients using a majority filter. In the common case, at each pixel of each image, the coefficient to be fused, a nxn window is centered. The pixels in the neighborhood defined by the window are checked and the one of the largest absolute value is selected. This counts for each of the two images. If the selected pixel belongs to the first image then the corresponding coefficient of this image will contribute to the fused image, otherwise the coefficient of the second image will contribute to the final result. The window moves along each direction until the whole image is "scanned", which means that all the pixels-coefficients are processed.

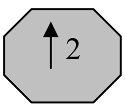
The MS fusion rule is the simplest and the fastest one and gives very good results for the sharp variations of the input images, which are preserved in a better way in the fused image. The other two fusion rules are more sophisticated and computational complex, as they are not applied directly to simple pixels but process the images in "windows"- areas, which can lead to better representation of certain regions in the final image.

2.1.1.3. Reconstruction of images

Having decomposed the source images and fused the coefficients we are able to reconstruct the final, fused image through the reconstruction (or synthesis) procedure, which means to apply the inverse wavelet transform. A pair of synthesis filters (a lowpass filter h´ and a highpass g´ one), capable to allow nearly perfect reconstruction, is recursively applied to the successive level subband images, starting from the last level and ending to the first one. The fused subbands constructed are upsampled in order to make the size of the final image equal to the first one. The general reconstruction scheme is the following:



Reconstruction procedure at level j



Upsampling of the corresponding subband (insert zeros at odd-indexed columns/rows)

h´ Convolution with lowpass filter h’

g´ Convolution with highpass filter g’

select Select the valid part in order to have the appropriate image size (see page 28)

As described above, within the three stages of the image fusion analysis using the DWT, the general procedure has some very basic parameters which need to be properly defined and selected. We make a brief analysis to this basic point:

- **The input images**: source images must, of course, have the same size and be fully registered. This means that the corresponding pixels have to represent the same part of the information to be processed, in order to fuse the appropriate data and not irrelevant ones.
- The number **N** of decomposition/reconstruction levels: this parameter is firstly limited by the size of the original images, it cannot overcome the value $\log_2$(image size). In theory, larger number of levels means detailed decomposition, as even the smallest area/window of the image contributes to the calculation of the information to be fused. On the other hand, this leads to computation complexity and system delay, especially in high-dimensional images.
- **The structure of** analysis (decomposition) and synthesis (reconstruction) **filters**: Filters, used to "simulate" the behavior of wavelet mother functions, have to be properly designed. Their values must be carefully determined in order to allow perfect reconstruction and be associated with the corresponding wavelet family. Wavelets can be realized by iteration of filters with rescaling. The resolution of the image, which is a measure of the amount of detail information in it, is determined by the filtering operations, and the scale is determined by upsampling and downsampling (subsampling) operations.

  There are two categories of filters:

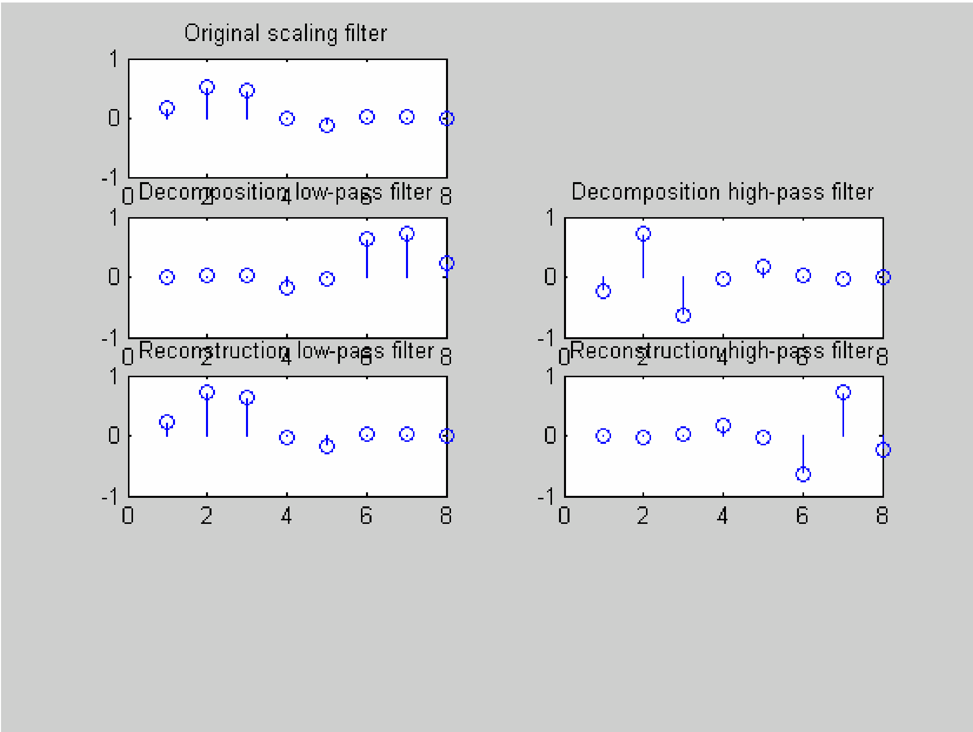
  a) **orthogonal**, whose coefficients are real numbers, filters are of the same size and are not symmetric. The lowpass and highpass filter are alternated flip of each other. They have regular structure which leads to easy implementation and scalable architecture. The analysis and synthesis filters are inverse. An example of orthogonal filters is illustrated below:

  b) **biorthogonal**. In the case of the biorthogonal wavelet filters, the low pass and the high pass filters do not have the same length. The low pass filter is always symmetric, while the high pass filter could be either symmetric or antisymmetric. The coefficients of the filters are either real numbers or integers. For perfect reconstruction, biorthogonal filter bank has all odd length or all even length filters. The two analysis filters can be symmetric with odd length or one symmetric and the other antisymmetric with even length. Also, the two sets of analysis and synthesis filters must be dual. The linear phase biorthogonal filters are the most popular filters for data compression applications.

  An example of orthogonal and biorthogonal filters is illustrated in the next page.
- **The fusion rule**: depending on the application and the structure of the source images, the suitable fusion rule can be followed for best result.

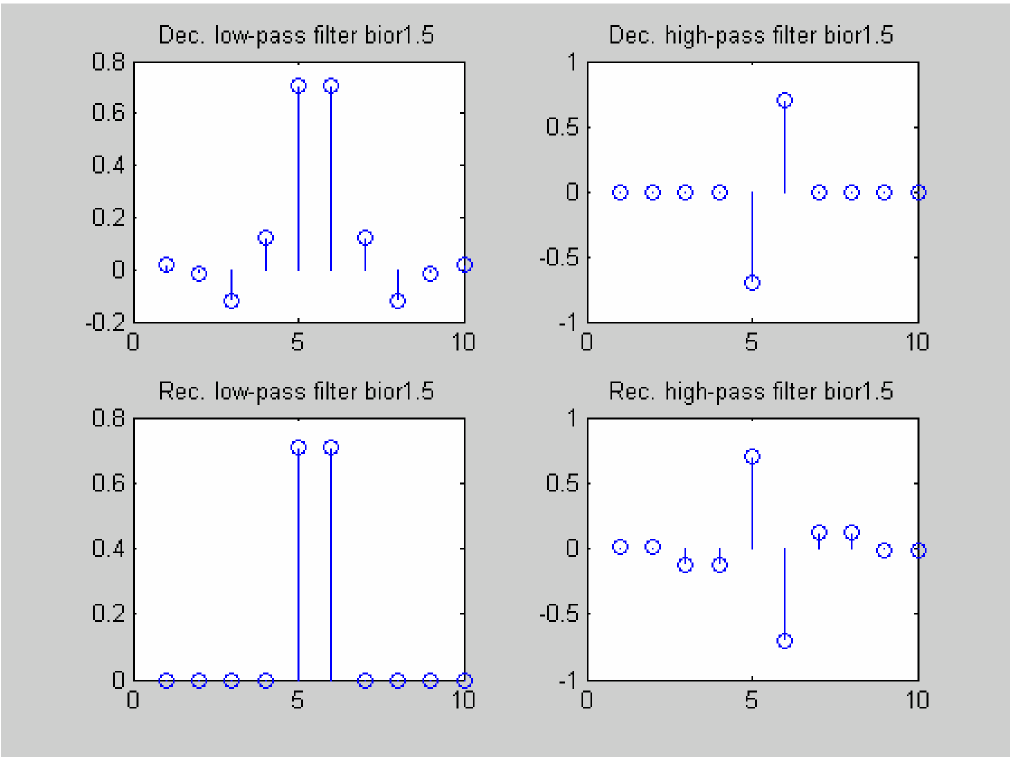
Orthogonal analysis/synthesis filters

Lowpass decomposition filter							
-0.0106	0.0329	0.0308	-0.1870	-0.0280	0.6309	0.7148	0.2304
Highpass decomposition filter							
-0.2304	0.7148	-0.6309	-0.0280	0.1870	0.0308	-0.0329	-0.0106
Lowpass reconstruction filter							
0.2304	0.7148	0.6309	-0.0280	-0.1870	0.0308	0.0329	-0.0106
Highpass reconstruction filter							
-0.0106	-0.0329	0.0308	0.1870	-0.0280	-0.6309	0.7148	-0.2304



Biorthogonal analysis/synthesis filters

Lowpass decomposition filter									
0.0166	-0.0166	-0.1215	0.1215	0.7071	0.7071	0.1215	-0.1215	-0.0166	0.0166
Highpass decomposition filter									
0	0	0	0	-0.7071	0.7071	0	0	0	0
Lowpass reconstruction filter									
0	0	0	0	0.7071	0.7071	0	0	0	0
Highpass reconstruction filter									
0.0166	0.0166	-0.1215	-0.1215	0.7071	-0.7071	0.1215	0.1215	-0.0166	-0.0166



2.1.1.4 Implementation of Regular Discrete Wavelet

2.1.1.4.1. Analysis/synthesis process

In this section the results of the DWT application to medical images and its use to image fusion is presented. The implementation of the above was made using the Matlab Toolbox. It is attempted, through a great number of tests using different parameter values, to export, as much as possible, conclusions about the algorithm and determine the conditions under which this method turns out to be optimum.

At first, we attempt to test and evaluate the analysis/synthesis procedure. To do so we make use of two quality metrics: a) the Root Mean Square Error and b) the quality index Q valued in the range [-1 1] (see Chapter 6 for more details) as proposed by Wang and Bovik [10]. The number N of decomposition levels and the structure of filters (size and coefficients) are the parameters to be determined. The values of the applied filters are given by the function wfilters() of Matlab which associates different wavelet families with filters.

Continuing, we attempt to test and evaluate the whole fusion process. To do so we make use of the quality metrics Qf (see Chapter 6 for more details) as proposed by Gemma Piella and Henk Heijmans [11]. The number N of decomposition levels, the fusion rule and the structure of filters (size and coefficients) are the parameters to be determined.

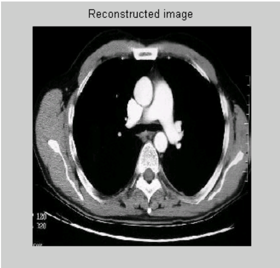
To evaluate the contribution of each parameter to the final result, at every test step we change the value of one, keeping the others fixed.

Another point to focus on is the value range of the coefficients calculated by the application of the discrete wavelet transform. As known, for an 8-bit image, the value range of its pixels is the interval [0 255]. **We noticed that the wavelet coefficients, depending on the algorithm, tend to have a much greater range, sometimes taking negative values**. In order to compare the input and output images, it is wise to set their pixels values in the same domain. So, histogram stretching takes place in order to cut off the pixel values over 255 and under 0. That's why in our evaluation tests two sets of results are presented, one using the coefficients as they are obtained by the application of the transform and one after the normalization has taken place.

A sequence of results is presented in the following pages. The first four testing sets check the efficiency of the DWT and the inverse DWT applied to a single image. The remaining sets evaluate the whole fusion process of two input medical images.

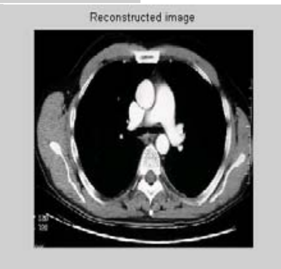
TEST 1 : Defining the number of levels

Original image

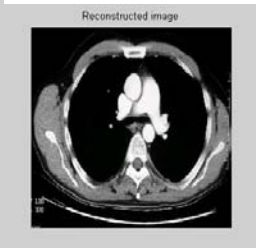


Level 1

Level 2

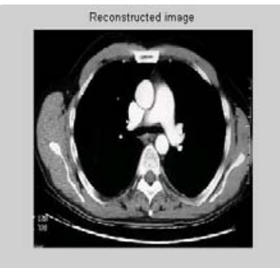


Level 3



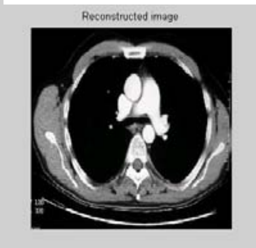
Level 4

Level 5



Level 6

Level 7



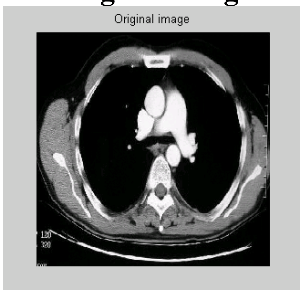
Level 8

Number Of Levels	RMSE Non normalized	RMSE normalized values	Q Non normalized	Q normalized values
N=1	6.5545e-013	0.0060	0.9852	0.8043
N=2	1.3756e-012	0.0059	0.9853	0.8233
N=3	2.1524e-012	0.0058	0.9852	0.8333
N=4	2.9583e-012	0.0055	0.9852	0.8851
N=5	3.8398e-012	**0.0054**	0.9853	0.9148
N=6	4.6868e-012	0.0055	0.9852	**0.9186**
N=7	5.0051e-012	0.0056	0.9852	0.9185
N=8	5.1987e-012	0.0057	0.9852	0.9163

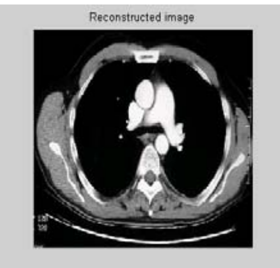
System parameters: **size of filters = 20**
wavelet family = Daubechies

TEST 2 : Defining the size of the filter

Original image

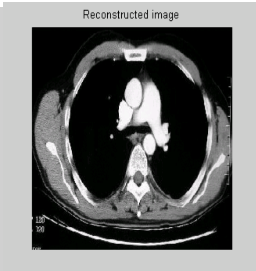


size=2

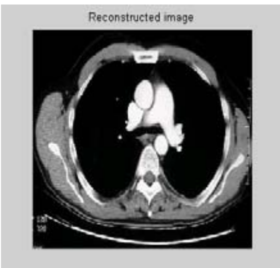


size=4

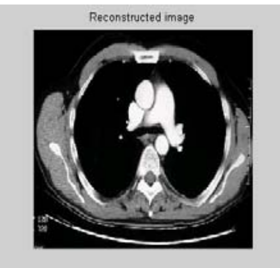
size=6



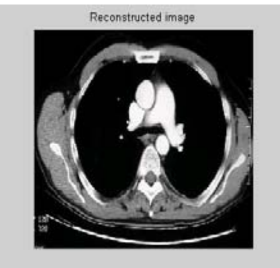
size=8



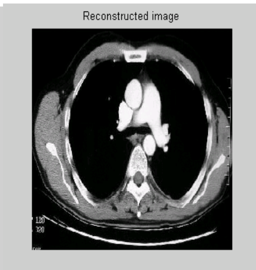
size=10



size=20



size=40



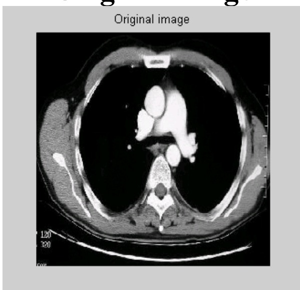
size=80

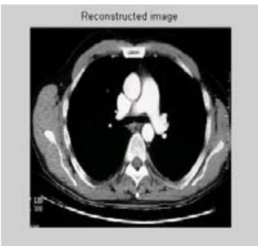
Size Of Filters	RMSE		Q	
	Non normalized	normalized values	Non normalized	normalized values
S=2	5.7346e-016	**0.0039**	0.9853	0.9087
S=4	7.7022e-013	0.0055	0.9852	0.9175
S=6	9.7243e-012	0.0060	0.9852	0.8156
S=8	1.8239e-012	0.0056	0.9852	0.9184
S=10	2.8431e-012	0.0060	0.9853	0.8156
S=20	4.6868e-012	0.0055	0.9852	**0.9186**
S=40	4.6864e-012	0.0058	0.9852	0.9031
S=80	3.4361e-005	0.0059	0.9852	0.8156

System parameters: number of levels = 6
wavelet family = Daubechies

TEST 3 : Defining the wavelet family

Original image

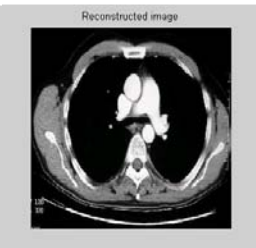




Daubechies

Coiflets

Symlets



Biothogonal

Reverse biorthogonal

other

Wavelet Family	RMSE Non normalized	RMSE normalized values	Q Non normalized	Q normalized values
Daubechies	2.8431e-012	0.0060	0.9852	0.8156
Coiflets	3.2403e-009	0.0058	0.9852	0.8498
Symlets	1.9853e-013	0.0055	0.9852	**0.9146**
Biorthogonal	6.2543e-016	0.0045	0.9853	0.8649
Reverse Biorthogonal	6.0376e-016	**0.0044**	0.9852	0.8779
Other (defined by Oliver Rockinger)	7.9238e-016	0.0071	0.9853	0.8236

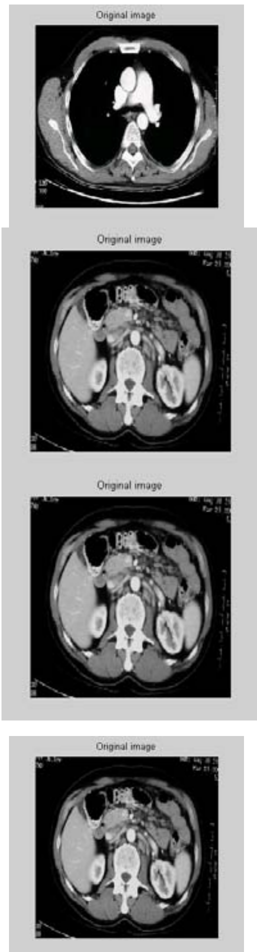
System parameters: number of levels = 6
size of filter = 10

TEST 4 : Testing the input

Input Image	RMSE Non normalized	RMSE normalized values	Q Non normalized	Q normalized values
Image1	6.0376e-016	0.0044	0.9852	0.8779
Image2	6.1832e-016	0.0035	0.8941	0.9398
Image 3	7.1044e-016	0.0041	0.9998	**0.9366**
Image 4	6.8036e-016	0.0040	0.9988	0.9592
Image 5	6.0376e-016	**0.0044**	0.9852	0.8779
Image 6	7.9238e-016	0.0071	0.9853	0.8236

System parameters: number of levels = 6
size of filter = 10
wavelet family = reverse biorthogonal

Original image



Reconstructed image

From the analysis above we can have a first impression of the problem. At first sight no very obvious conclusion can be made. Performance is a combination of many parameters and its evaluation is not an easy part. The only obvious is that the coefficients have to be normalized in order to have a clear comparison. Taking a closer look at the results summarized in the tables, and focusing on the peak values of the quality factor Q we can assume that the following set of parameters seems to produce a result very close to the optimum possible

Parameter set: **size of filter =20** (efficient enough, not so much calculation complexity)
wavelet family = Symlet
number of levels=6-7

Let's attempt to confirm this hypothesis:

Number Of Levels	RMSE Non normalized	RMSE normalized values	Q Non normalized	Q normalized values
N=1	3.2662e-014	0.0059	0.9860	0.8038
N=2	6.0180e-014	0.0060	0.9853	0.8013
N=3	9.0157e-014	0.0059	0.9852	0.8171
N=4	1.1768e-013	0.0057	0.9851	0.8434
N=5	1.5491e-013	0.0056	0.9853	0.8874
N=6	1.9853e-013	0.0055	0.9852	0.9146
N=7	2.0694e-013	0.0055	0.9852	0.9170
N=8	2.1902e-013	0.0058	0.9852	0.9091
System parameters:	**size of filters**	=	**10**	
	wavelet family	=	**Symlet**	

It is getting clear, comparing to the tables above, that increasing the number of levels, the performance of the algorithm is improved (notice that we have not chosen the "optimum" filter size), which is confirmed by both evaluation measures. The RMSE is eliminated, which means that the reconstructed image gets closer to the original one and the quality factor Q is increased, which means that the output image is more similar to the input one.

Number Of Levels	RMSE Non normalized	RMSE normalized values	Q Non normalized	Q normalized values
N=1	1.2913e-016	0.0040	0.9941	0.9156
N=2	2.3322e-016	0.0040	0.9855	0.9150
N=3	3.3150e-016	0.0041	0.9853	0.9091
N=4	4.2521e-016	0.0043	0.9852	0.8858
N=5	5.1218e-016	0.0044	0.9853	0.8817
N=6	6.0376e-016	0.0044	0.9852	0.8779
N=7	6.7199e-016	0.0042	0.9852	0.8797
N=8	7.6204e-016	0.0039	0.9853	0.9116
System parameters:	size of filters = 10		wavelet family = reverse biorthogonal	

Using the same filter size but different filter coefficients we notice that increasing the number of levels the performance is decreased!

Number Of Levels	RMSE Non normalized	RMSE normalized values	Q Non normalized	Q normalized values
N=1	4.0574e-015	0.0058	0.9852	0.8676
N=2	8.6017e-015	0.0060	0.9853	0.8570
N=3	1.3681e-014	0.0061	0.9852	0.8376
N=4	1.9018e-014	0.0062	0.9852	0.8240
N=5	2.5523e-014	0.0061	0.9852	0.8174
N=6	3.1646e-014	0.0060	0.9853	0.8155
N=7	3.4270e-014	0.0060	0.9852	0.8156
N=8	3.5757e-014	0.0059	0.9853	0.8169
System parameters:	size of filters = 20		wavelet family = Symlet	

If we try to combine the values that seem to give the best performance we end to the conclusion that increasing the number of levels, the performance is decreased! Notice that the filter coefficients are the same as in the table of TEST 3 but the filter size is doubled. The results are totally different!!

Conclusion: Taking into consideration all these we notice that attempting to uniquely determine what is "optimum" can be misleading, it is application dependent. That's why in the whole bibliography researchers provide their own suggestion about the parameters of the wavelet analysis/synthesis procedure.

2.1.1.4.2. Fusion process

In the previous section the Discrete Wavelet Transform was introduced. In this section its application in image fusion is presented. To evaluate the performance of the algorithm, a quality metric Q_f (see chapter 6) for image fusion is used. The results are extracted through a similar procedure as presented before: the decomposition/reconstruction parameters- number of levels, filter size, wavelet family- are gradually changed in order to attempt to approximate the "optimum" system structure.

The three fusion rules mentioned (maximum selection, weighted average using variance and window based verification in a 3x3 window) are implemented. The fusion of the LL subband coefficients, also called the "base" or the "lowpass" image, is a different topic. On the grounds that the lowpass image contains much of the information the original images carry out, as it is simply a downsampled, blurred copy of them, the fusion rule can take the following simple forms: a) select the base subband of the one image, b) select the base subband of the other image, or c) take their average. The third choice is the optimum and the usual, as in most cases the structure of the input images is not predefined so as to determine which image can have better contribution to the final information. As for the implementation of the fusion process in the highpass subbands the following can be mentioned:

a) maximum selection: simply, the coefficient with the largest magnitude is selected

b) weighted average (Burt's method): firstly, a small local area (window) size is selected. Moving this window pixel-to-pixel within each image a salience factor is computed within each area according to the relation

$$\mathbf{factor(i,j)} = \frac{\Sigma \mathbf{image1(i,j)*image2(i,j)}}{\Sigma [\mathbf{image1(i,j)}^2 + \mathbf{image2(i,j)}^2]}$$

where (i,j) is the position where the window is applied.

A threshold **T is** then defined and a selection factor is calculated

$$\mathbf{weight} = 0.5 - \frac{\mathbf{0.5*(1\text{-}factor)}}{\mathbf{1\text{-}T}}$$

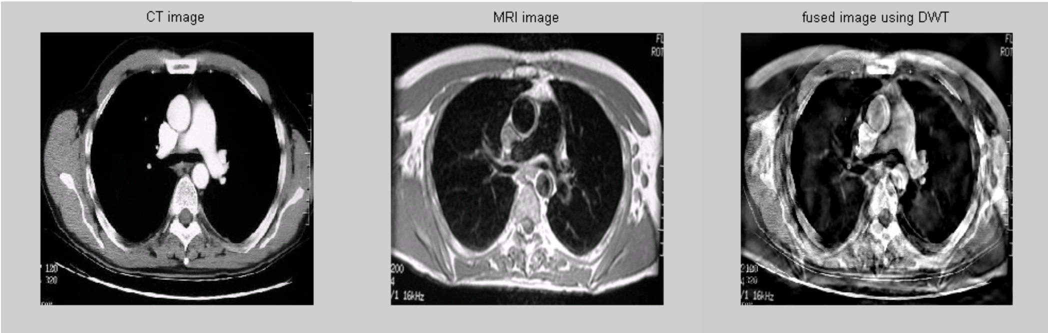
If the value of the factor at position (i,j) is over the threshold T and $\Sigma \mathbf{image1(i,j)}^2 > \Sigma [\mathbf{image2(i,j)}^2$ then the coefficients of image1 are selected multiplied by the factor **(1- weight),** otherwise the coefficients of image2 weighted by the factor **weight** are selected as the coefficients of the fused image at position (i,j).

c) window based verification (Li's method): a window is defined again and centered at each position (i,j). The coefficient of the image that has the largest magnitude within the window, will define which image will contribute to the fused result. Then, the coefficient at position (i,j) of the chosen image (simply its value and NOT its magnitude) is selected for the fused image

A sequence of results follows:

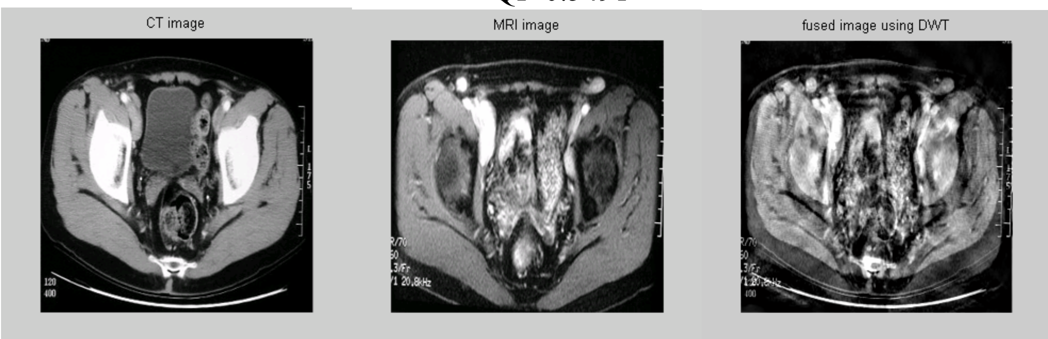
Test1:

wavelet family=daubechies, filter size=20



Qf=0.5230

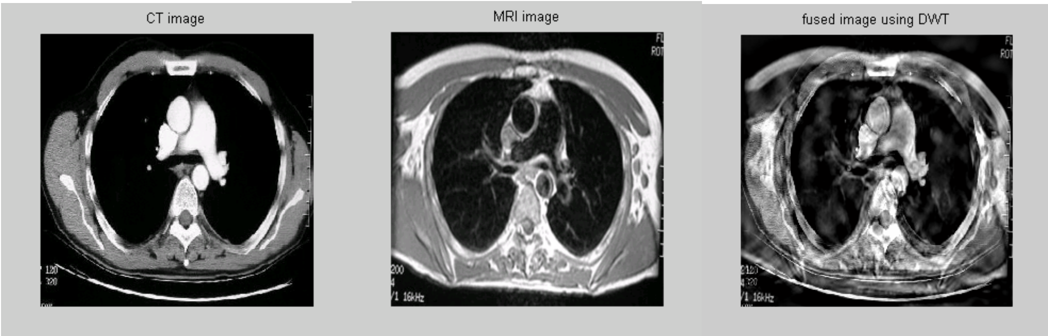
Qf=0.5491



Qf=0.5821

Qf=0.5671

Fusion rule=maximum selection



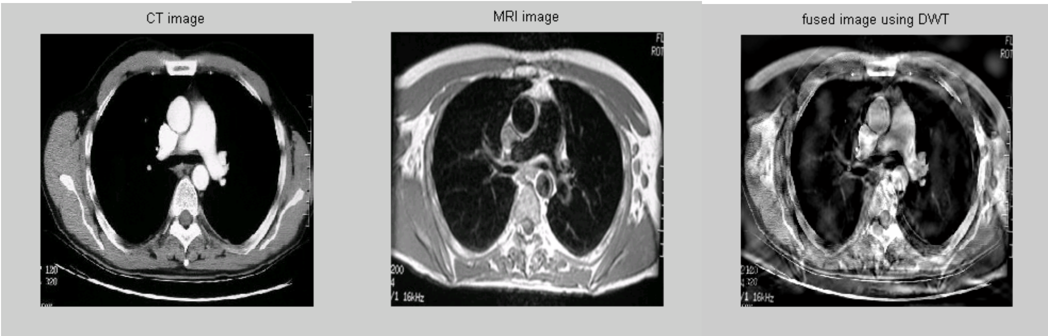
Qf=0.4787

Qf=0.5109

Qf=0.5456

Qf=0.5282

Fusion rule= weighted average (using variance) within a 3x3 window



Qf=0.4751

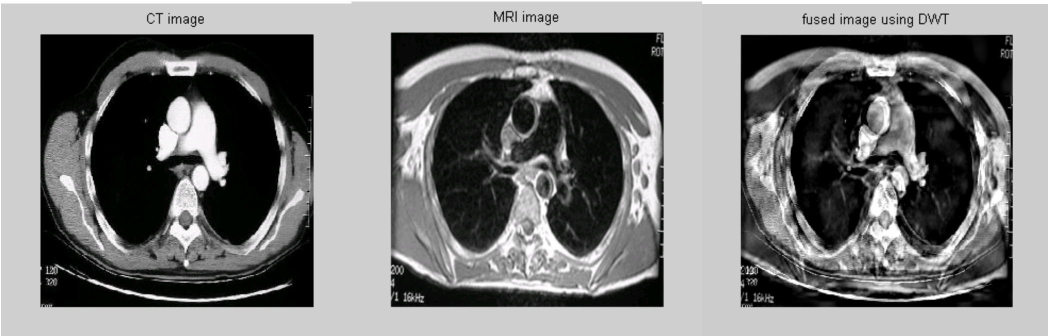
Qf=0.5063

Qf=0.5477

Qf=0.5290

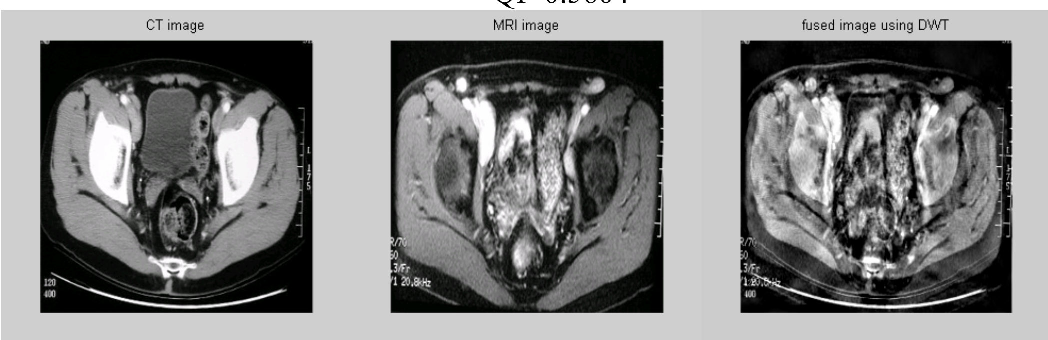
<u>Fusion rule= window based verification within a 3x3 window</u>

Test2:
wavelet family=daubechies, filter size=10



Qf=0.5418

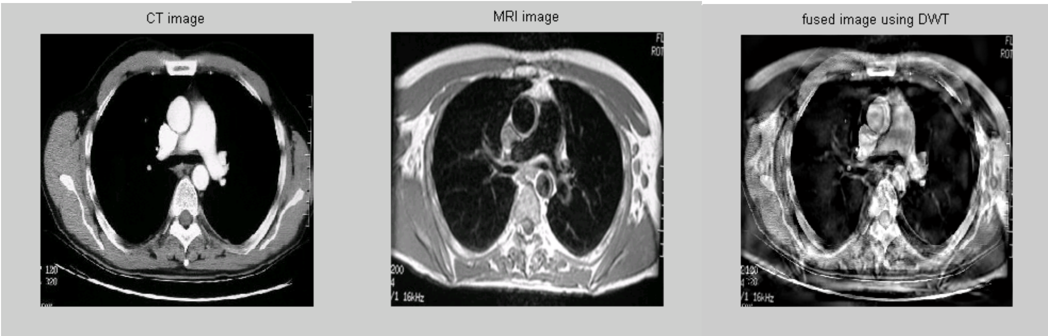
Qf=0.5604



Qf=0.6008

Qf=0.5881

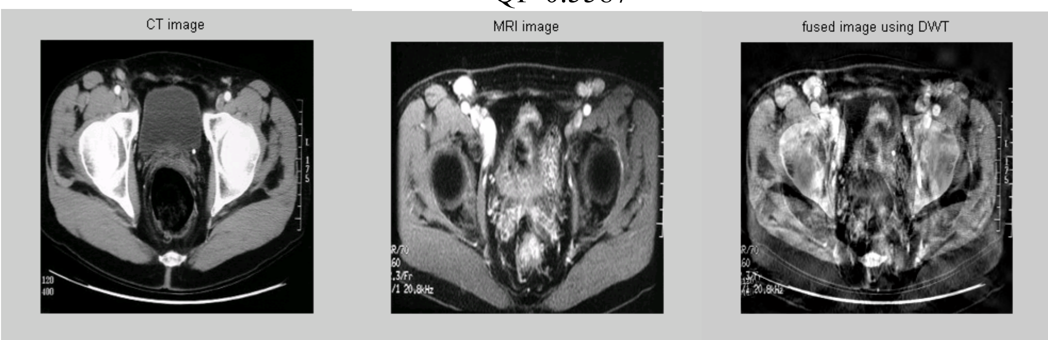
Fusion rule =maximum selection



Qf=0.5009

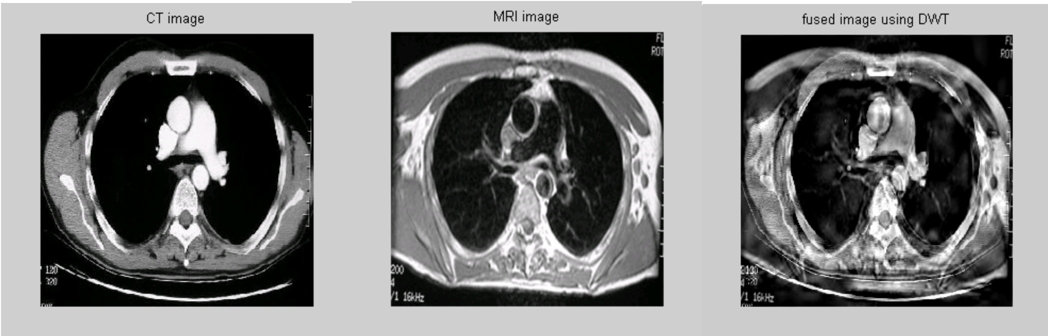
Qf=0.5249

Qf=0.5587



Qf=0.5564

<u>Fusion rule= weighted average (using variance) within a 3x3 window</u>



Qf=0.5014

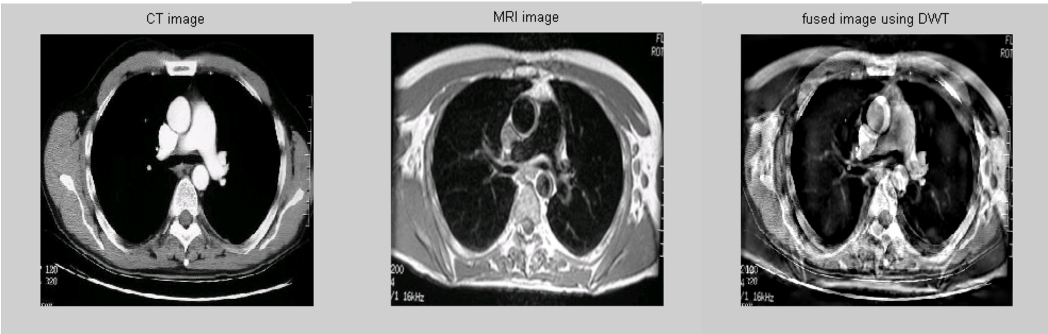
Qf=0.5185

Qf=0.5614

Qf=0.5506

<u>Fusion rule= window based verification within a 3x3 window</u>

Test3:
wavelet family=Symlet, filter size=10



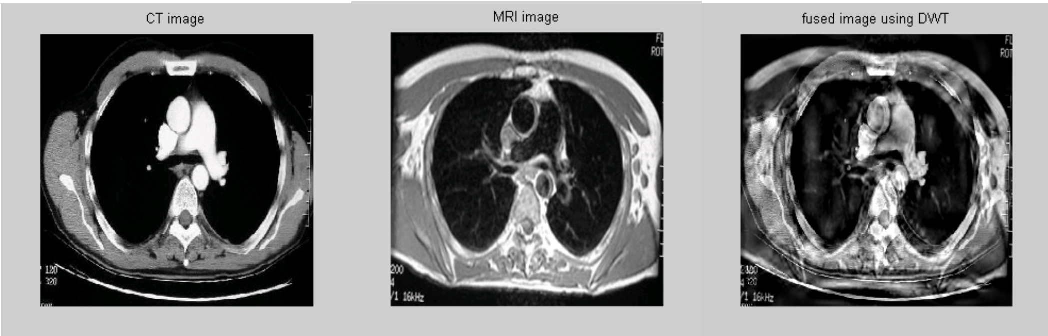
Qf=0.5504

Qf=0.5689

Qf=0.5908

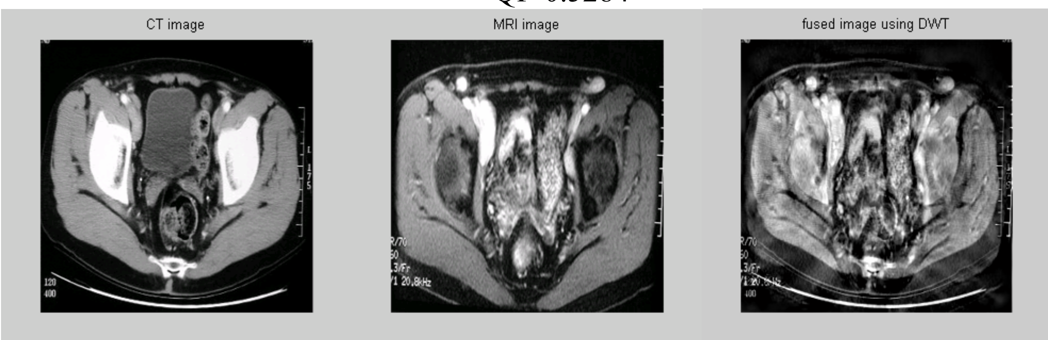
Qf=0.5928

Fusion rule =maximum selection

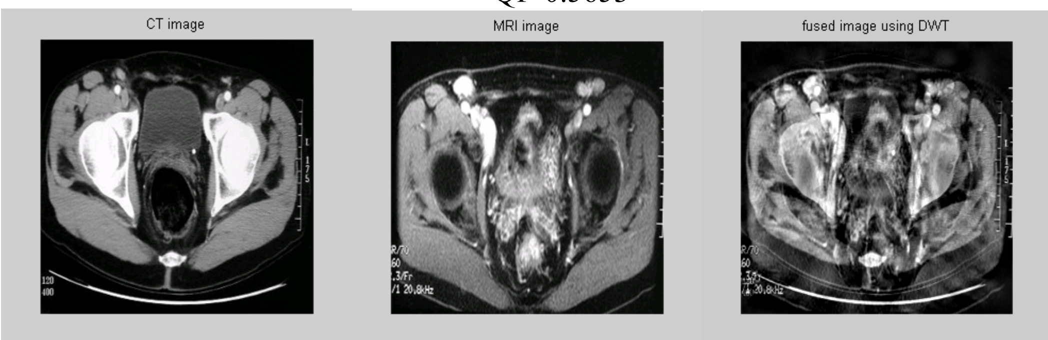


Qf=0.5016

Qf=0.5284

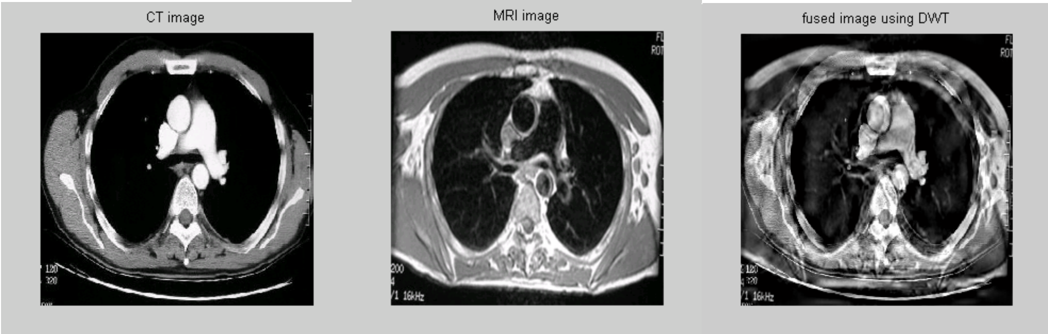


Qf=0.5653



Qf=0.5542

<u>Fusion rule= weighted average (using variance) within a 3x3 window</u>



Qf=0.5118

Qf=0.5226

Qf=0.5629

Qf=0.5517

Fusion rule= window based verification within a 3x3 window

The above are summarized in the following table in order to have a clearer view of the results:

Image pair	Fusion rule 1 Maximum Selection	Fusion rule 2 Weighted Average	Fusion rule 3 WBV
Parameters: Number of levels=5,filter size=20,wavelet family=daubechies			
1	**0.5230**	0.4787	0.4751
2	**0.5491**	0.5109	0.5063
3	**0.5821**	0.5456	0.5477
4	**0.5671**	0.5282	0.5290
Parameters: Number of levels=5,filter size=10,wavelet family=daubechies			
1	**0.5418**	0.5009	0.5014
2	**0.5604**	0.5249	0.5185
3	**0.6008**	0.5587	0.5614
4	**0.5881**	0.5564	0.5506
Parameters: Number of levels=5,filter size=10,wavelet family=Symlets			
1	**0.5504**	0.5016	0.5118
2	**0.5689**	0.5284	0.5226
3	**0.5908**	0.5653	0.5629
4	**0.5928**	0.5542	0.5517

2.1.1.5 Conclusions

As mentioned in the sections above evaluation is not an easy issue and sometimes the results extracted cannot be objective, unfortunately they can be misleading. In addition, it is much more difficult to determine the parameters (number of levels, structure of filters) of the decomposition-fusion-reconstruction scheme, as performance is application and input dependent. Besides, one parameter value can give unsatisfactory results but in combination with another suitable parameter value it is likely to have a much duckier effect.

The last table gave us a very clear conclusion: **the maximum selection rule outperforms the other two implemented fusion rules, no matter what the "system parameters" are.**

As regards the number of levels, a value in the range [4-6] seems to be optional (the images used were of size 256x256, so the maximum number of levels is 8). In theory level number one should produce the perfect reconstruction results.

The structure of filters is also a key point. A key point is the size of the filters to be used. A filter of large size could produce redundant information and a filter of small size could not represent the sufficient information, data loss would take place. A size of 10 seems to be a satisfactory selection.

Finally, the wavelet family to determine the value of the filter is also a topic to mention. Daubechies wavelets are the most commonly used, the Symlets wavelet family had the best performance in our measures, biorthogonal filters were suggested in bibliography for perfect reconstruction. It depends on the researcher to decide the optimal solution.

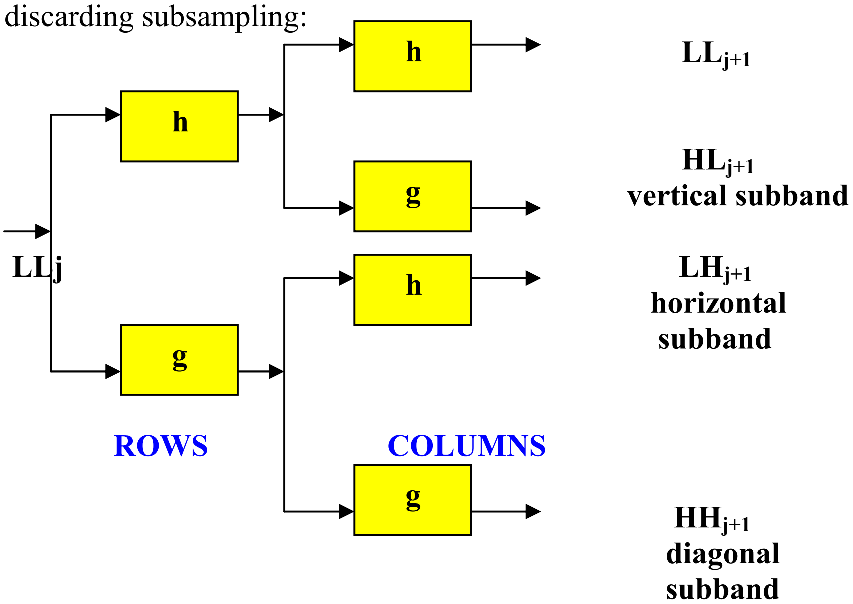
In our case the triple 5(levels)-10(size)-Symlets (wavelet family) was proved sufficient enough.

2.1.2. The Shift Invariant Discrete Wavelet Transform

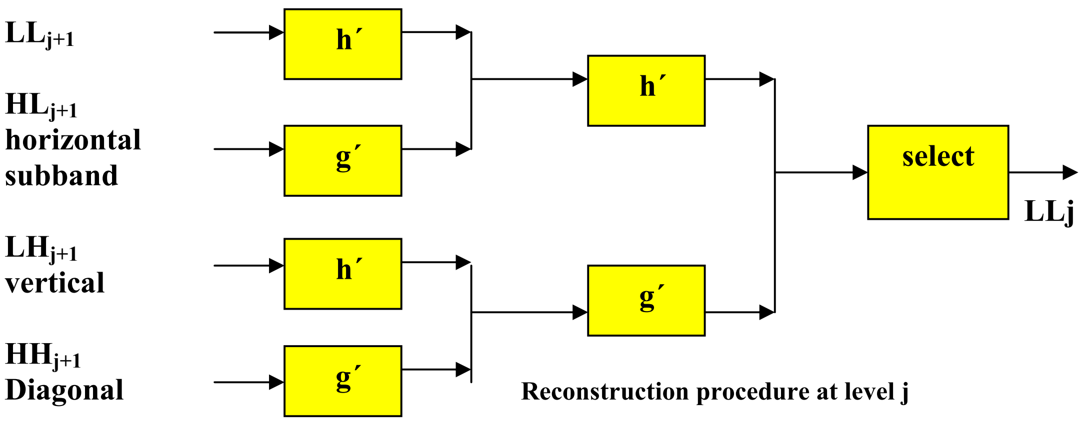
2.1.2.1. Introduction

The Shift Invariant Discrete Wavelet Transform (SIDWT) evolved as an extension of the DWT, an improved version of it in order to overcome its limitations. In a normal wavelet decomposition small shifts of the input image are able to move energy between subbands, which is a result of the subsampling necessary for critical decimation. Sampling takes place in order to reduce the amount of data that has to be analyzed and to enforce the implicit time-frequency uncertainty of the analysis. However, this results in coefficients that are highly dependent on their location, which can lead to small shifts in the input causing large changes in the wavelet coefficients, large variations in the distribution of energy at different scales and possibly large changes in the reconstructed images [12]. The SIDWT overcomes shift invariance by discarding all subsampling.

The decomposition/reconstruction scheme of the SIDWT is the same as the DWT discarding subsampling:



Decomposition procedure at level j

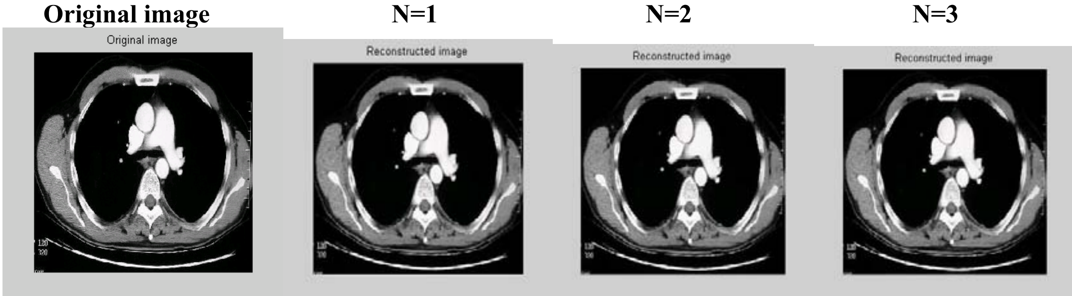


Reconstruction procedure at level j

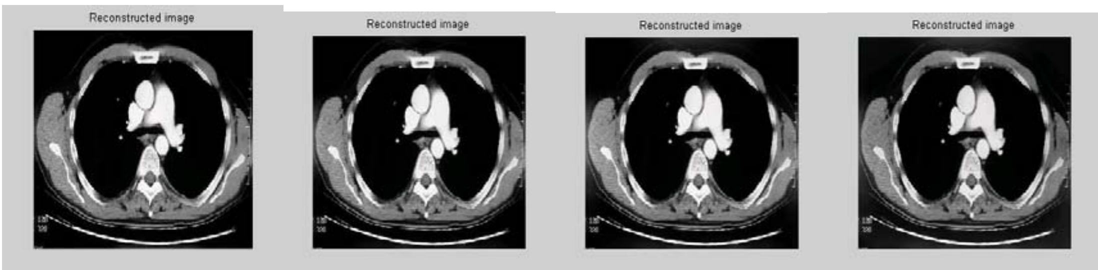
2.1.2.2. Implementation of the Shift Invariant Wavelet Transform

Having in mind the above and following the same procedure as in the previous section, we implement the fusion using the SIDWT. The three, already known and analyzed, fusion rules are adopted, a number of 4 levels is chosen, the "Daubechies" wavelet and a filter size of 4 are selected. At each level the filter is zero-padded, a number of zeros, depending on the level the decomposition/reconstruction takes place, are inserted between the coefficients. The same quality indexes, Q and Qf, are adopted (see Chapter 3 for details).

Original image **N=1** **N=2** **N=3**



N=4 **N=5** **N=6** **N=7**



Number Of Levels	RMSE normalized values	Q normalized values
N=1	**0.0051**	**0.9179**
N=2	0.0058	0.9174
N=3	0.0075	0.9141
N=4	0.0151	0.9089
N=5	0.0240	0.8896
N=6	0.0388	0.8585
N=7	0.0506	0.7796

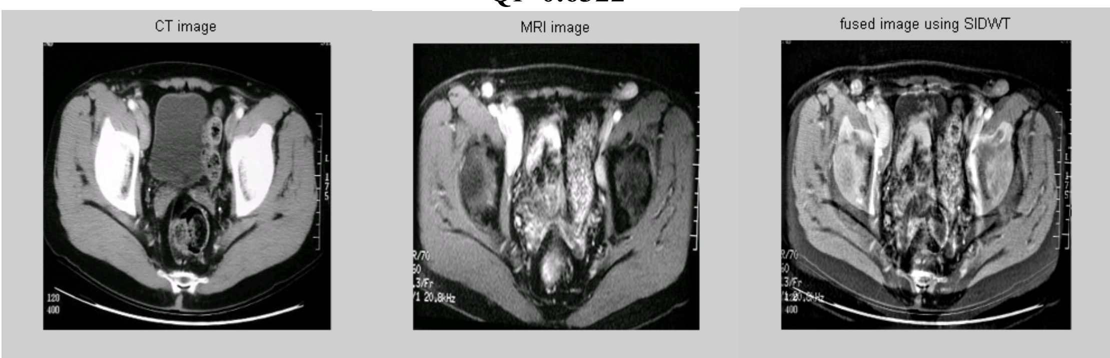
System parameters: **size of filters = 4 (at first)**
wavelet family = Daubechies

T E S T 1 : NUMBER OF LEVELS=4

Fusion rule=maximum selection

Qf=0.6238

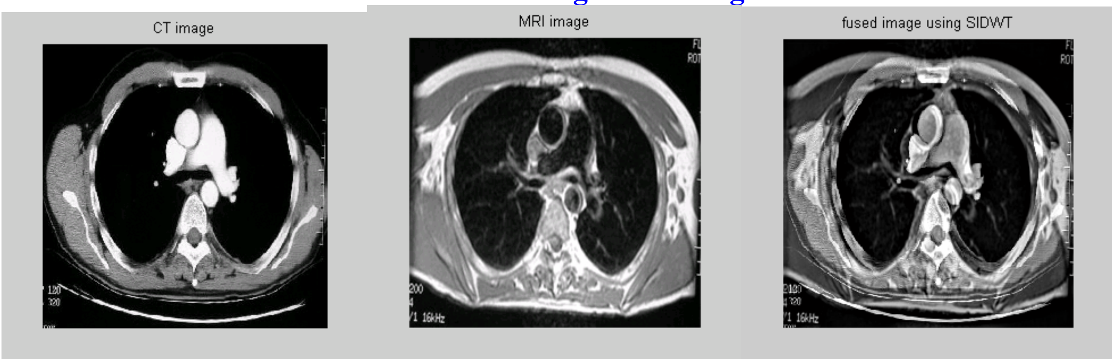
Qf=0.6322



Qf=0.6804

Qf=0.6848

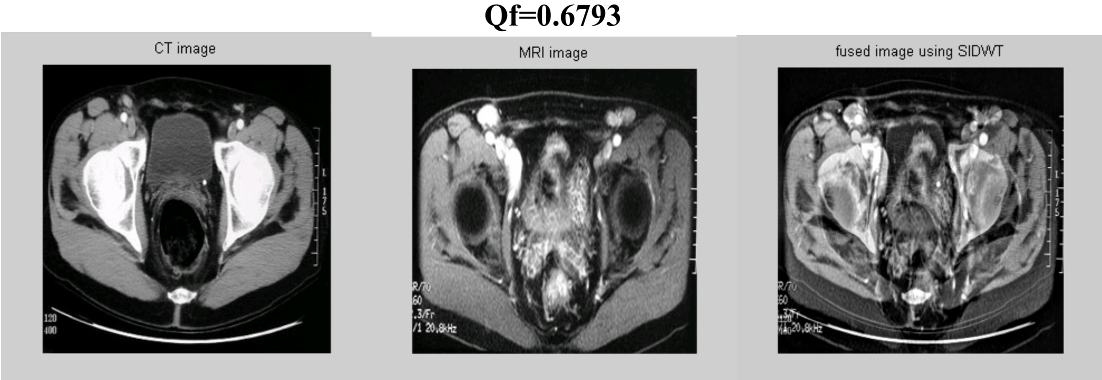
Fusion rule=weighted average



Qf=0.6244

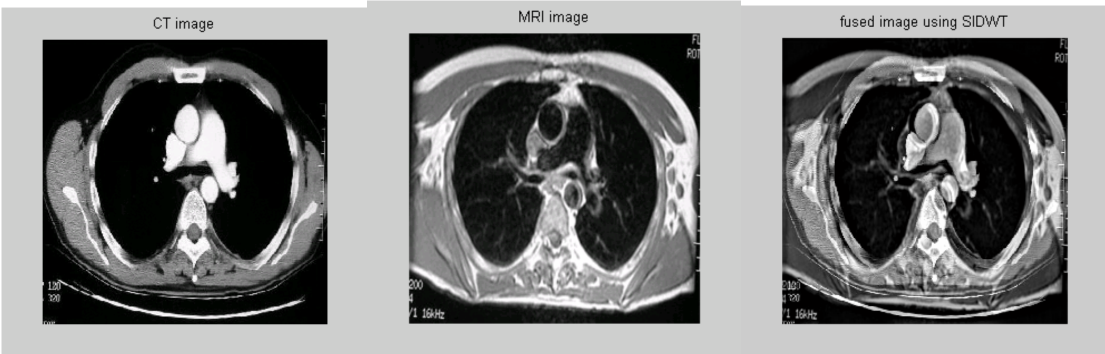
Qf=0.6302

Qf=0.6793



Qf=0.6838

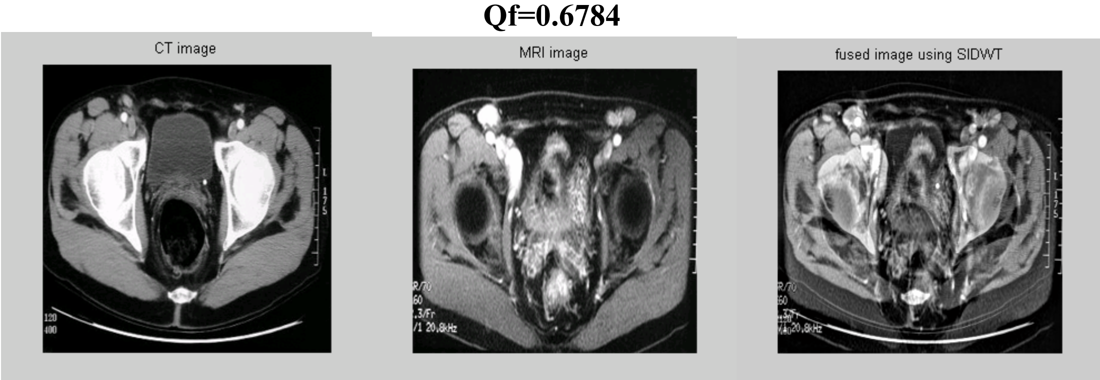
Fusion rule=window based verification



Qf=0.6217

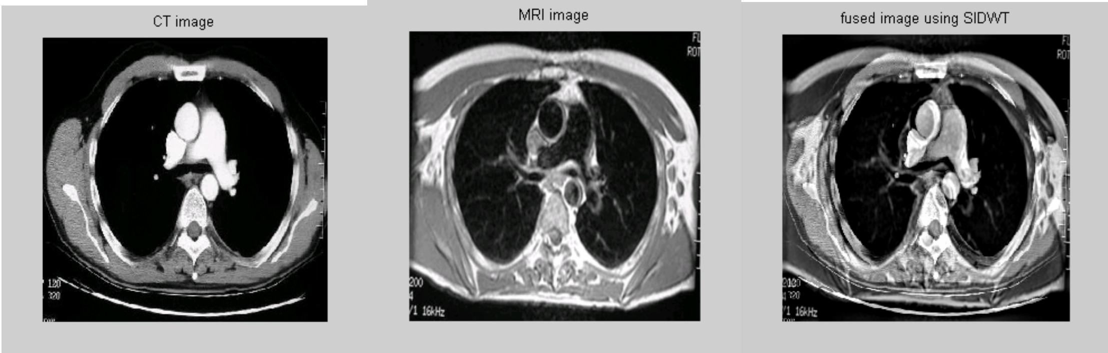
Qf=0.6281

Qf=0.6784



Qf=0.6822

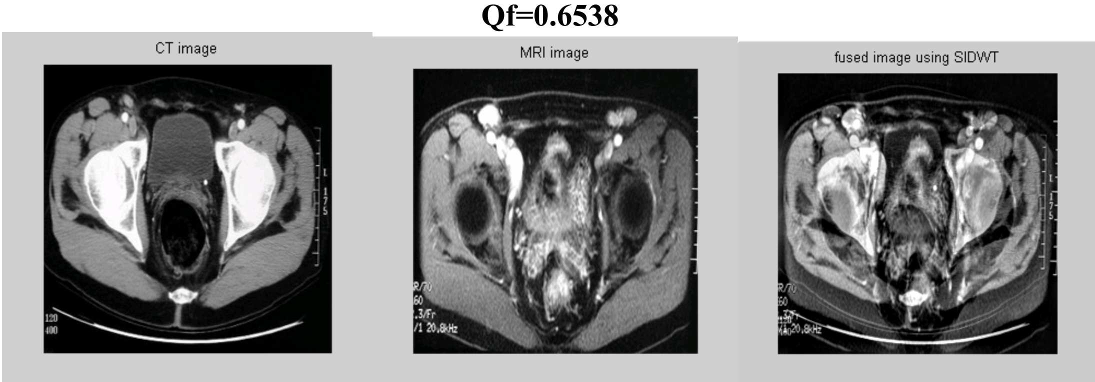
TEST 2 : NUMBER OF LEVELS=5
Fusion rule=maximum selection



Qf=0.5803

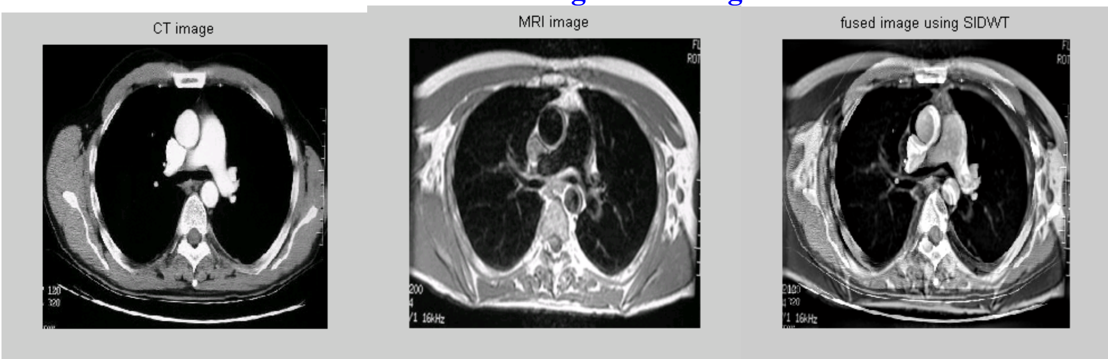
Qf=0.6281

Qf=0.6538



Qf=0.6638

Fusion rule=weighted average



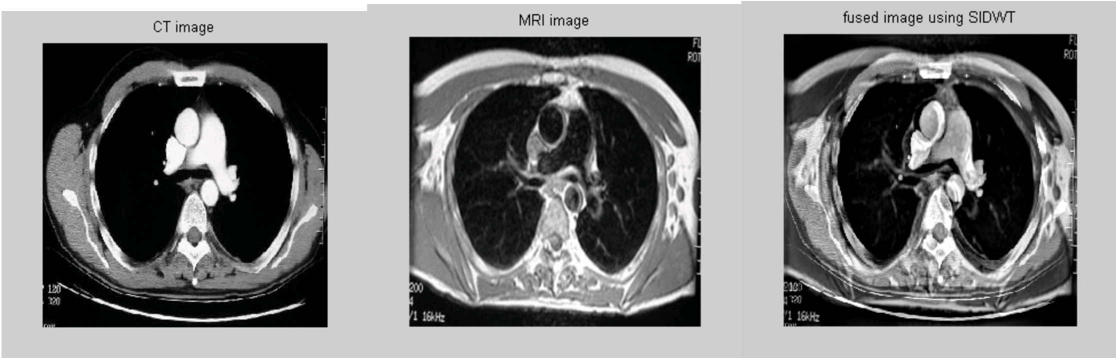
Qf=0.5844

Qf=0.6270

Qf=0.6541

Qf=0.6655

Fusion rule=window based verification



Qf=0.5777

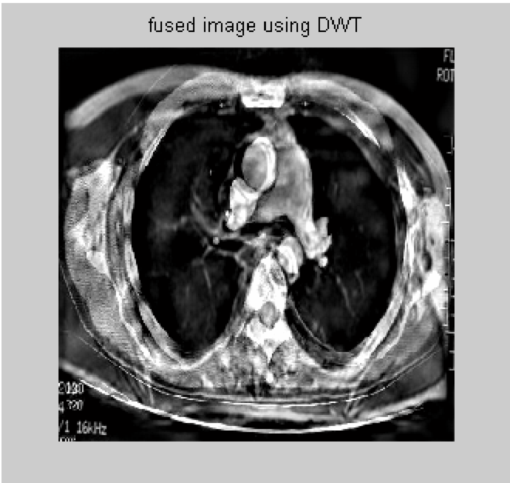
Qf=0.6242

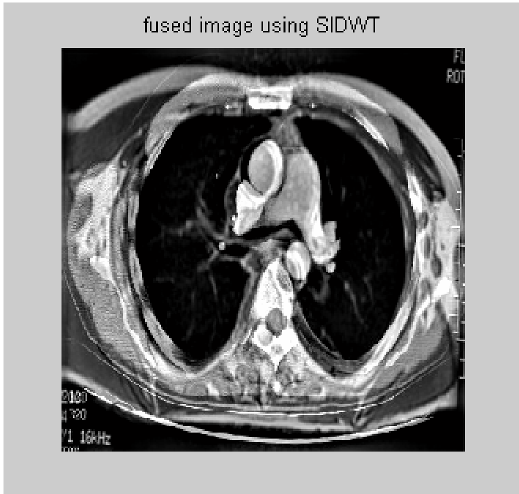
Qf=0.6515

Qf=0.6605

2.1.2.3. Conclusions

The performance evaluation of the Shift Invariant Wavelet Transform just validated the theoretical background: it is an improved version of the Discrete Wavelet Transform. The improvement is obvious at the first sight, and the evaluation through the quality metric Q_f confirms that the SIDWT outperforms the DWT. It is also clear that increasing the number of decomposition levels increases the Root Mean Square Error and decreases the quality factor, which means that the performance of the algorithm tails away. It is also clear that although the Maximum Selection fusion rule provides the highest fusion performance in the DWT, in this case the three fusion rules produce very similar results!!
.

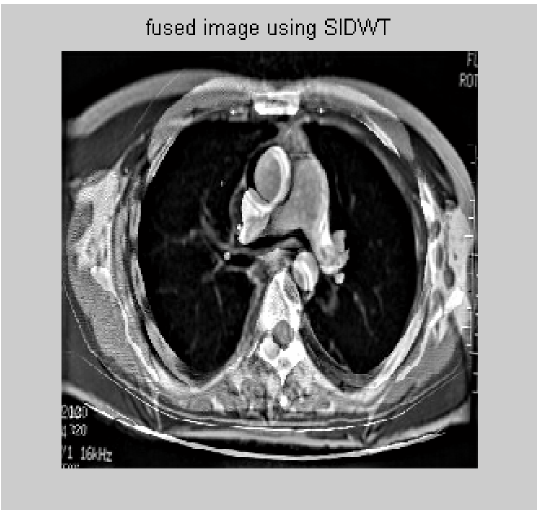


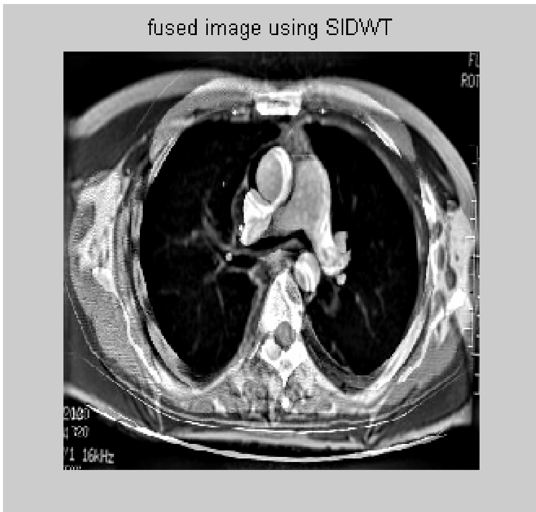


Qf=0.5418 DWT

Qf=0.5803 SIDWT

Wavelet family= Daubechies, Number of levels=5





Qf=0.6238 SIDWT with N=4

Qf=0.5803 SIDWT with N=5

Parameters: filter size at first level=4,wavelet family=daubechies **Number of levels=4**			
Image pair	**Fusion rule 1** **Maximum Selection**	**Fusion rule 2** **Weighted Average**	**Fusion rule 3** **WBV**
1	0.6238	**0.6244**	0.6217
2	**0.6322**	0.6302	0.6281
3	**0.6804**	0.6793	0.6784
4	**0.6848**	0.6838	0.6822
Parameters: filter size at first level=4,wavelet family=daubechies **Number of levels=5**			
1	0.5803 ↓	**0.5844** ↓	0.5777 ↓
2	**0.6281** ↓	0.6270 ↓	0.6242 ↓
3	0.6538 ↓	**0.6541** ↓	0.6515 ↓
4	0.6638 ↓	**0.6655** ↓	0.6605 ↓
SIDWT at different number of decomposition levels and fusion rules			

Of course it is not all perfect! The improvement the algorithm offers comparing to the regular DWT has the cost of excessive redundancy, as the extracted data are not subsampled but preserve the size of the input images. The memory requirements are increased by a factor of 2^j for each input (j defines the level of decomposition) at each stage comparing to the regular DWT and the computational complexity is high in a similar way. This is not evident in powerful computing systems, where computations are performed in milliseconds. However, an input of considerable size, a complicated filter structure, a great number of levels and a medium or slow computing system certainly outlines this drawback.

2.1.3 Dual Tree Complex Wavelet Transform

2.1.3.1. Introduction

Another suggestion to integrate shift invariance into the Discrete Wavelet Transform is the Dual Tree Complex Wavelet Transform [Kingsbury, 1998]. The name of the algorithm implies that complex filters are used in both the decomposition and reconstruction stage. Unfortunately, many experiments have shown that it is very difficult to design an inverse transform, based on complex filters which can guarantee perfect reconstruction of the output signal and good frequency selectivity. Although complex filters can be designed to give perfect reconstruction quite easily at level 1 by applying the constraint that the reconstructed signal must be real, a similar constraint cannot be applied at further levels where inputs and outputs are complex [16]. Hence, a different approach should be followed.

The solution proposed by Kingsbury was the development of the Dual Tree Complex Wavelet Transform by noting that shift-invariance can be achieved with a real Discrete Wavelet Transform. Instead of discarding subsampling as in the SIDWT, the DT-CWT doubles the sampling rate at each level of the tree (the term will be analyzed below) by eliminating the downsampling by 2 after the first level of filtering. This is equivalent to having two parallel fully decimated trees (each performing a Discrete Wavelet Transform) after the first level of decomposition. The filter structure at each level should then satisfy some conditions.

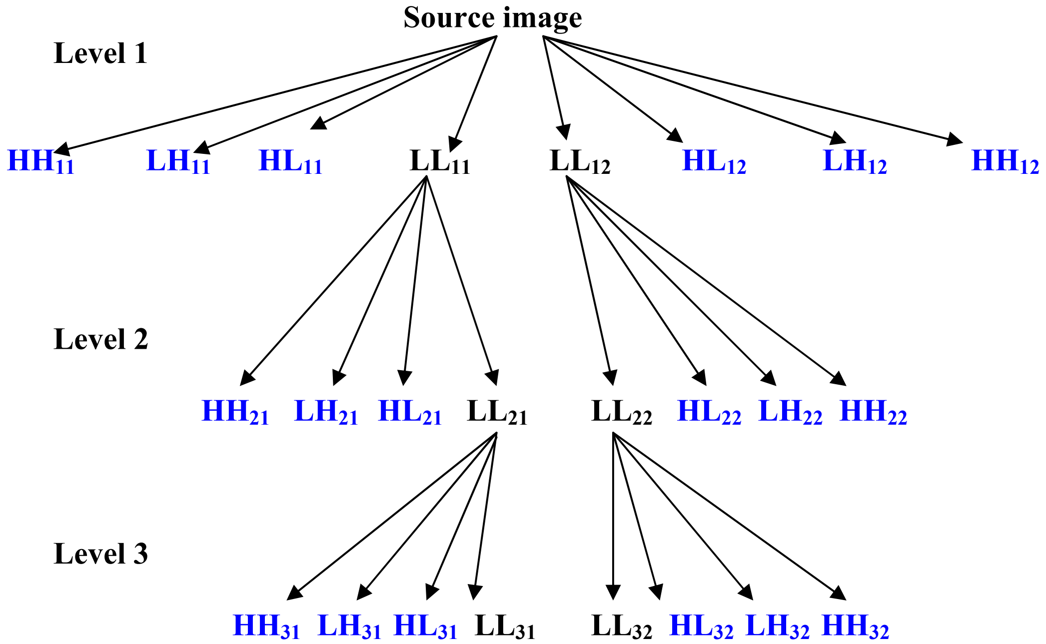
The Dual-Tree Complex Wavelet Transform iteratively applies separable spatial filters to produce frequency subbands as in the Discrete Wavelet Transform [9]. Two fully decimated trees are constructed, the one containing the even and the other containing the odd samples after the first level of filtering. It was found that, to get uniform intervals between samples from the two trees below level 1, the filters in one tree must provide delays that are half a sample different from those in the other tree. This requires odd-length filters in the one tree and even-length filters in the other. In order to achieve greater symmetry between the two trees, the one tree could use odd and the other even filters alternately from level to level. **However, this is optional, not essential and will not be considered in our implementation**.

Thus far, the Dual Tree transform does not seem to be complex. However, we can make it, in a "pseudo-way": we can assume that the outputs from the two trees are interpreted as the real and imaginary parts of complex wavelet coefficients. The filters are selected from a perfect reconstruction biorthogonal set and the impulse responses can be considered as the real and imaginary part of a complex wavelet [Kingsbury, 1998].

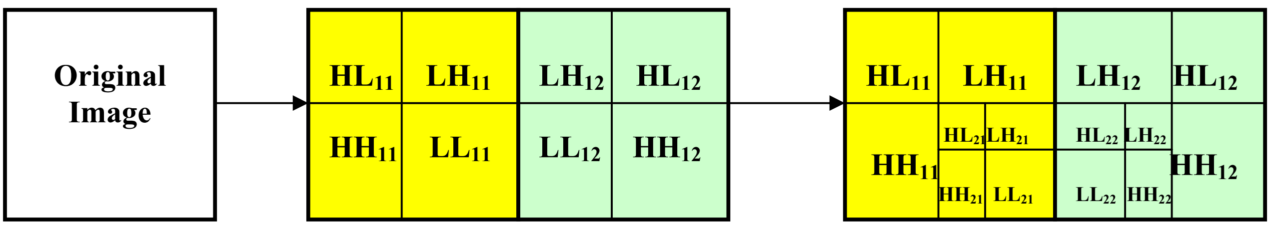
The filters, as in the case of the DWT, are applied to the two dimensions, firstly to the rows and then to the columns of the image data vector. As a result eight subbands are constructed now, four in each tree, three highpass and a lowpass one. This stands for real filters. In the case of complex ("pseudo-complex") wavelet transform the subbands are doubled; eight subbands representing the real part and eight subbands representing the imaginary part. The lowpass (LL) subband at each level constitutes the input for the next decomposition one, so only the lowpass approximation of the last level is finally preserved.

Fusion takes place in a similar way as in the discrete wavelet transform, combining the corresponding coefficients at each tree. The three fusion rules used with the DWT can also be used in this case. However they must be applied to the magnitude of the coefficients as they are complex numbers. Theoretically, the DT-CWT should produce improved results over the corresponding DWT scheme due to its shift-invariance. This has, and will be, tested in the sections that follow.

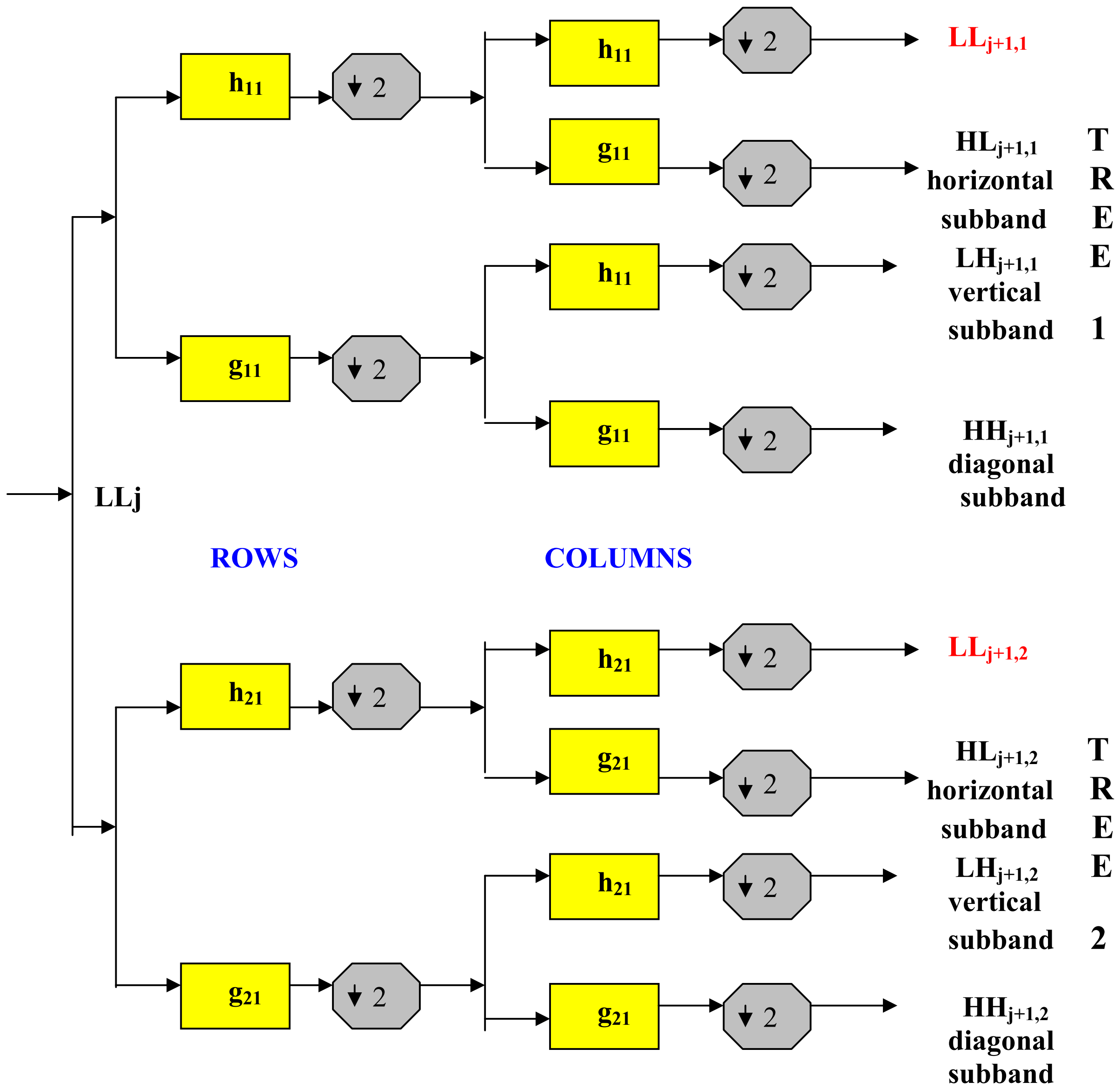
Putting these all together in a tangible and complete scheme gives us a clearer point of view:



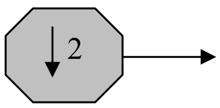
In this way the image structure, level by level, gets the following sequence:



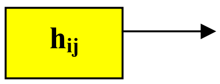
The only limitation in the filter structure is that different filters are used for all the stages after the first level of decomposition and that different filters are applied to each tree. In the synthesis procedure, in a similar way, the last decomposition level uses different filters from the previous ones and each tree has its own reconstruction filter structure. The schematic representation of the analysis/synthesis procedure gets the following form:



Decomposition procedure at level j

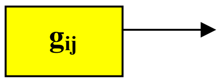


downsapling of the corresponding subimage



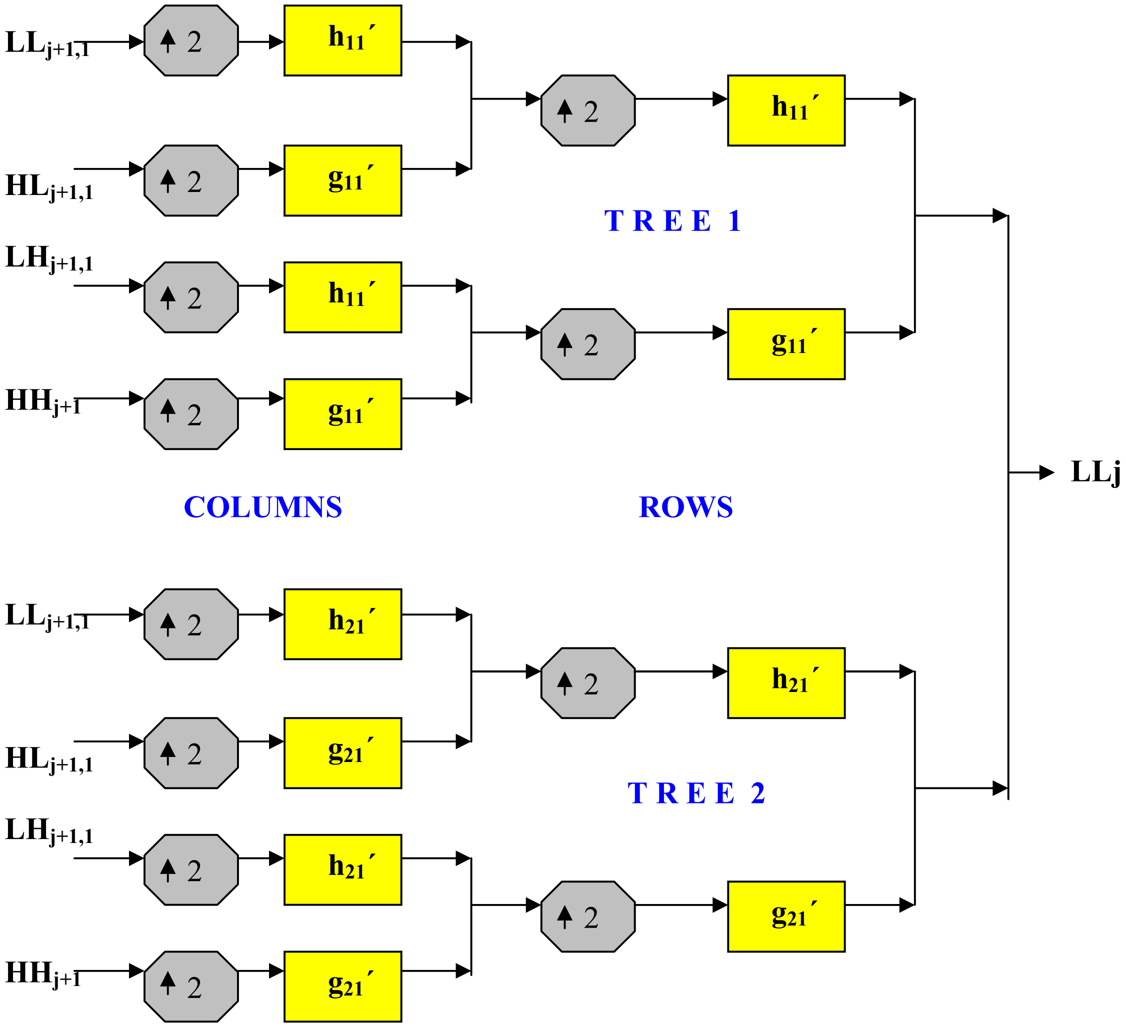
j=current decomposition level, **j+1**=next decomposition level
lowpass analysis/decomposition filter, convolution with the input image
h_{ij} lowpass filter for tree i (i=1, 2) at level j

(j=1 for the first decomposition level and 2 for the other levels)

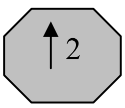


highpass analysis/decomposition filter, convolution with the input image
g_{ij} highpass filter for tree i (i=1, 2) at level j

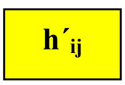
Having decomposed the original images, the wavelet coefficients in each of the six highpass subbands and the final lowpass one are fused and the reconstruction procedure is performed in a similar way, by upsampling and filtering the fused coefficients level to level.



Reconstruction procedure at level j

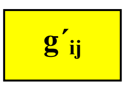


Upsampling of the corresponding subband (insert zeros at odd-indexed columns/rows)



Convolution with lowpass filter **h′**$_{ij}$ for tree i (i=1,2) at level j

(j=1 for the last reconstruction level and 2 for the other levels)



Convolution with highpass filter **g′**$_{ij}$ for tree i (i=1, 2) at level j

Remark: if we use "complex" wavelets then 12 highpass subbands will be created, 6 for the real part of the coefficient and six for the imaginary part of it. To eliminate this great number of subbands, which adds complexity, we can take the magnitude of the coefficients, or use real number as coefficients, leading to the **Dual-Tree Real Wavelet Transform**.

2.1.3.2. Implementation of the Dual-Tree Complex Wavelet Transform

There are two versions of the Dual Tree Complex Wavelet Transform: the Real one, which is 2-times expansive (comparing to the regular Discrete Wavelet Transform presented in the previous section) and the Complex one, which is 4-times expansive. As met in theory the filters of the first level differ from the ones of the other stages. The filters used for the first level and the other stages are chosen according to the analysis of Kingsbury [13].

Decomposition stage Lowpass filters			
TREE 1		TREE 2	
1st level	other	1st level	other
0	0.03516384000000	0.01122679215254	0
-0.08838834764832	0	0.01122679215254	0
0.08838834764832	-0.08832942000000	-0.08838834764832	-0.11430184000000
0.69587998903400	0.23389032000000	0.08838834764832	0
0.69587998903400	0.76027237000000	0.69587998903400	0.58751830000000
0.08838834764832	0.58751830000000	0.69587998903400	0.76027237000000
-0.08838834764832	0	0.08838834764832	0.23389032000000
0.01122679215254	-0.11430184000000	-0.08838834764832	-0.08832942000000
0.01122679215254	0	0	0
0	0	0	0.03516384000000

Reconstruction stage Lowpass filters			
TREE 1		TREE 2	
1st level	other	1st level	other
0	0	0	0.03516384000000
0.01122679215254	0	0	0
0.01122679215254	-0.11430184000000	-0.08838834764832	-0.08832942000000
-0.08838834764832	0	0.08838834764832	0.23389032000000
0.08838834764832	0.58751830000000	0.69587998903400	0.76027237000000
0.69587998903400	0.76027237000000	0.69587998903400	0.58751830000000
0.69587998903400	0.23389032000000	0.08838834764832	0
0.08838834764832	-0.08832942000000	-0.08838834764832	-0.11430184000000
-0.08838834764832	0	0.01122679215254	0
0	0.03516384000000	0.01122679215254	0

Decomposition stage Highpass filters			
TREE 1		TREE 2	
last level	other	last level	other
0	0	0	-0.03516384000000
-0.01122679215254	0	0	0
0.01122679215254	-0.11430184000000	-0.08838834764832	0.08832942000000
0.08838834764832	0	-0.08838834764832	0.23389032000000
0.08838834764832	0.58751830000000	0.69587998903400	-0.76027237000000
-0.69587998903400	-0.76027237000000	-0.69587998903400	0.58751830000000
0.69587998903400	0.23389032000000	0.08838834764832	0
-0.08838834764832	0.08832942000000	0.08838834764832	-0.11430184000000
-0.08838834764832	0	0.01122679215254	0
0	-0.03516384000000	-0.01122679215254	0

Reconstruction stage Highpass filters			
TREE 1		TREE 2	
last level	other	last level	other
0	-0.03516384000000	-0.01122679215254	0
-0.08838834764832	0	0.01122679215254	0
-0.08838834764832	0.08832942000000	0.08838834764832	-0.11430184000000
0.69587998903400	0.23389032000000	0.08838834764832	0
-0.69587998903400	-0.76027237000000	-0.69587998903400	0.58751830000000
0.08838834764832	0.58751830000000	0.69587998903400	-0.76027237000000
0.08838834764832	0	-0.08838834764832	0.23389032000000
0.01122679215254	-0.11430184000000	-0.08838834764832	0.08832942000000
-0.01122679215254	0	0	0
0	0	0	-0.03516384000000

Taking a closer look to the coefficients of the filters it is noticed that the synthesis filters, for both trees and in all levels, are the analysis filters of the corresponding trees and level reversed. In addition, the analysis filters after the first level of tree 1 are the synthesis filters of tree 2 and vice versa! Finally, at the first level of analysis and the last level of synthesis, the reconstruction filters of tree 2/tree1 are the decomposition filters of tree1/tree 2 delayed by one sample. These conditions meet the requirements indicated by theory.

Let's analyze each version:

Real Dual Tree Wavelet Transform

The Real Dual Tree Wavelet Transform (or simply Dual Tree Transform) is implemented using two separable DWTs in parallel, as in the scheme above. After the decomposition stage, the fusion process takes places and the fused image is reconstructed through the synthesis scheme illustrated before. The quality metric Qf is adopted again to evaluate the performance of the algorithm.

Complex Dual Tree Wavelet Transform

The Complex Dual Tree Wavelet Transform also produces six distinct directions, subbands, with the difference that there are two wavelets in each direction, the real and imaginary part of the complex wavelet. Because the complex version has twice as many wavelets as the real version, the complex version is 4-times expansive (comparing to the regular Discrete Wavelet Transform), which means four separable DWTs in parallel. The filter structure is the same as presented before. Different filters are applied at the first/last level of decomposition/reconstruction and at the intermediate levels. In addition, each tree uses different filters too. In our implementation we added one more condition to avoid dublication of results comparing to the real Dual Tree Transform: **different filters are applied to the rows and the columns**.

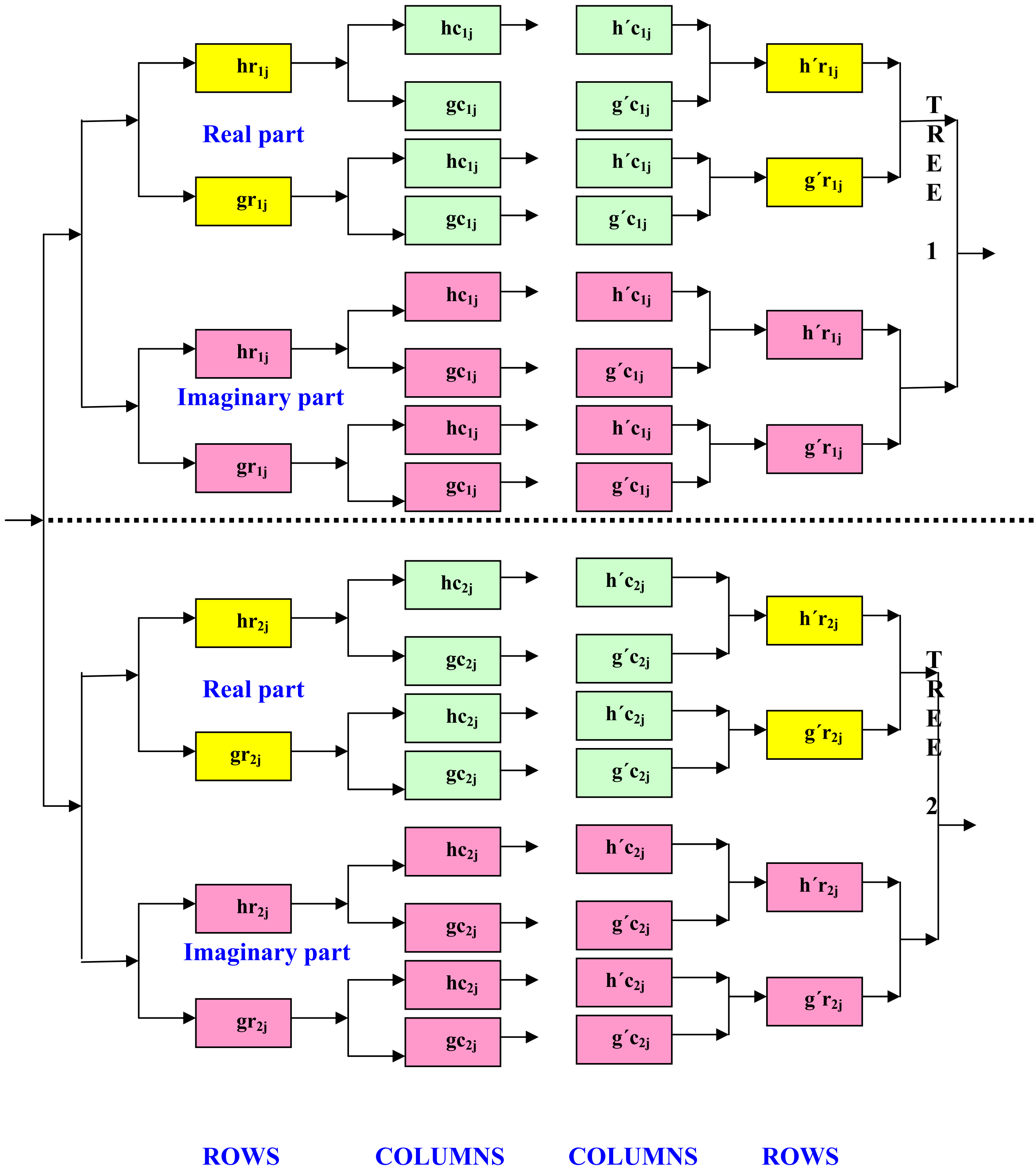
Remark: In fact, the Dual-Tree Wavelet Transform is a "pseudo-complex" wavelet transform because it does not make use of complex wavelet families. Complex analysis and synthesis filters are very difficult to be created by complex wavelets, only theoretically. Trying to implement the algorithm using the Matlab toolbox, only real filters, derived from wavelet families, are supported. Thus, we "simulate" the complex transform using real wavelets-filters. This is achieved by considering the outputs of the analysis procedure as the real and imaginary parts of complex coefficients. We use a pair of real lowpass and highpass filters to simulate each part of a hypothetical complex coefficient: the real and imaginary part.

The structure of these real filters is a key point for the efficiency of the algorithm. They must satisfy some certain conditions. This analysis will not be mentioned here. Instead, two very good sources for further reading is proposed: **"Image processing using complex wavelets"** by Nick Kingsbury, 1999, pp 2543-2560 and **"Complex wavelet transforms with allpass filters"**, Felix C. Fernandes and Ivan W. Selesnick, Rutger L.C. van Spaendonck, C. Sidney Burrus, December 18, 2002

In general, the DT-CWT has excellent directionality, reduced shift sensitivity and explicit phase information but has the drawback that is redundant because the transform coefficients require more storage space than the input signal.

In the following figure the structure of a tree produced by the Dual Tree Complex Wavelet Transform is illustrated. Due to limited space and in order to make the decomposition procedure clear to the reader only the one tree is illustrated. In addition, the downsampling procedure is not illustrated but, of course, is a key part of the decomposition process. The structure of the second tree is similar with the difference that different filters are applied. Notice that different filters are applied to the two directions (rows and columns of the input image matrix). The notation used is described below:

Analysis		Synthesis	
hr_{ij}	,	$h'r_{ij}$	→ Lowpass filter for tree i (i=1, 2) at level j applied to rows
hc_{ij}	,	$h'c_{ij}$	→ Lowpass filter for tree i (i=1, 2) at level j applied to columns
gr_{ij}	,	$g'r_{ij}$	→ Highpass filter for tree i (i=1, 2) at level j applied to rows
gc_{ij}	,	$g'c_{ij}$	→ Highpass filter for tree i (i=1, 2) at level j applied to columns

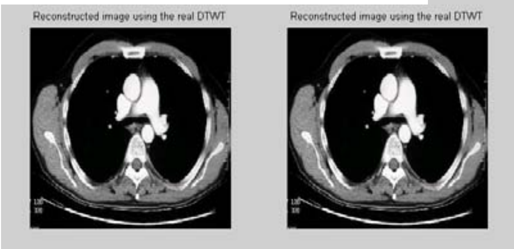


***the implementation of the real and complex filter part is different, that's why a different color is used.

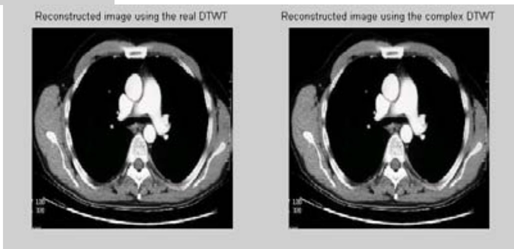
Decomposition/reconstruction stages

TEST 1

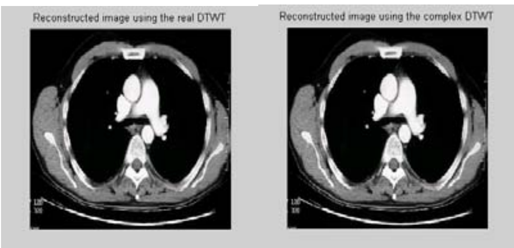
Real and Complex Dual Tree Wavelet Transform in the analysis/synthesis process



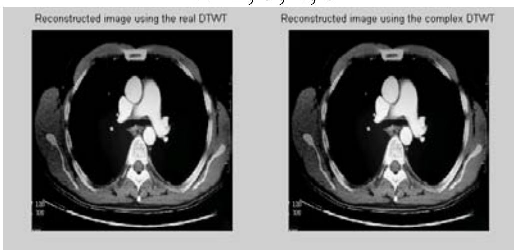
N=1



N=2, 3, 4, 5



N=6



N=7

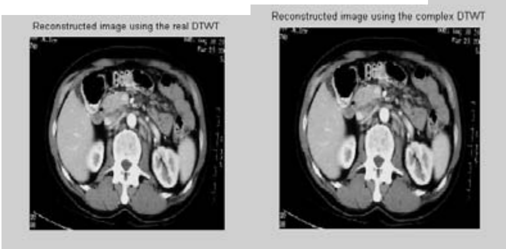
Number Of Levels	**RMSE normalized values**		**Q normalized values**	
DTWT	**Real**	**Complex**	**Real**	**Complex**
N=1	0.0073	0.0073	0.8152	0.8152
N=2	**0.0038**	**0.0038**	**0.9188**	**0.9188**
N=3	**0.0038**	**0.0038**	**0.9188**	**0.9188**
N=4	**0.0038**	**0.0038**	**0.9188**	**0.9188**
N=5	**0.0038**	**0.0038**	**0.9188**	**0.9188**
N=6	0.0086	0.0086	0.8529	0.8529
N=7	0.1081	0.1076	0.5596	0.5583

System parameters: **size of filters = 10**
wavelet family = Kingsbury defined

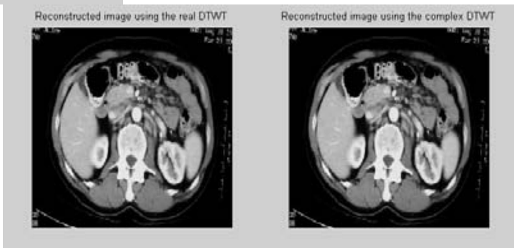
TEST 2

Real and Complex Dual Tree Wavelet Transform in the analysis/synthesis process

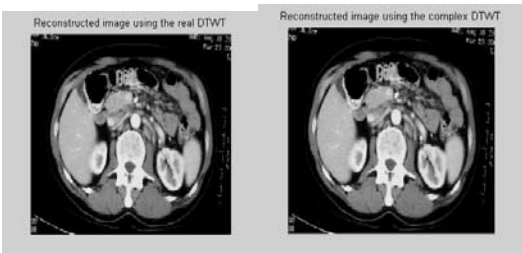
Change the input image



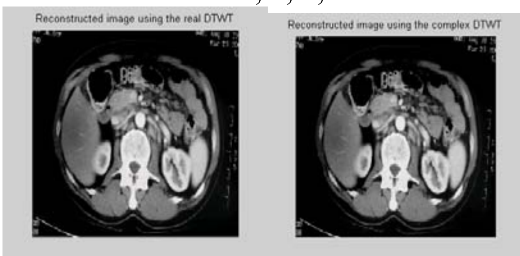
N=1



N=2, 3, 4, 5



N=6

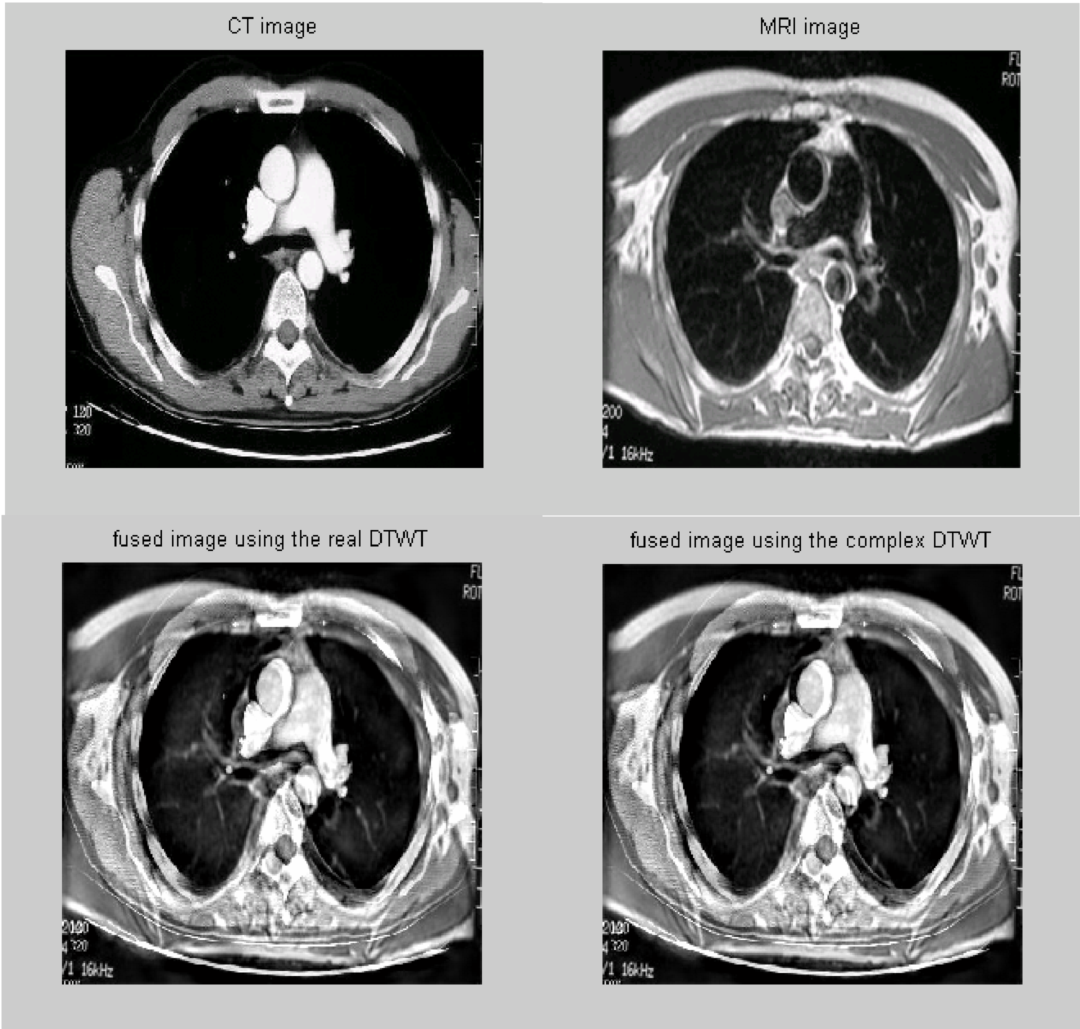


N=7

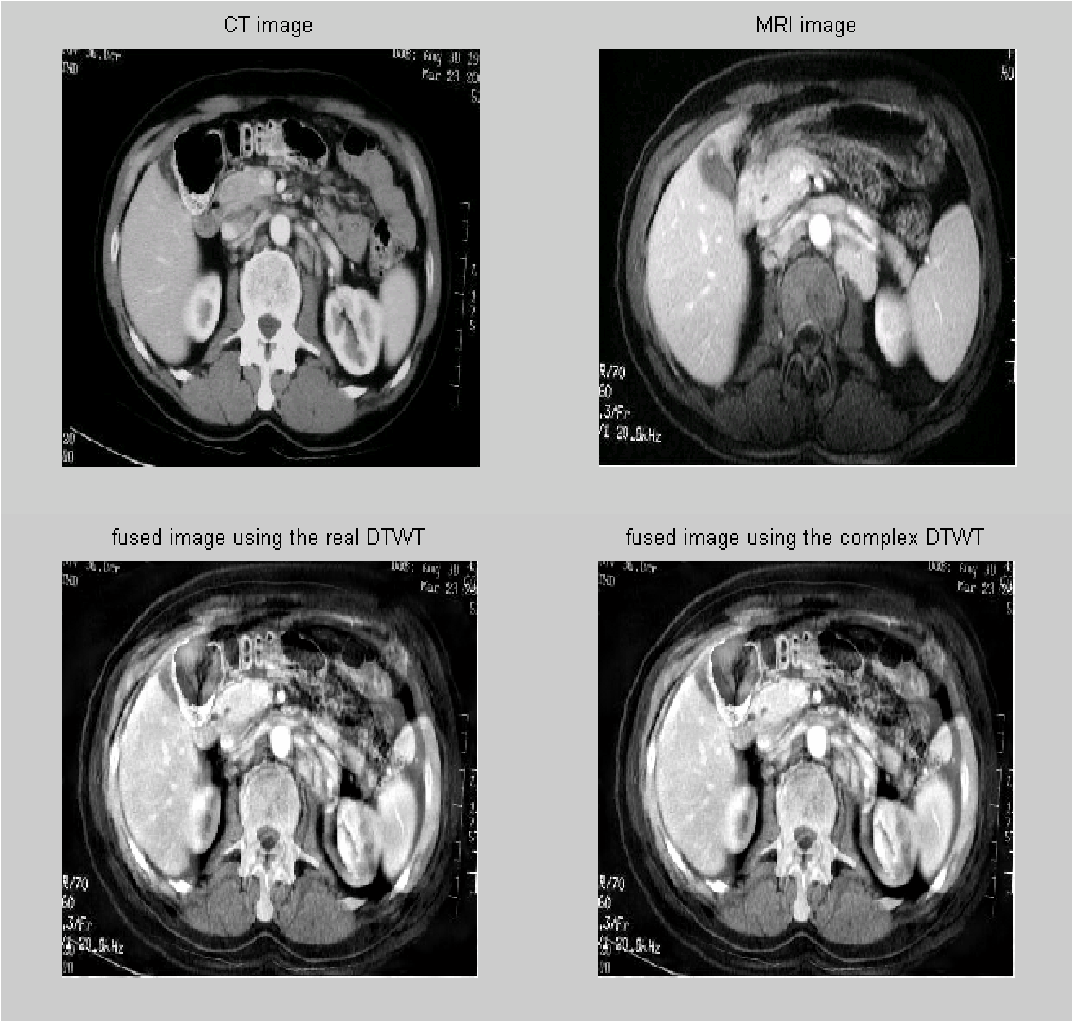
Number Of Levels	**RMSE normalized values**		**Q normalized values**	
DTWT	**Real**	**Complex**	**Real**	**Complex**
N=1	0.0066	0.0066	0.9104	0.9104
N=2	**0.0034**	**0.0034**	**0.9469**	**0.9469**
N=3	**0.0034**	**0.0034**	**0.9469**	**0.9469**
N=4	**0.0034**	**0.0034**	**0.9469**	**0.9469**
N=5	**0.0034**	**0.0034**	**0.9469**	**0.9469**
N=6	0.0069	0.0069	0.8942	0.8940
N=7	0.0933	0.0941	0.7295	0.7350

System parameters: **size of filters** = **10**
wavelet family = **Kingsbury defined**

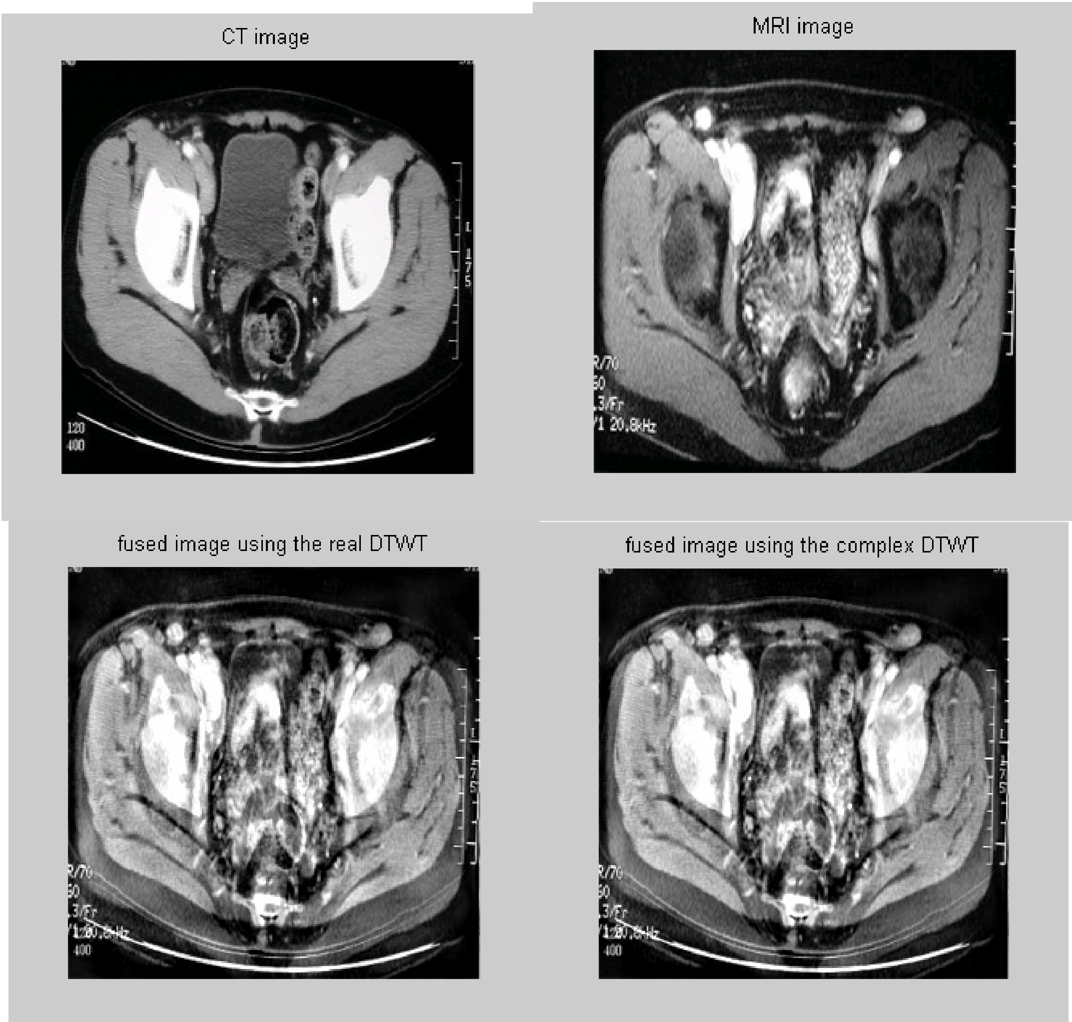
TEST 3
Real and Complex Dual Tree Wavelet Transform in image fusion
Number of levels=5



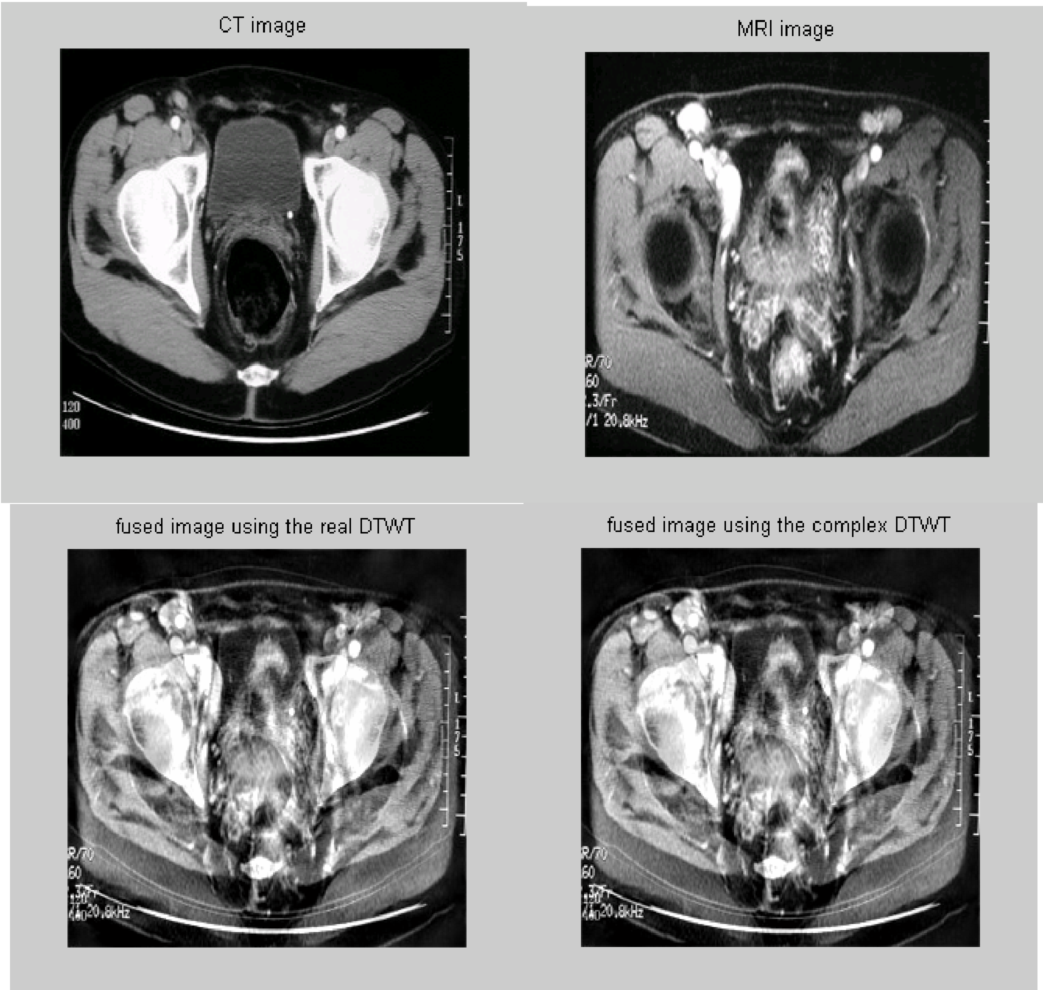
Qf=0.6289 **Qf=0.6534**



Qf=0.6161 **Qf=0.6295**



Qf=0.6509 **Qf=0.6697**

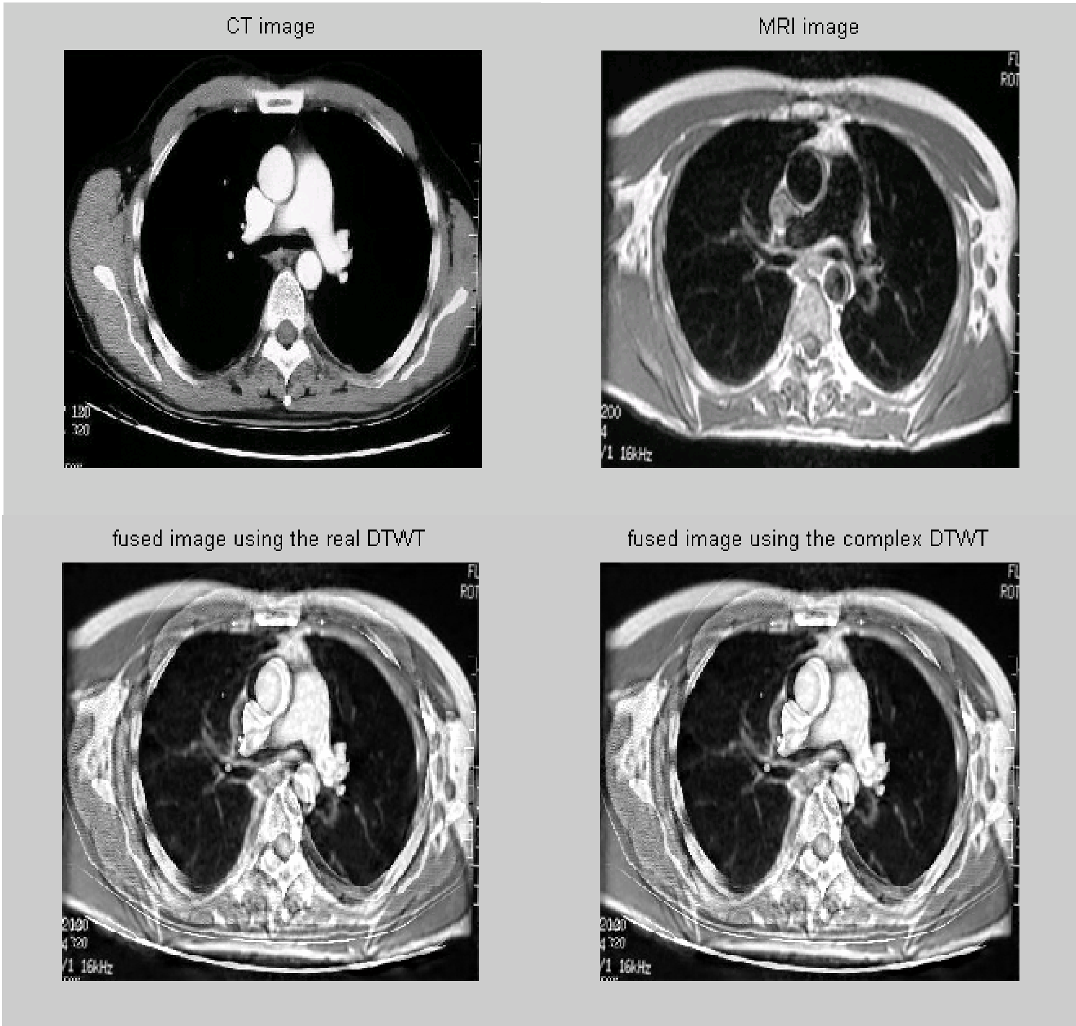


Qf=0.6380 **Qf=0.6643**

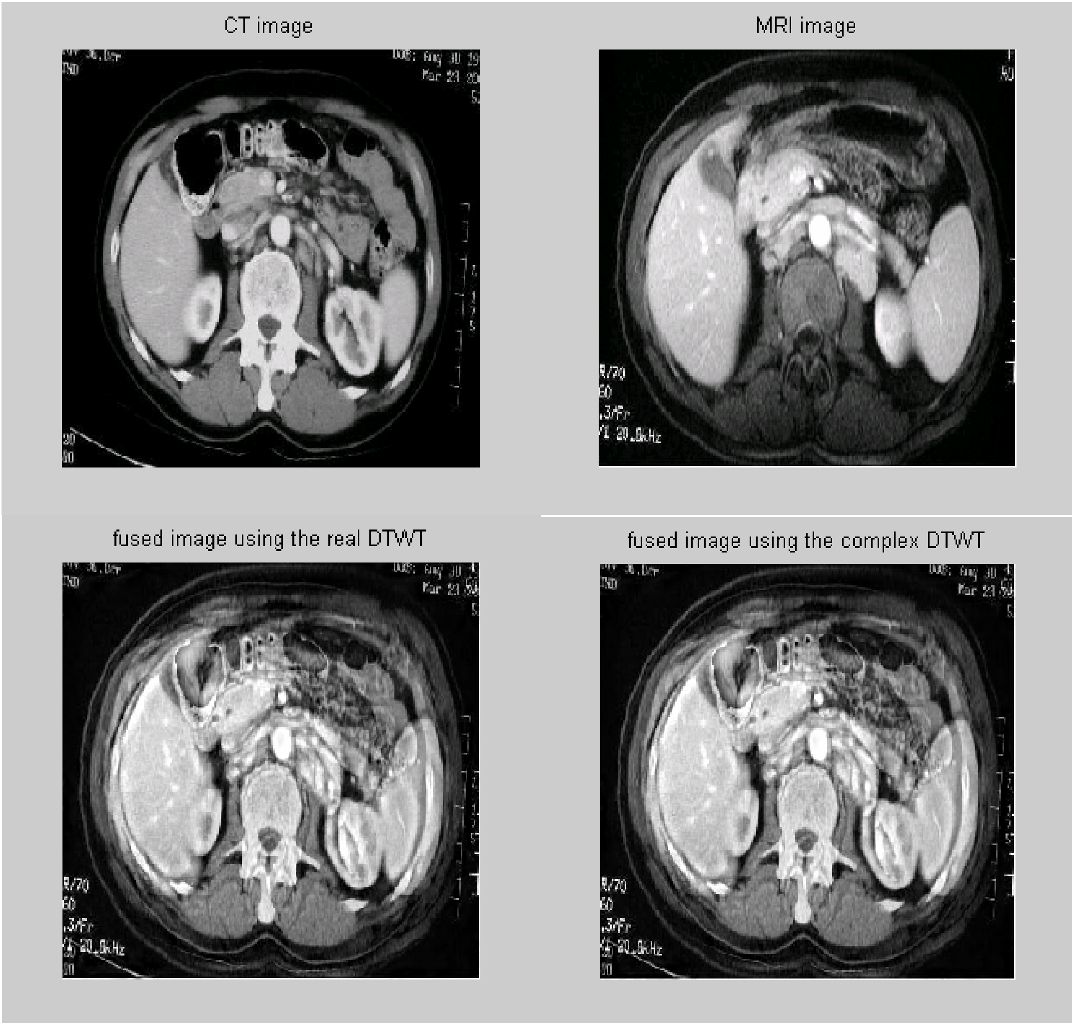
TEST 4

Real and Complex Dual Tree Wavelet Transform in image fusion

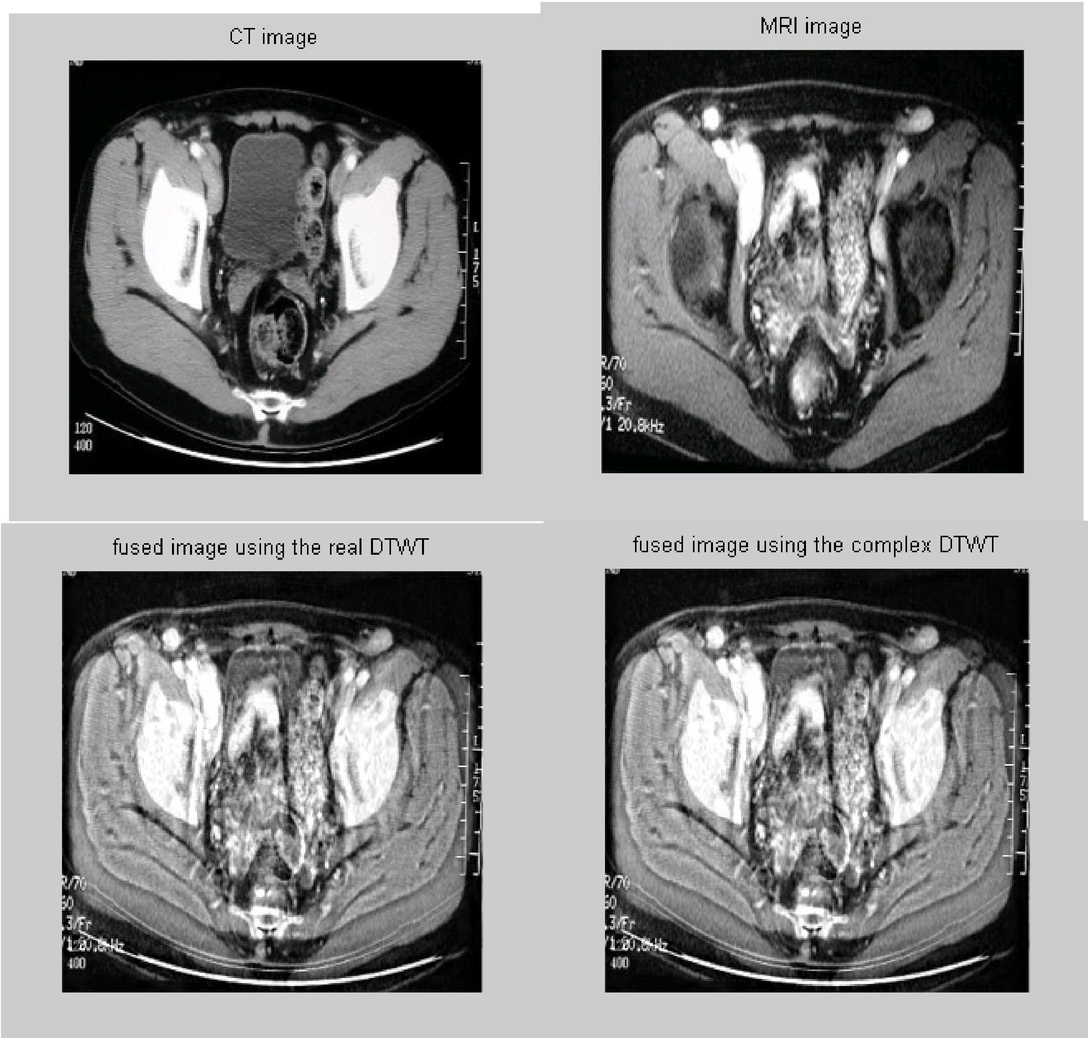
Number of levels=3



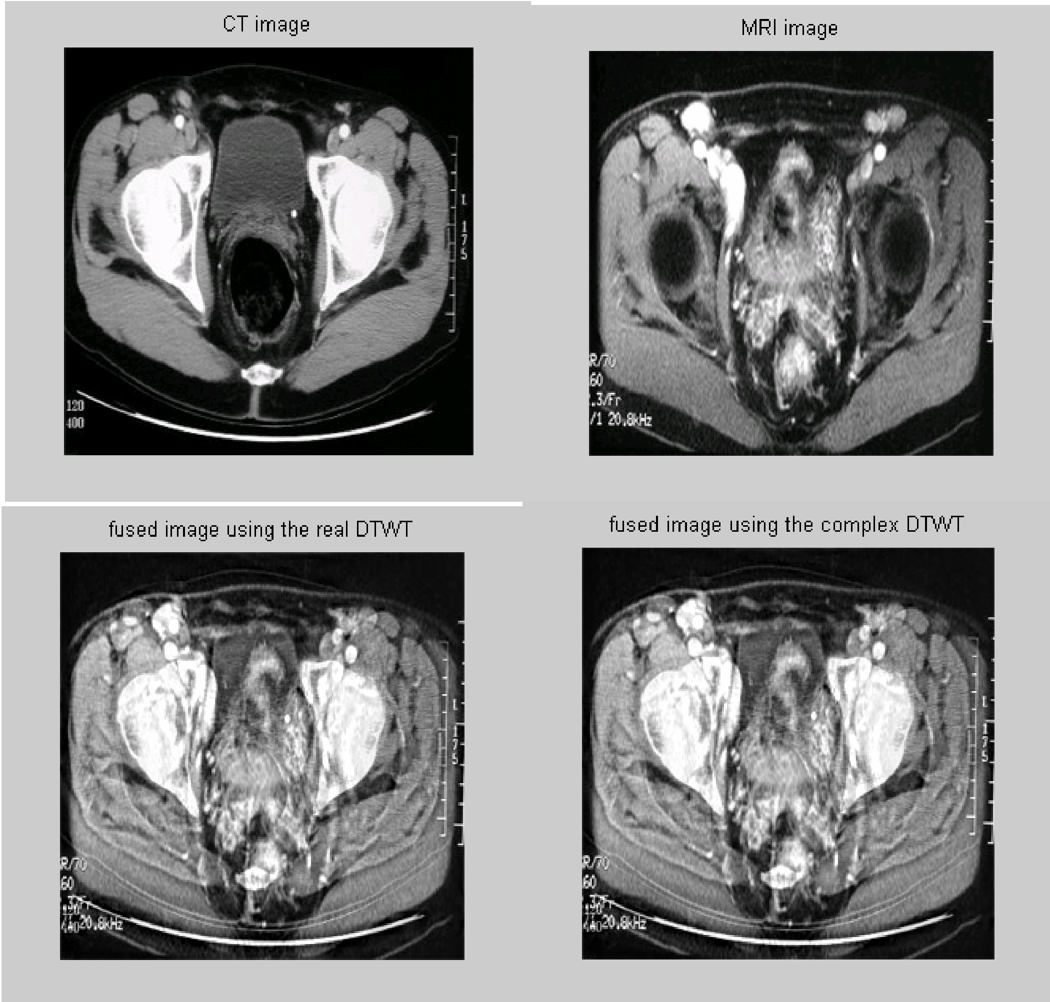
Qf=0.7065 **Qf=0.7163**



Qf=0.6173 **Qf=0.6258**



Qf=0.6689 **Qf=0.6773**



Qf=0.6610 **Qf=0.6691**

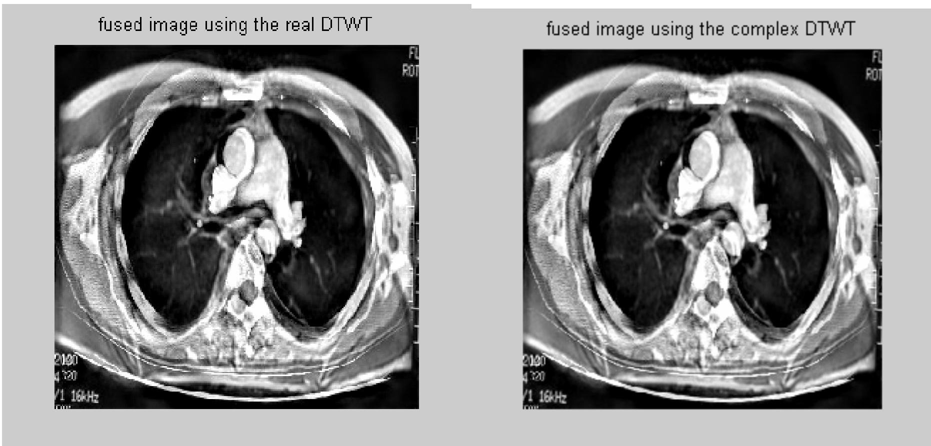
Quality metric Qf			
Levels	**Image pair**	**Real DTWT**	**Complex DTWT**
N=3	**1**	**0.7065**	**0.7163**
	2	**0.6173**	**0.6258**
	3	**0.6689**	**0.6773**
	4	**0.6610**	**0.6691**
N=5	**1**	**0.6289**	**0.6534**
	2	**0.6161**	**0.6295**
	3	**0.6509**	**0.6697**
	4	**0.6380**	**0.6643**

2.1.3.3 Conclusions

From the section before we could make some useful and remarkable conclusions. First of all we notice the impressive result (from TEST 1 and TEST 2) that the two versions of the Dual-Tree Complex Wavelet Transform perform equally well during the decomposition/reconstruction process but the Complex version outperforms the Real one when the fusion process is involved. Another remarkable result is that the two transforms have exactly the same behavior, the same values of metrics, at levels 2 to 5!!! Overcoming this range the difference in performance is obvious in the reconstructed images too. The fact that in the analysis/synthesis procedure the same values of evaluation are met between the levels 2-5 is not a coincidence; the same result was extracted using another input image in TEST 2. This is due to the similar structure of the decomposition/reconstruction process. Using totally different filters along the real and imaginary "parts", along the rows and columns, this can change.

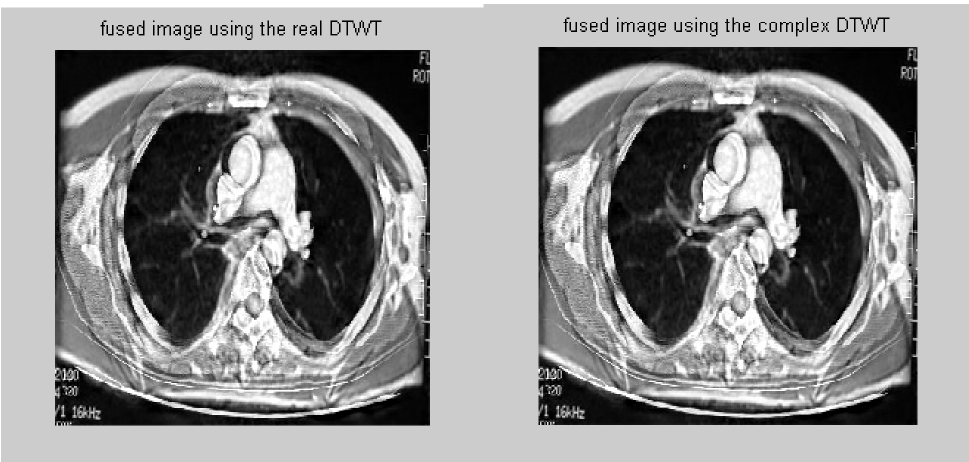
Finally, changing the number of levels, in particular reducing the number from 5 to 3, has a small effect in the final result. Selecting a lower level the performance gets slightly improved.

N=5



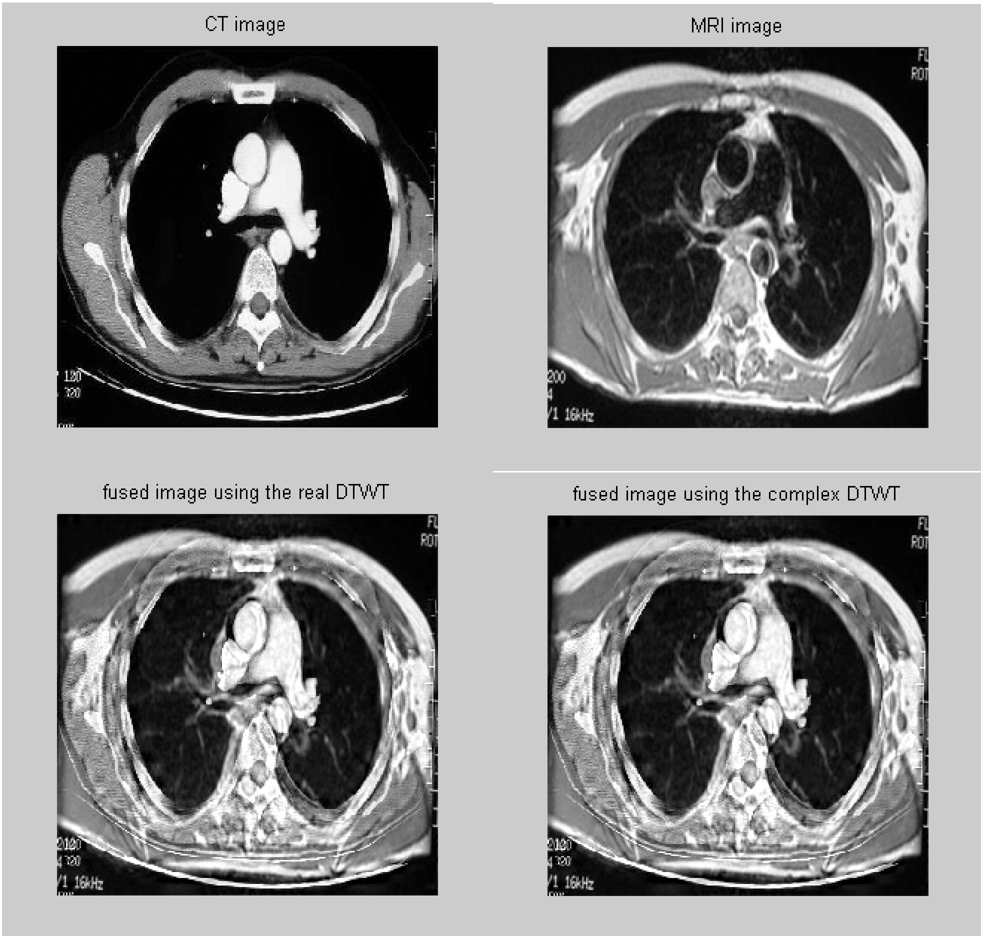
Qf=0.6289 **Qf=0.6534**

N=3

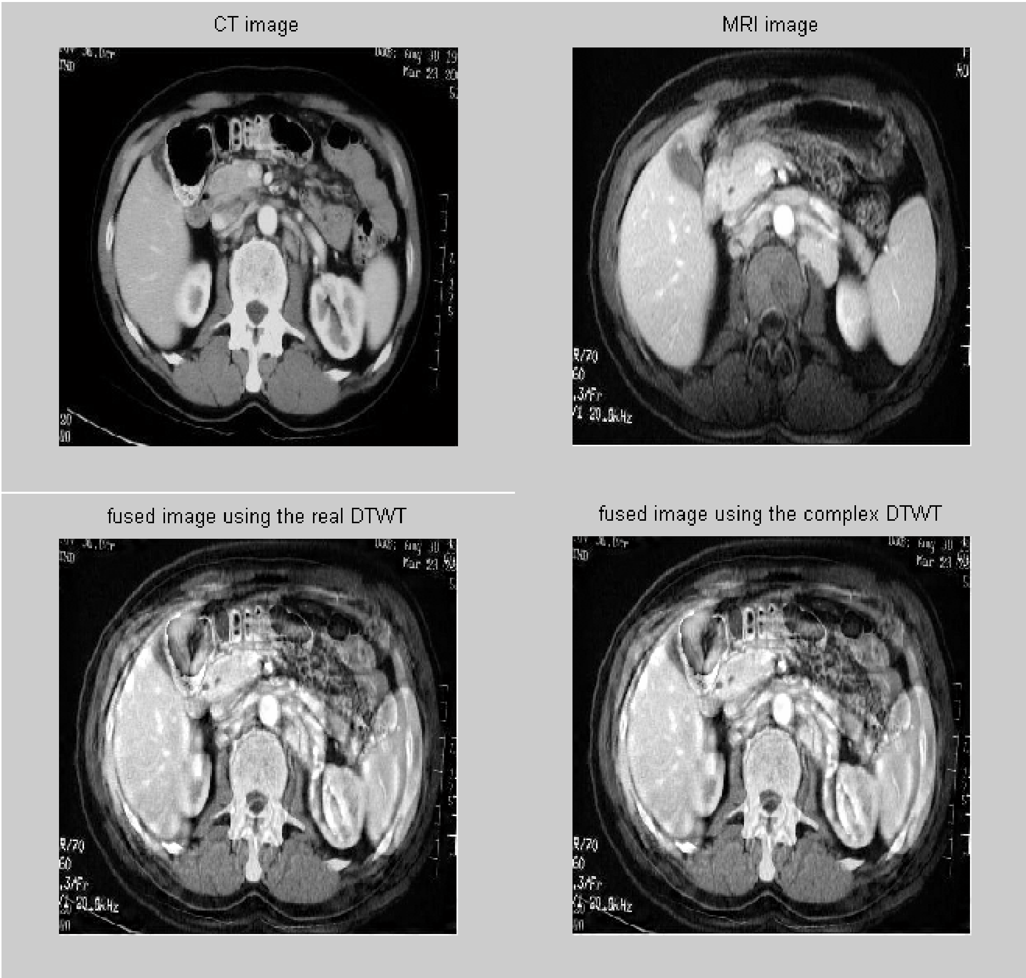


Qf=0.7065 **Qf=0.7163**

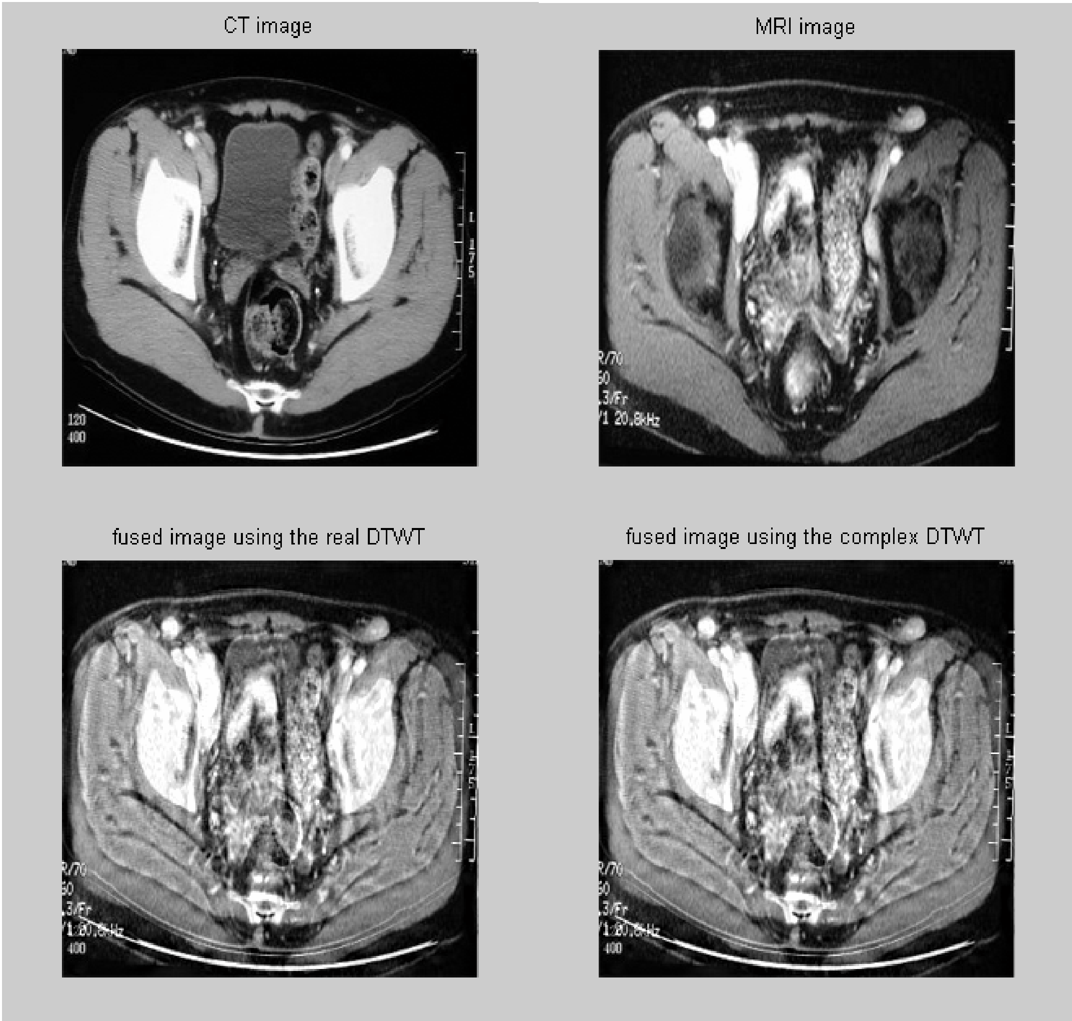
Another issue that is worth discussing is the contribution of the filters at the different stages. As mentioned in theory, it is preferable to use different filters in the first/last decomposition/reconstruction procedure from the ones used in the intermediate stages. This is proved practically, evaluating the performance of the Dual Tree Discrete Wavelet Transform when the same filter structure is used among all the analysis/synthesis stages. The results from the application of the above concept are illustrated below:



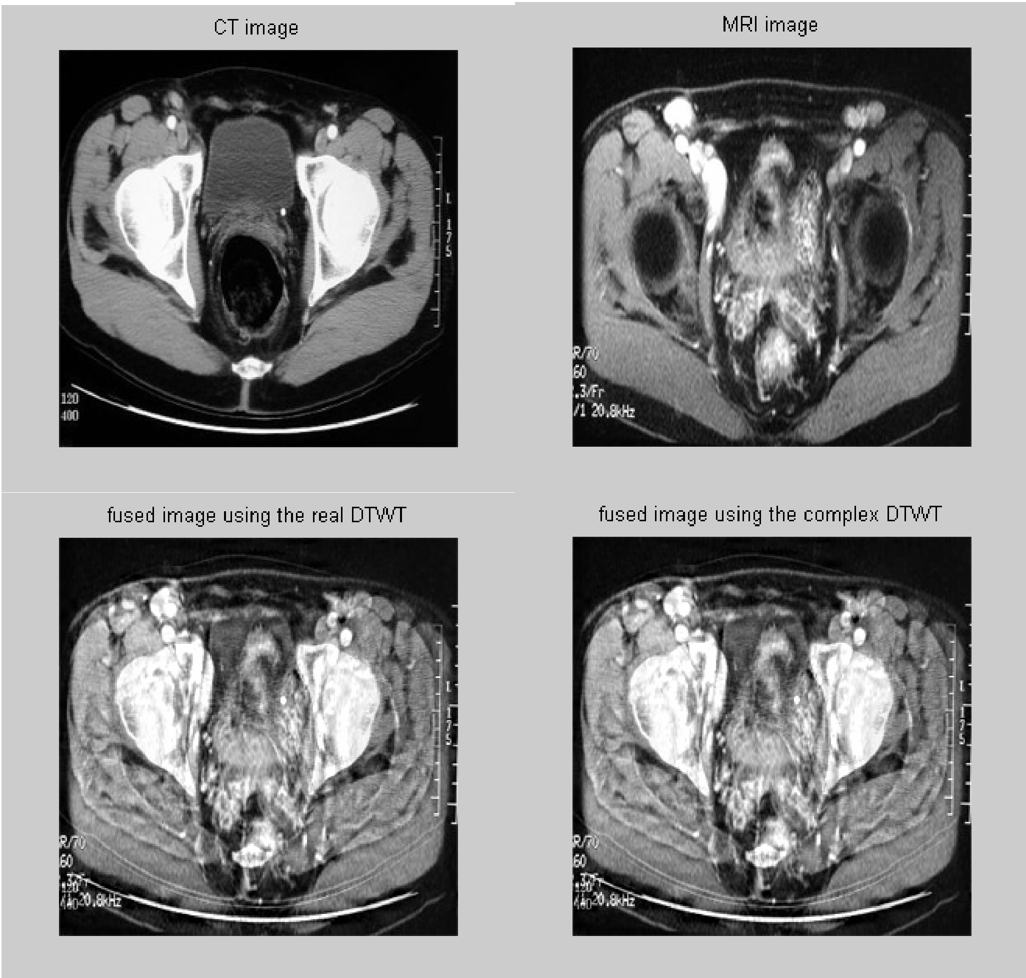
Qf=0.6979 **Qf=0.7050**



Qf=0.6109 **Qf=0.6156**



Qf=0.6580 **Qf=0.6634**

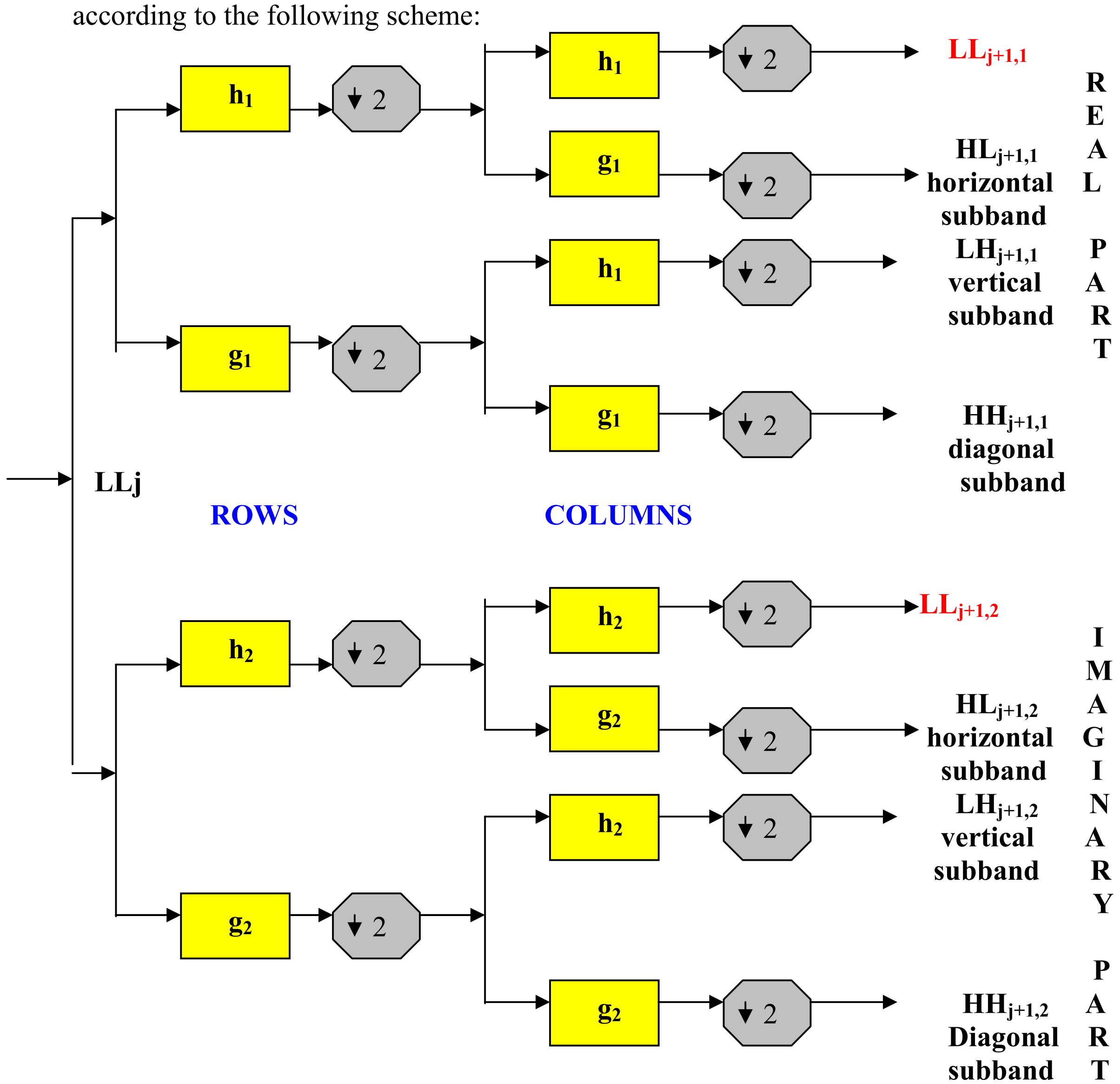


Qf=0.6511 **Qf=0.6566**

Levels = 3		**Quality metric Qf**	
	Image pair	**Real DTWT**	**Complex DTWT**
Use different stage filters	1	0.7065	0.7163
	2	0.6173	0.6258
	3	0.6689	0.6773
	4	0.6610	0.6691
Use same stage filters	1	0.6979	0.7050 ↓
	2	0.6109	0.6156 ↓
	3	0.6580	0.6534 ↓
	4	0.6511	0.6566 ↓

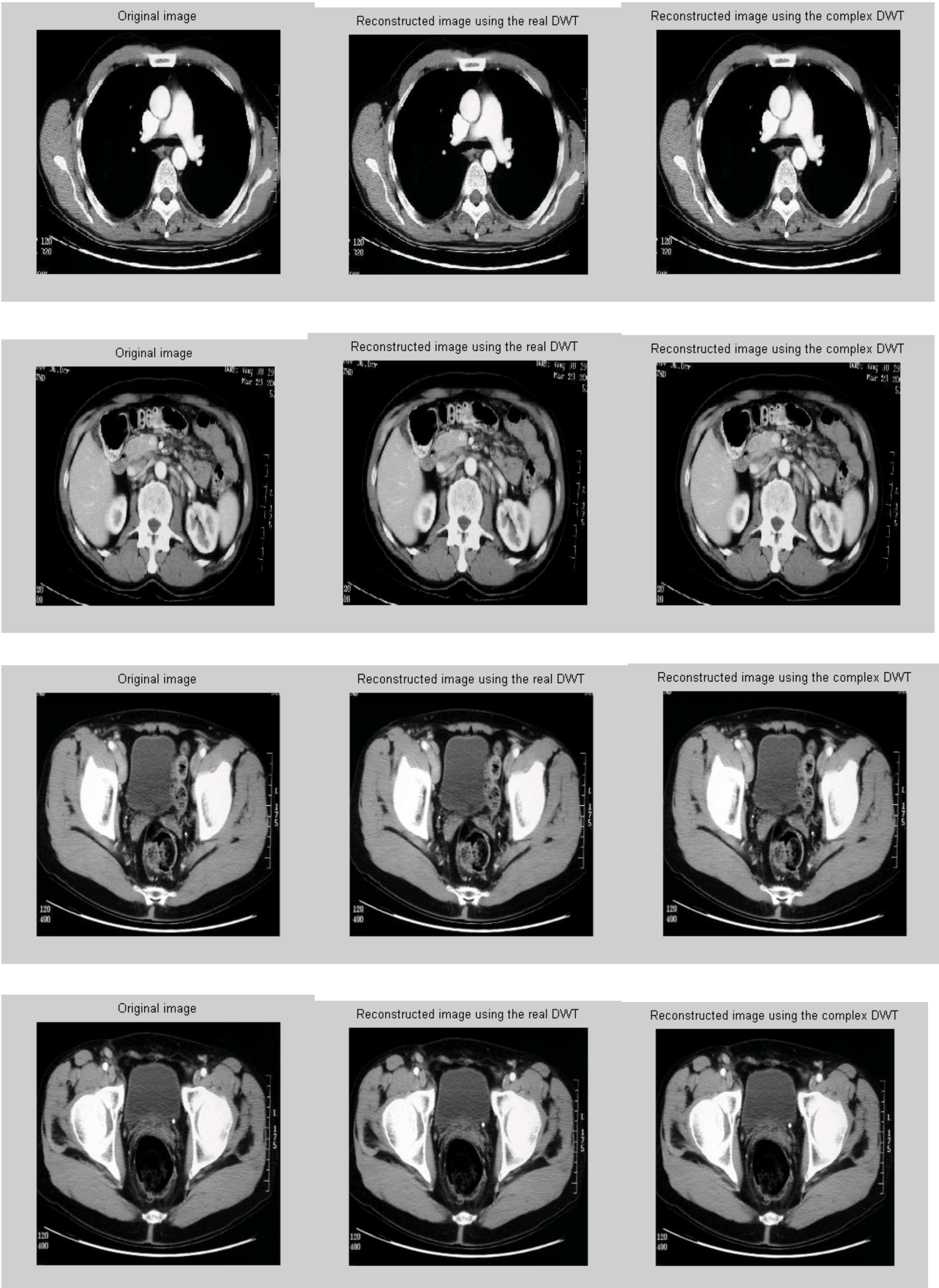
2.1.4. Real versus Complex wavelet transform implementation

As mentioned in section 2.1.1, complex wavelet transforms offer basic advantages over real ones but are not easily applicable, especially in the decomposition-filtering-reconstruction procedure described. **In practice, trying to obtain the filter coefficients using the Matlab toolbox was not possible as only real wavelet analysis/synthesis filters were supported**. However, complex wavelet transforms can be "simulated' using two real wavelets for each component produced considering them as the real and imaginary part of complex coefficients. In this section the two versions of the Discrete Wavelet Transform described (regular DWT and DT-CWT) are compared using the same filter structure, "simulating" a complex wavelet transform version of them. The other version, the Shift-Invariant Discrete Wavelet Transform, is preferable not to be expanded in a complex form, because the structure of the filters needed, with zero-padding along different levels, is a little complicated to be achieved. The DT-CWT has already been introduced and the complex Regular Discrete Wavelet Transform is constructed according to the following scheme:

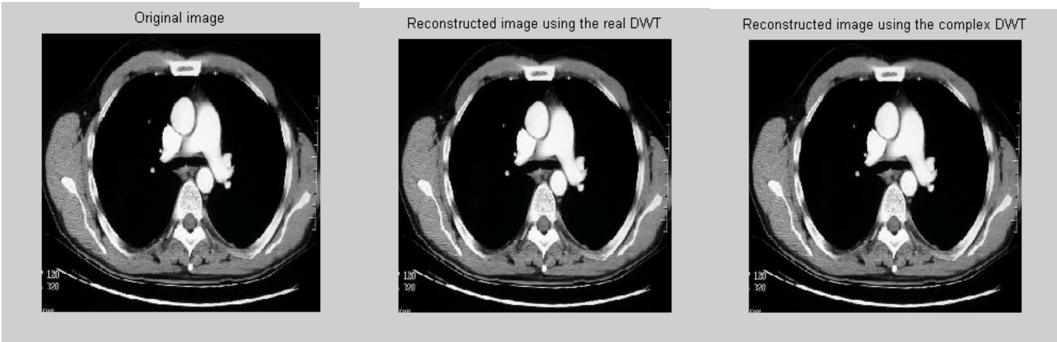


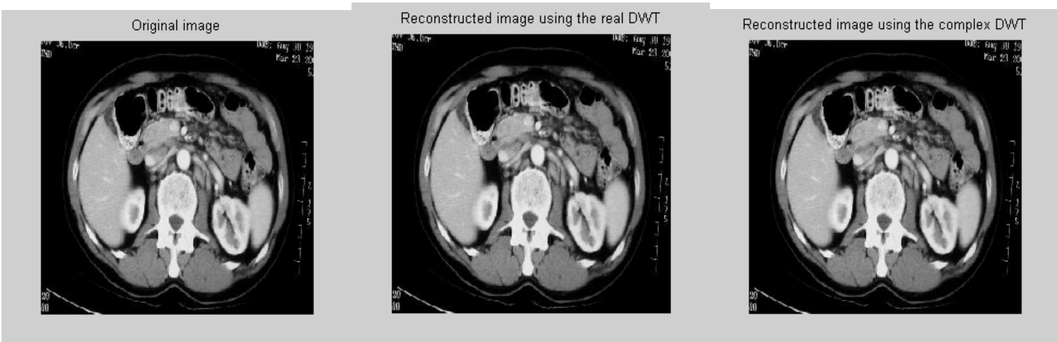
At first, it is important to compare the real and complex version of the Regular Discrete Wavelet Transform, both in the decomposition-reconstruction and the analysis-fusion-synthesis-phase:

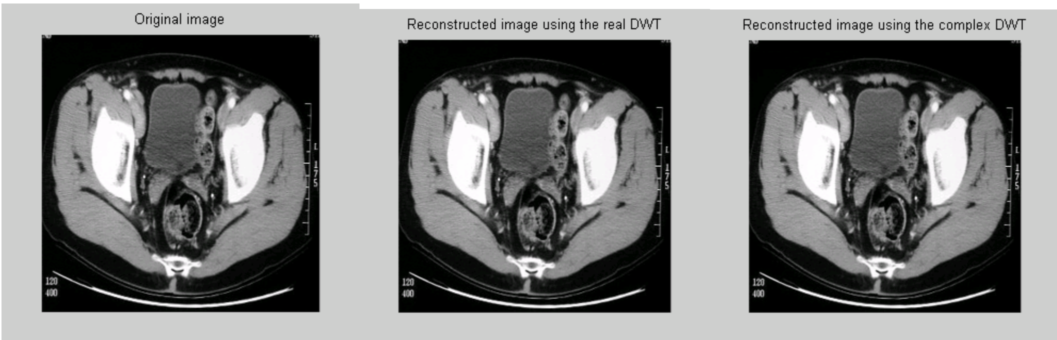
Number of levels=4

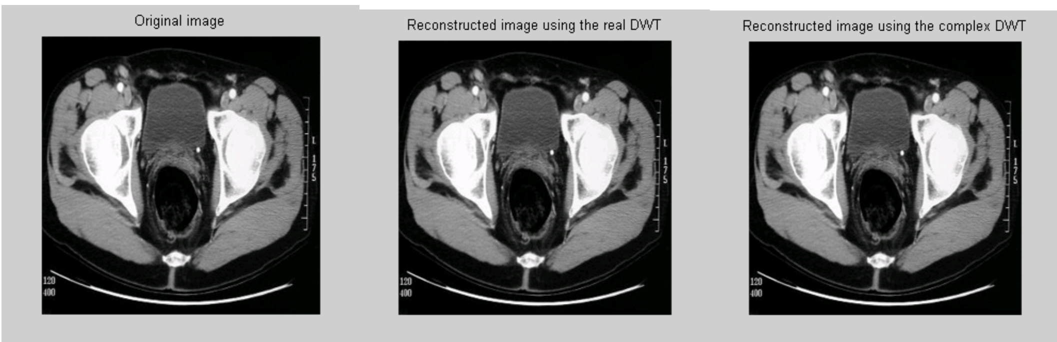


Number of levels=6









The results are summarized in the following table:

Levels	Image pair	Real DWT Q	Real DWT RMSE	Complex DWT Q	Complex DWT RMSE
4	1 2 3 4	0.8152 0.9104 0.9008 0.9211	0.0073 0.0066 0.0075 0.0076	0.8152 0.9104 0.9008 0.9211	0.0073 0.0066 0.0075 0.0076
6	1 2 3 4	0.8431 0.9031 0.8913 0.9199	0.0071 0.0062 0.0071 0.0069	0.8708↑ 0.9278↑ 0.9144↑ 0.9300↑	0.0058↓ 0.0051↓ 0.0060↓ 0.0060↓

Conclusion: It is very interesting to notice that the two versions of the regular Discrete Wavelet Transform perform exactly equally, for all inputs, until a certain number of levels is reached. When the number of levels is increased enough it is getting clear that the complex transform outperforms the real one in a satisfactory way. The theoretical expectation tends to be proved.

It's worth attempting to define the filter structure of the analysis/synthesis procedure for the complex Discrete Wavelet Transform. Different approaches were tested: using different filters for the first/last and the upper/lower decomposition/reconstruction levels and different filter application along rows and columns, as adopted in the implementation of the Dual Tree Complex Wavelet Transform. The results clearly showed that the only condition that gives efficient results is the use of different filters for the imaginary and real parts of the transformation scheme presented above. The conditions of the filter structure are:

- the synthesis filters are the analysis filters inversed
- the analysis filters of the "imaginary part" of the transform are the synthesis filters of the "real part " of the transform delayed by one sample and vice veersa.

The filter structure adopted is the following (it the first stage filter structure of the DT-CWT implementation!):

The filter structure applied to the rows and the columns of the input image matrix is retained the same and the filter structure within the first level and the remaining ones is also kept in the original, fixed form.

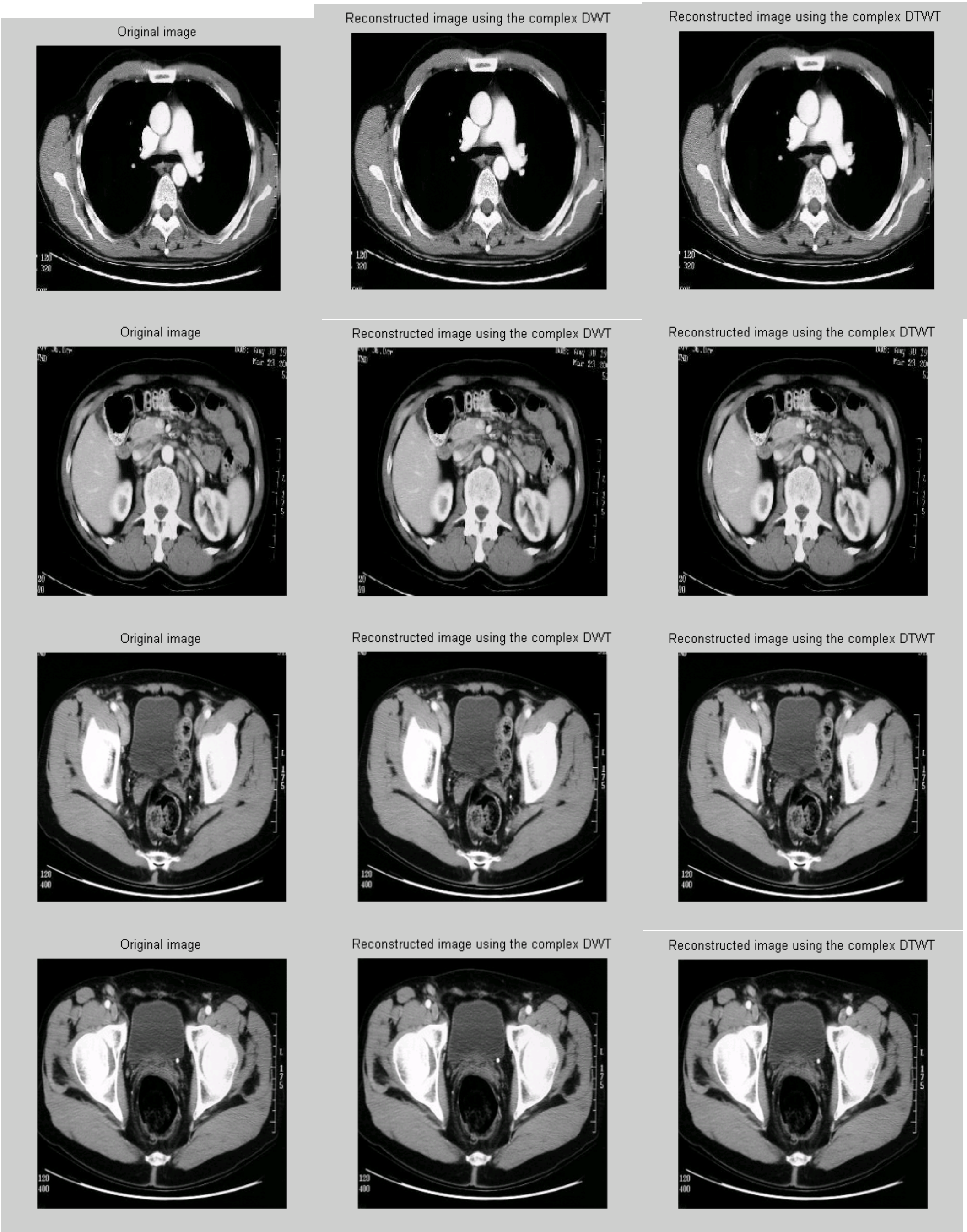
Filter structure for the implementation of the complex DWT

Decomposition stage			
TREE 1 (Real part)		TREE 2(Imaginary part)	
Lowpass filter	Highpass filter	Lowpass filter	Highpass filter
0	0	0.01122679215254	0
-0.08838834764832	-0.01122679215254	0.01122679215254	0
0.08838834764832	0.01122679215254	-0.08838834764832	-0.08838834764832
0.69587998903400	0.08838834764832	0.08838834764832	-0.08838834764832
0.69587998903400	0.08838834764832	0.69587998903400	0.69587998903400
0.08838834764832	-0.69587998903400	0.69587998903400	-0.69587998903400
-0.08838834764832	0.69587998903400	0.08838834764832	0.08838834764832
0.01122679215254	-0.08838834764832	-0.08838834764832	0.08838834764832
0.01122679215254	-0.08838834764832	0	0.01122679215254
0	0	0	-0.01122679215254

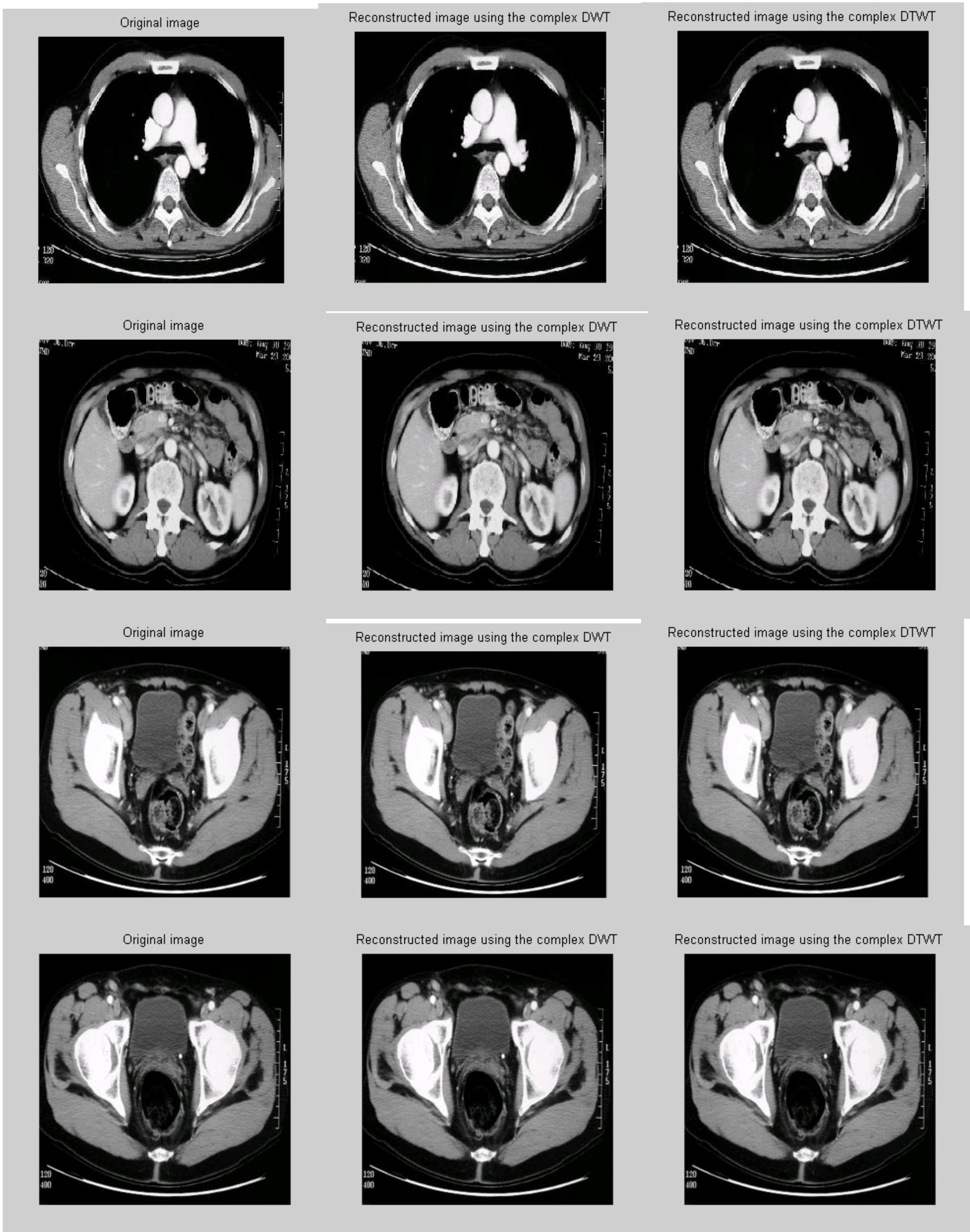
Reconstruction stage Lowpass filters			
TREE 1		TREE 2	
Lowpass filter	Highpass filter	Lowpass filter	Highpass filter
0	0	0	-0.01122679215254
0.01122679215254	-0.08838834764832	0	0.01122679215254
0.01122679215254	-0.08838834764832	-0.08838834764832	0.08838834764832
-0.08838834764832	0.69587998903400	0.08838834764832	0.08838834764832
0.08838834764832	-0.69587998903400	0.69587998903400	-0.69587998903400
0.69587998903400	0.08838834764832	0.69587998903400	0.69587998903400
0.69587998903400	0.08838834764832	0.08838834764832	-0.08838834764832
0.08838834764832	0.01122679215254	-0.08838834764832	-0.08838834764832
-0.08838834764832	-0.01122679215254	0.01122679215254	0
0	0	0.01122679215254	0

Let’s try to compare the Regular DWT and the DT-CWT using complex wavelets in the decomposition/reconstruction procedure (* the filter structure remains the same):

Number of levels =4



Number of levels =6



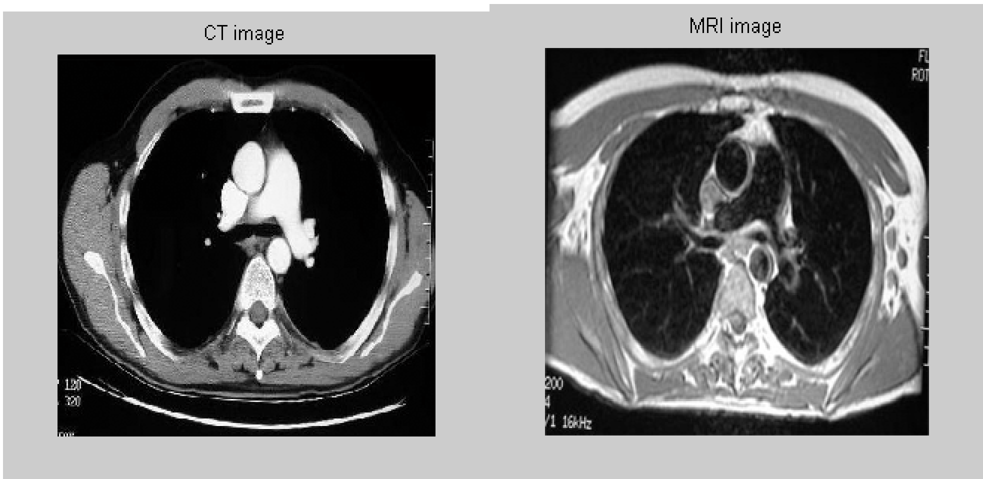
The results are summarized in the following table:

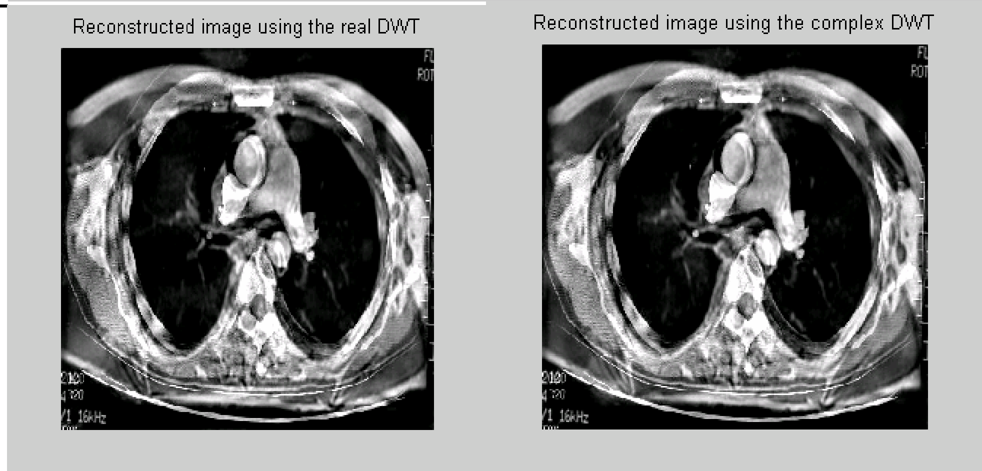
		Complex Dual Tree WT		Complex DWT	
Levels	**Image pair**	**Q**	**RMSE**	**Q**	**RMSE**
4	1	0.9188	0.0038	0.8152	0.0073
	2	0.9469	0.0034	0.9104	0.0066
	3	0.9464	0.0039	0.9008	0.0075
	4	0.9622	0.0039	0.9211	0.0076
6	1	↓ 0.8529	↑0.0086	0.8708↑	0.0058 ↓
	2	↓ 0.8940	↑0.0069	0.9278↑	0.0051 ↓
	3	↓ 0.8596	↑0.0084	0.9144↑	0.0060 ↓
	4	↓ 0.8929	↑0.0082	0.9300↑	0.0060 ↓

Conclusion:

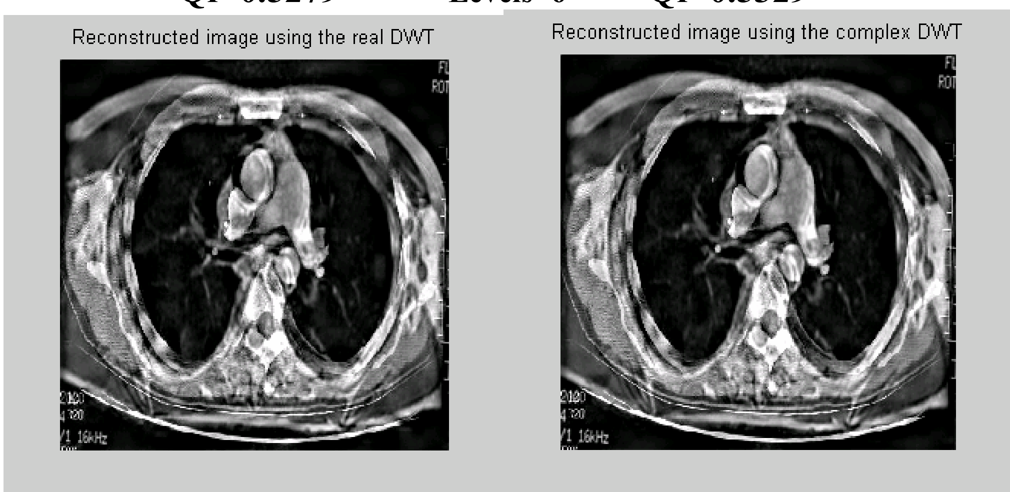
We notice that in the lower levels, the DT-CWT performs in a better way while the complex Regular Discrete Wavelet Transform is recommended for higher levels, as its performance increases as the number of levels gets higher!!!!

We now expand the above procedure taking into consideration the fusion process. The structure of the filters remains the same in both transforms, the fusion rule as well: select the highpass coefficients of maximum amplitude and take the average of the created lowpass coefficients. **We have already seen, in the previous sections, that the Dual Tree Complex Wavelet Transform outperforms the corresponding Real Dual Tree Transform and that the Dual Tree structure outperforms the Real Discrete Wavelet Transform**. We expect that the regular Discrete wavelet transform using complex wavelets would give better results in compare with the Real one. The conclusions are summarized in the following pages:

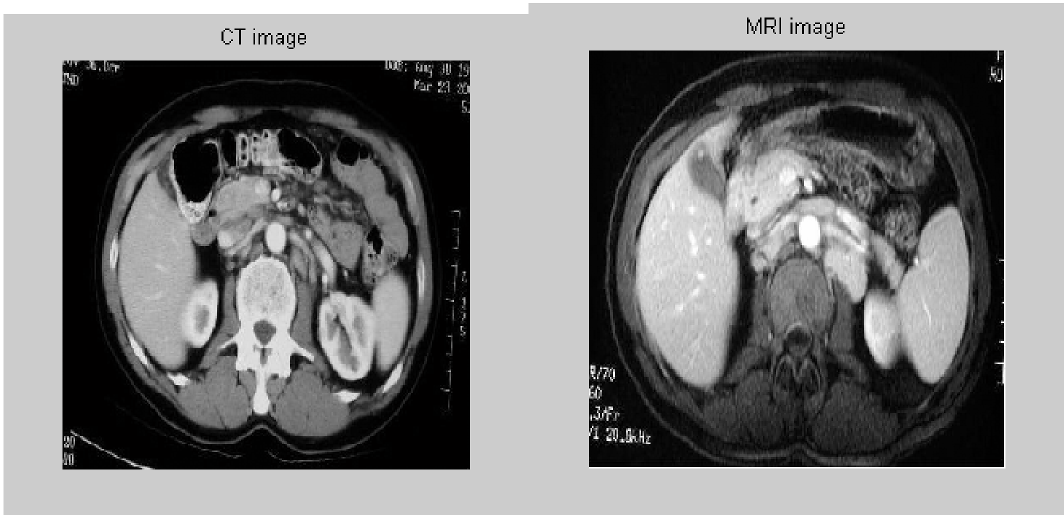


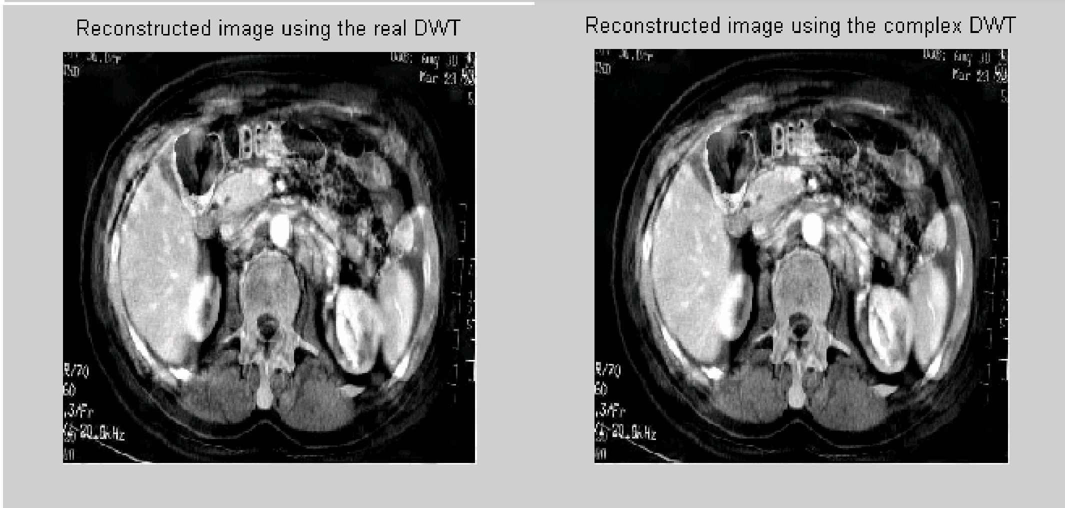


Qf=0.5279 Levels=6 Qf=0.5329

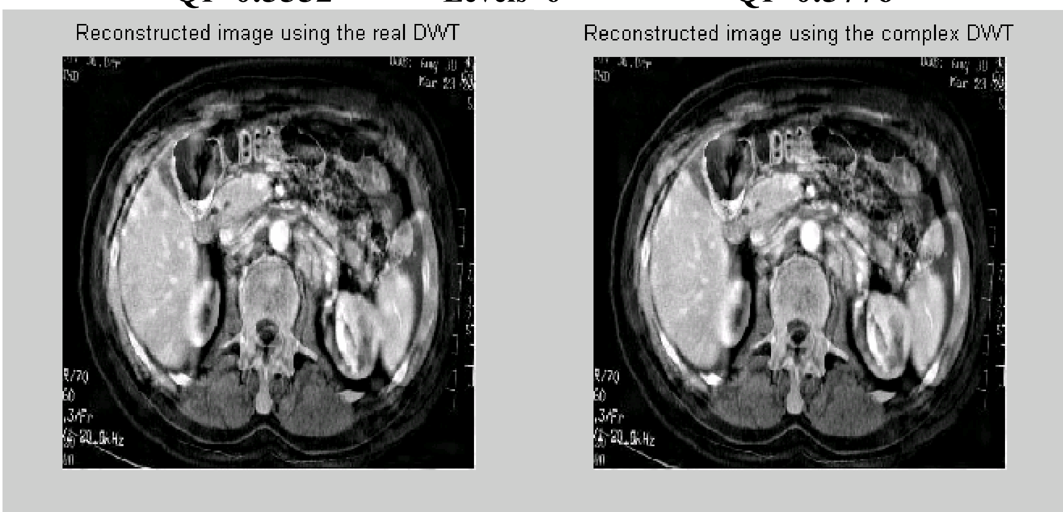


Qf=0.5784 Levels=4 Qf=0.5940

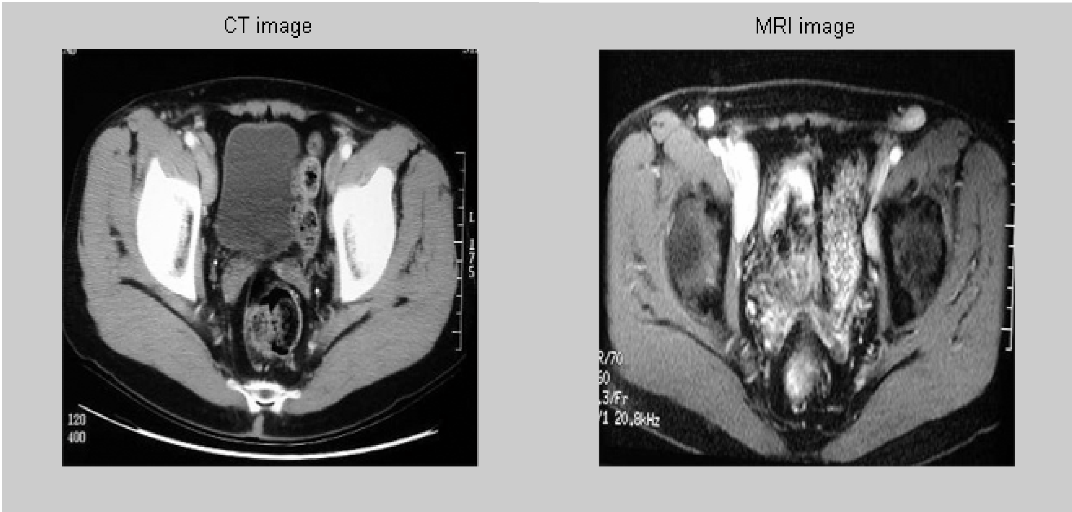


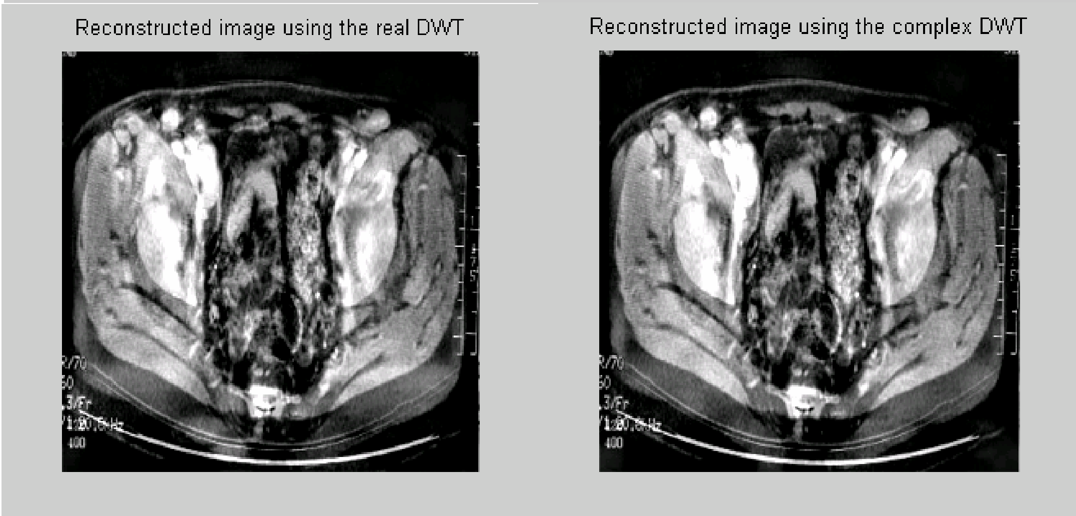


Qf=0.5552 **Levels=6** **Qf=0.5776**

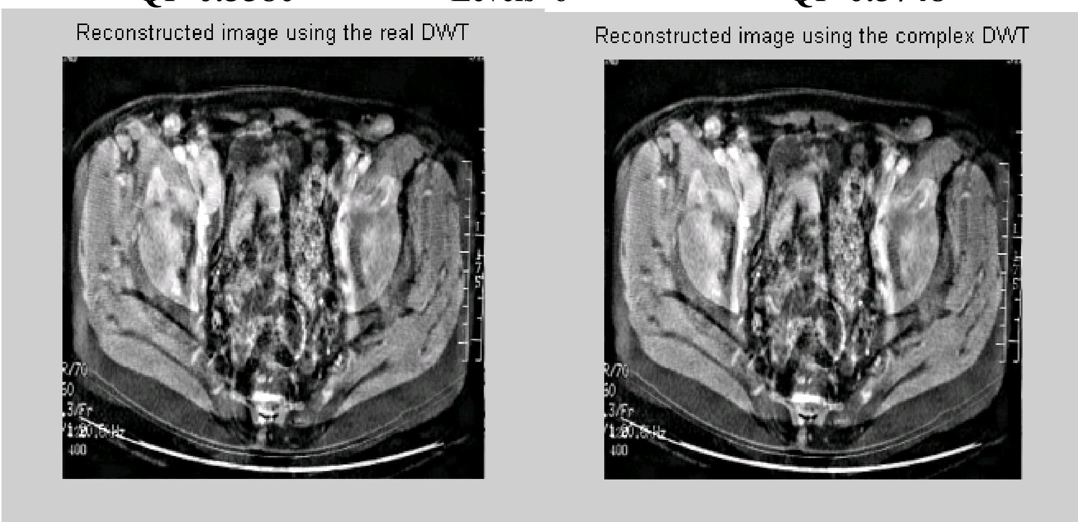


Qf=0.5817 **Levels=4** **Qf=0.6005**

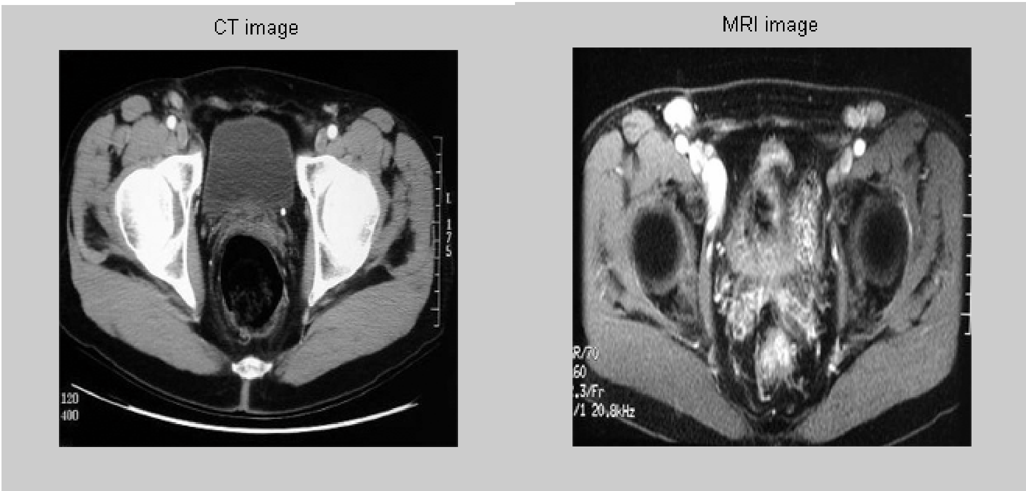


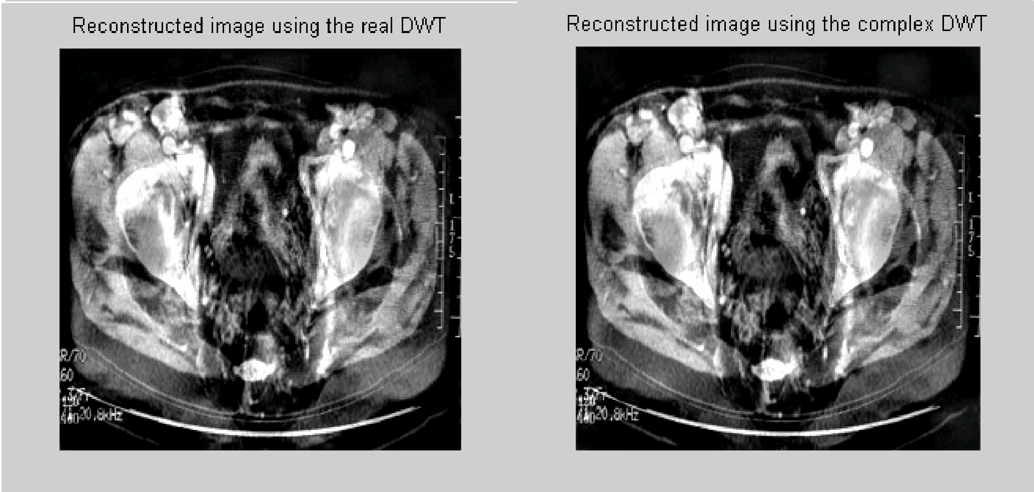


Qf=0.5580 **Levels=6** **Qf=0.5748**

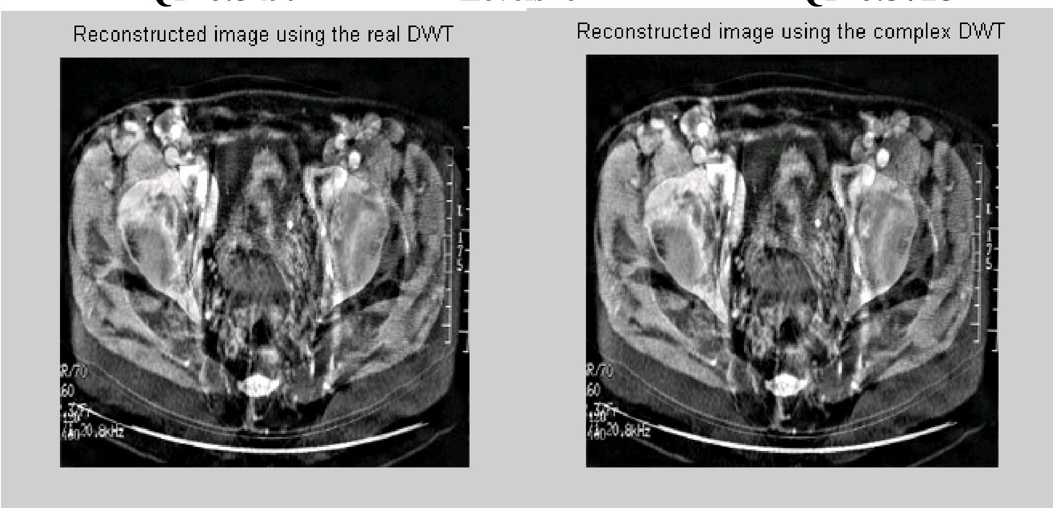


Qf=0.6279 **Levels=4** **Qf=0.6510**





Qf=0.5499 **Levels=6** **Qf=0.5725**



Qf=0.6224 **Levels=4** **Qf=0.6409**

The above are summarized in the following table:

	Image pair	Quality metric Qf Real DWT	Complex DWT
Number of levels = 4	1	0.5784	0.5940
	2	0.5817	0.6005
	3	0.6279	0.6510
	4	0.6224	0.6409
Number of levels =6	1	0.5279 ↓	0.5329
	2	0.5552 ↓	0.5776
	3	0.5580 ↓	0.5748
	4	0.5499 ↓	0.5725

In this section it was attempted to compare the complex and the real version of the Discrete Wavelet Transform. Two useful conclusions were extracted:

- The complex version of the wavelet transform, in the pseudo-simulation process implemented, clearly outperforms the corresponding real one, both in the regular Discrete Wavelet Transform and the Dual Tree Wavelet Transform.
- In the analysis-fusion-synthesis procedure, the fusion performance, for all the inputs used, deteriorates as the number of decomposition levels is getting higher (the metric Q_f is decreased), although the complex Discrete Wavelet Transform reconstructed the original image, without fusing the coefficients, in a better way in the upper levels.

		Real DWT		Complex DWT	
Levels	Image pair	Q	RMSE	Q	RMSE
4	1	0.8152	0.0073	0.8152	0.0073
	2	0.9104	0.0066	0.9104	0.0066
	3	0.9008	0.0075	0.9008	0.0075
	4	0.9211	0.0076	0.9211	0.0076
6	1	0.8431	0.0071	0.8708 ↑	0.0058 ↓
	2	0.9031	0.0062	0.9278 ↑	0.0051 ↓
	3	0.8913	0.0071	0.9144 ↑	0.0060 ↓
	4	0.9199	0.0069	0.9300 ↑	0.0060 ↓

Discarding the fusion process, the reconstructed image using the complex DWT was a closer approximation to the input image in the upper level 6 than in the lower level 4

Chapter 3
Non-perfect reconstruction wavelet transforms

In this chapter, we deal with wavelet-based algorithms and techniques that do not have the perfect reconstruction property, which means that the reconstructed image (signal in the general purpose) is not identical to the input one. The results extracted clearly showed that the fusion process cannot be efficiently implemented. However, these methods offer other basic advantages. In section 3.1 the Mallat-Zhong Discrete Wavelet transform (generally mentioned as Dyadic Discrete Wavelet Transform) is presented and its application to medical images is tested. A new concept is introduced in sections 3.1.2, 3.2: multiscale edge detection and its application to image fusion. Following this approach, input data can be represented and manipulated using their edge points, which are extracted through wavelet-based transforms. This representation brings data into a more abstract and subtractive description, which can give the implementer less computational load to manipulate and information about the parts where the input signal has sharp variations. Reconstructing the output image from edges was not perfectly achieved using our proposed wavelet-based techniques. A much more complicated and deeply theoretical algorithm is needed for the synthesis procedure. This technique was proposed by Mallat in [20] but is not implemented in our case. Instead, a much simpler technique is used with the cost of poor performance.

3.1. The Mallat-Zhong Discrete Wavelet Transform

3.1.1. The analysis-fusion-synthesis process

The Discrete Dyadic Wavelet Transform (this is the general notation, not absolutely correct, met in bibliography for the Mallat-Zhong Discrete Wavelet Transform) is another simplified "version" of the Discrete Wavelet Transform, basically used for edge detection along different scales/levels. It has a similar structure as the DWT described in the sections above. It can be implemented using two approaches:
a) the theoretical/analytical one, using only mathematical calculations by convolving the input signal (images in our case) with a wavelet scaling function
b) the practical one, applying low and high pass filters along each dimension through a certain decomposition/reconstruction scheme.

ANALYTICAL IMPLEMENTATION:

The two dimensional Discrete Dyadic Wavelet Transform of an image I(x, y) at scale 2^j and in orientation k is defined as

$$W_{2^j}^{k} I(x,y) = I(x,y) * \psi_{2^j}^{\kappa}(x,y)$$

with k=1, 2(horizontal and vertical dimension) and j=1, …, N (=maximum number of levels)

The orientated wavelets $\psi_{2^j}^{\kappa}(x,y)$ can be constructed by taking the partial derivatives of the scaling function

$$\psi^1(x,y) = \frac{\partial\theta(x,y)}{\partial x} \quad \text{and} \quad \psi^2(x,y) = \frac{\partial\theta(x,y)}{\partial y}$$

where θ(x, y) is a separable spline scaling function which plays the role of a smoothing filter. It can be shown that the DDWT gives the gradient of the image I(x, y) smoothed by θ(x, y) at dyadic scales:

$$\nabla_{2^j} I(x,y) \equiv \left(W^1_{2^j} I(x,y), W^2_{2^j} I(x,y)\right) = \frac{1}{2^{2^j}} \nabla\left(I * \theta_{2^j}\right)(x,y) = \frac{1}{2^{2^j}} \nabla\left(\theta_{2^j} * I\right)(x,y)$$

The smoothing function θ(x, y) can be defined as the product of two one dimensional smoothing functions, e.g. the prototypical example, the Gaussian

$$\theta(x,y) = \theta_1(x)\theta_2(y)$$

with the Gaussian functions $\theta_1(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$ **,** $\theta_2(y) = \frac{1}{\sqrt{\pi}} e^{-y^2}$

We can consider that at each scale $s=2^j$ the two orientated wavelets are given by the relation:

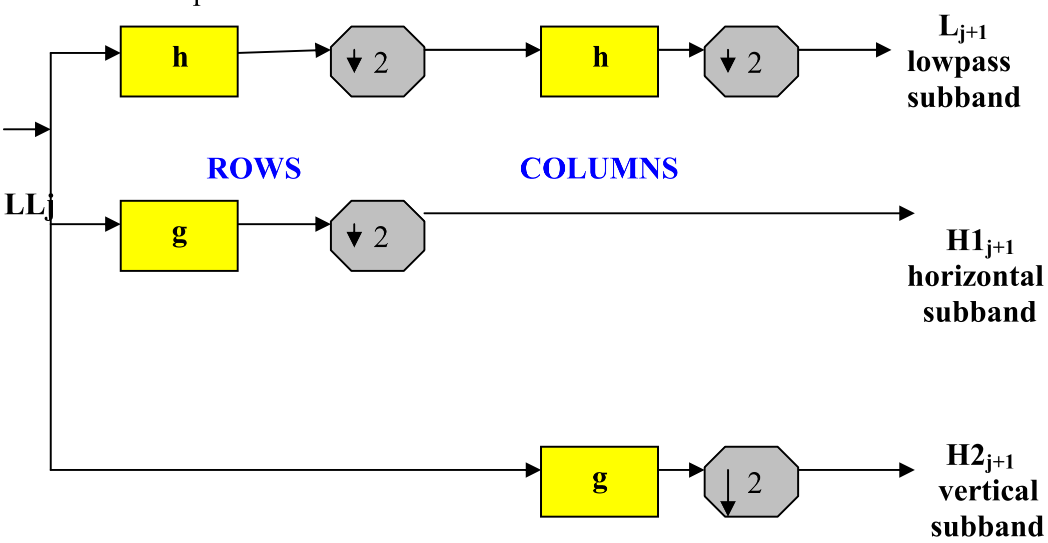
$$\psi^1_s(x,y) = \frac{1}{s^2}\psi^1\left(\frac{x}{s},\frac{y}{s}\right) \quad \textbf{and} \quad \psi^2_s(x,y) = \frac{1}{s^2}\psi^2\left(\frac{x}{s},\frac{y}{s}\right)$$

However, the lack of shift invariance and the aliasing associated with the use of non-redundant wavelet transforms, e.g. DWT, may introduce undesirable artifacts in the reconstructed images. Thus the use of the redundant wavelet representations such as the DDWTs is often justified [9] [18].

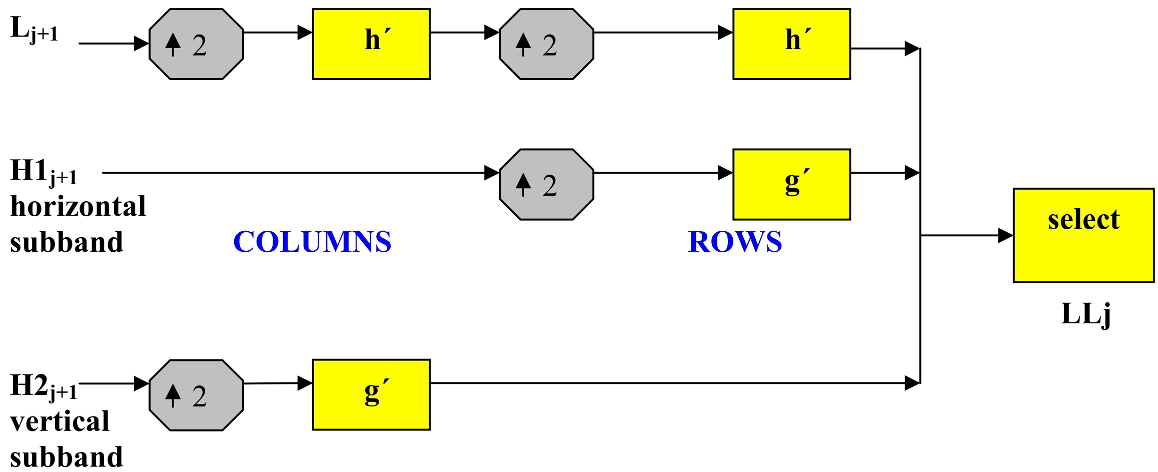
The synthesis procedure is complicated and deeply based on mathematical terms and calculations and will not be further analyzed. The reconstruction algorithm can be further approached at [20], [21]. Instead, the "practical" implementation, using cascaded application of filters is most frequently adopted.

PRACTICAL IMPLEMENTATION:

Following this approach, things are much more simplified. The basic decomposition/reconstruction scheme is similar to the one introduced in the previous sections and is presented below:



Decomposition procedure at level j



Reconstruction procedure at level j

↑↓2 → up/downsampling of the corresponding subimage (**optional in most cases**)

j=current decomposition level, **j+1**=next decomposition level

h → lowpass analysis/decomposition filter, convolution with the input image

g → highpass analysis/decomposition filter, convolution with the input image

select → select the valid part in order to obtain the appropriate image size due to convolution increment of subband size

In practice, the dyadic wavelet transform is computed by iterative filtering with a set of low and high pass filters h and g, associated with the wavelets ψ^1 and ψ^2 of the theoretical approach described before. These filters have finite impulse response, which makes the transform fast and easy to implement:

$$L_{2^{j+1}}(x, y)=[H_{j,x}*[H_{j,y}* L_{2^j}]](x, y)$$

$$D^1_{2^{j+1}}(x, y)=[D_{j,x}*[G_{j,y}* L_{2^j}]](x, y)= H1_{j+1}$$

$$D^2_{2^{j+1}}(x, y)=[G_{j,x}*[D_{j,y}* L_{2^j}]](x, y)= H2_{j+1}$$

where L_1 is the input image, L_{2^j} is the lowpass subimage at scale j, D is the Dirac filter whose impulse response is equal to 1 at 0 and 0 otherwise and $^{i}\mathbf{D}_{2}{}^{j}$ i=1,2, j=1, …, N are the detail, or highpass, subbands at scale j [18]. The orientation 1 refers to the rows and the orientation 2 refers to the columns.

In general, the up/downsampling procedure is discarded in the decomposition-reconstruction process of the previous scheme. Ivan Christov in [19] proposed the Mallat-Zhong Discrete Wavelet Transform which implements the dyadic transform using a quadratic spline lowpass filter h(n)={…, 0, 1, [3], 3, 1, 0, …}*√2/4 in order to generate the scaling function and the wavelets associated with it and speed-up computation and prevent group delays that would be introduced by non-symmetric filters. The decomposition filters at each resolution 2^j are computed by upsampling the filter $h_j[n]$ by j and the filters $g_j[n]$ are normalized finite difference filters meant to approximate the gradient of the image. The reconstruction filters h′, g′ are simply the decomposition filters flipped [19].

$$h_j[n]=(\uparrow j)\, h[-n]$$

$$g_j[-/+\, 2^{j-1}]=+/-\frac{\sqrt{2}}{2} \;,\; g_0[-1]=-\frac{\sqrt{2}}{2}$$

$$h'_j[n]=h_j[-n] \;,\; g'_j[n]=g[-n]$$

The dyadic transform is performed via a cascade of separable convolutions using the above filters:

$$A_{j+1}\ (x, y)=A_j * h_j[x] * h[y]$$

$$D^1_{j+1}(x,y) = A_j * g_j[x] \text{ (at rows)}$$

$$D^2_{j+1}(x,y)=A_j * g_j[y] \text{ (at columns)}$$

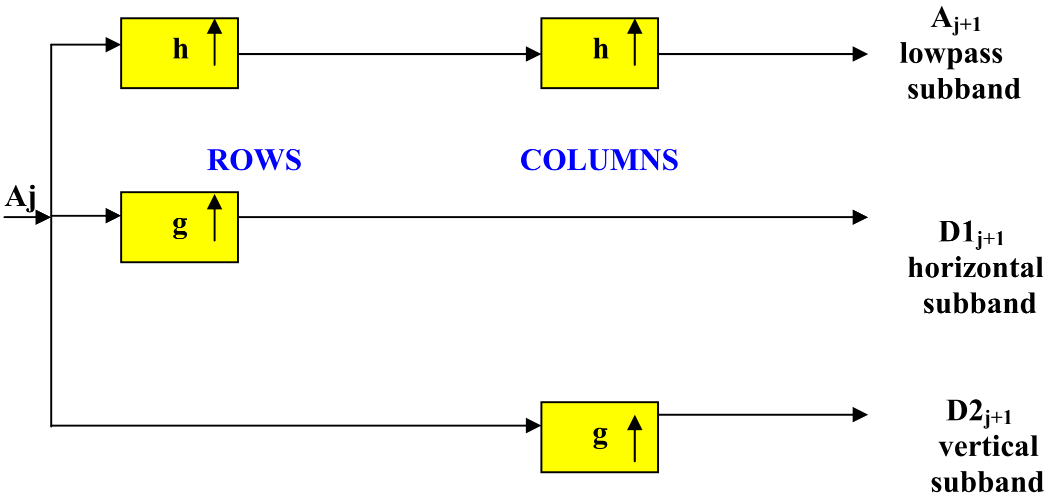
The coefficients A_{j+1} are the lowpass ones, while coefficients $D^1_{j+1}(x,y)$ and $D^2_{j+1}(x,y)$ are the highpass ones at horizontal and vertical orientation respectively.

The inverse transform is computed in a similar way

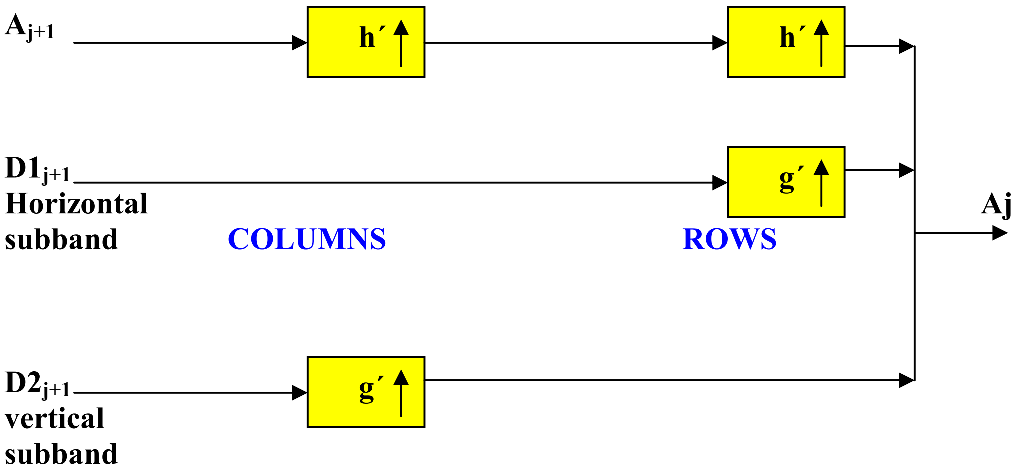
$$A_j(x, y)=A_{j+1} * h_j[x] * h_j[y] + D^1_{j+1} * g_j[x] + D^2_{j+1} * g_j[y]$$

****It must be remarked that the 2-D dyadic discrete wavelet transform does not have perfect reconstruction because the diagonal detail coefficients are not calculated. However, for edge detection it is sufficient to compute the horizontal and vertical wavelet coefficients.**

The Mallat-Zhong Discrete Wavelet Transform



Decomposition procedure at level j



Reconstruction procedure at level j

j=current decomposition level, **j+1**=next decomposition level

h, h′ → lowpass analysis/synthesis filter, upsampled, convolution with the input image

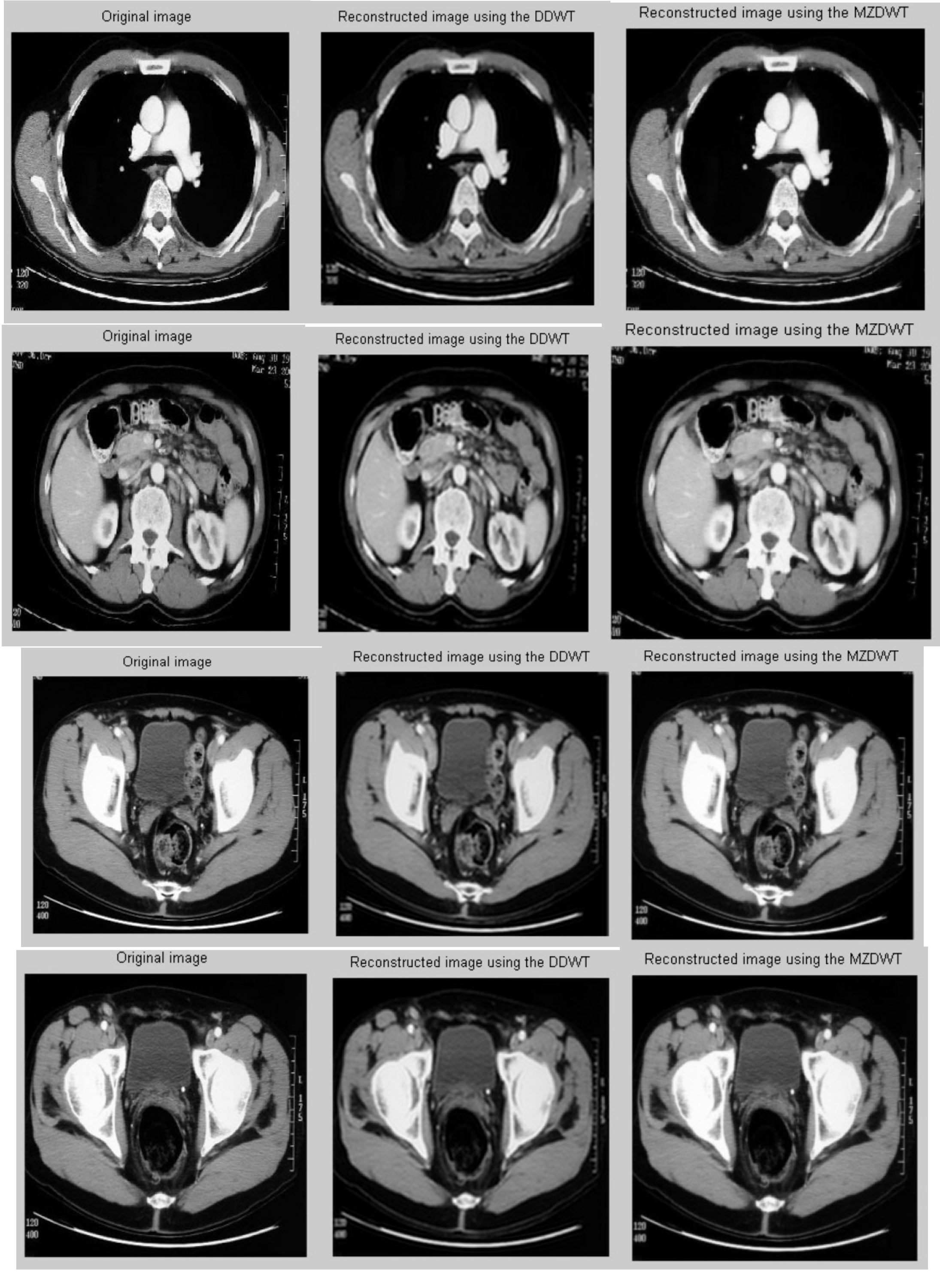
g, g′ → highpass analysis/synthesis filter, zero-padded, convolution with the input image

A list of tests is presented in the following pages in order to make a clearer view of the Dyadic Discrete Wavelet Transform. The two "practical implementations" are adopted with the difference that in the first one the up/downsampling procedure is discarded. Both the decomposition/reconstruction and the analysis-fusion-synthesis stages are implemented in order to compare the two versions. We expect that the second version, known as the Mallat-Zwong Wavelet Transform [19], which uses upsampling in the different stage filters, is proved more efficient. The fusion rules to be followed are fixed:

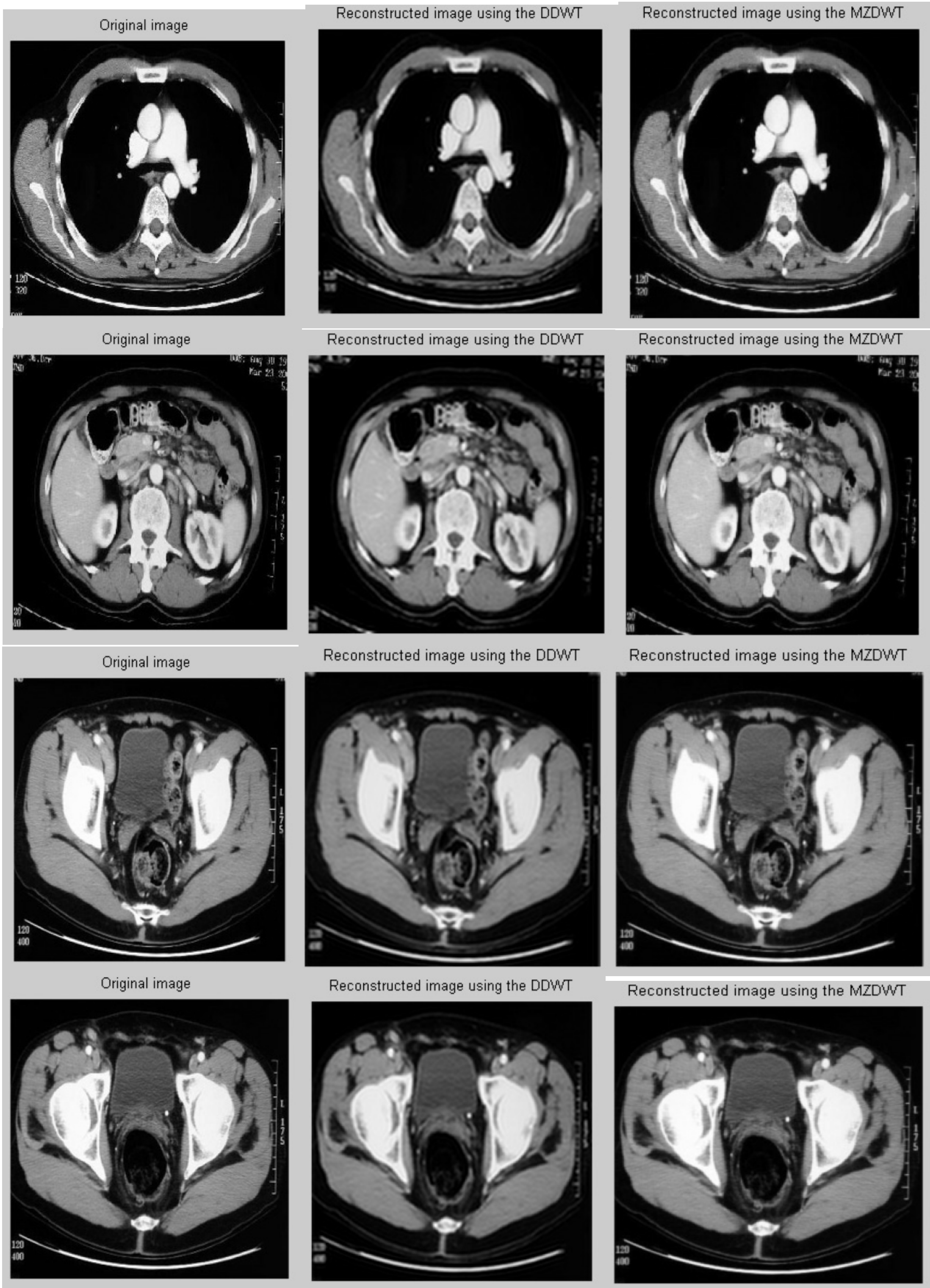
a) take the average of the lowpass coefficients and b) select the highpass coefficients of the maximum magnitude.

Test 1: evaluate the decomposition/reconstruction process

Number of levels = 4



Number of levels = 6



The above are summarized in the following table:

		Dyadic version1 WT		Mallat-ZwongWT	
Levels	Image pair	Q	RMSE	Q	RMSE
4	1	0.5981	0.0675	0.8430	0.0295
	2	0.7904	0.0443	0.9123	0.0430
	3	0.7098	0.0582	0.8962	0.0207
	4	0.7369	0.0422	0.9132	0.0165
6	1	0.5862	0.0691	0.8367	0.0295
	2	0.7815	0.0449	0.9099	0.0429
	3	0.6951	0.0598	0.8946	0.0207
	4	0.7203	0.0445	0.9111	0.0165

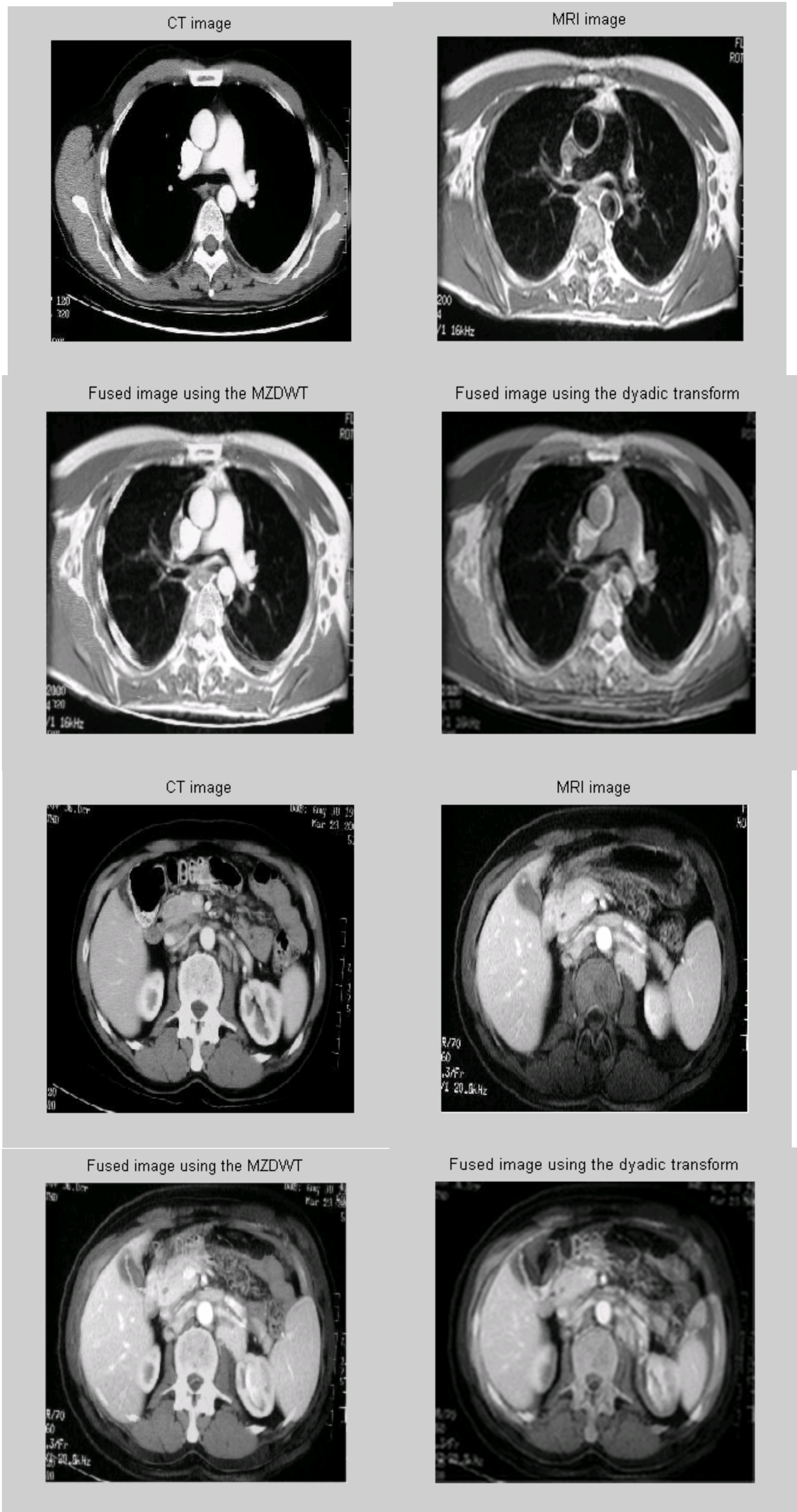
Conclusion: From the above, we notice that the expected theoretical results are confirmed in practice too. The performance deteriorates very slightly as the number of levels is increased and the MZDWT, which uses upsampled filters, is a much better version than the simple dyadic transform discarding the sampling process.

We implement the previous scheme for the needs of image fusion. Having decomposed the input images via the forward wavelet transform (the Dyadic Discrete wavelet transform), we fuse the corresponding coefficients of each subband produced using the Maximum Selection rule and finally we perform the inverse transform and reconstruct the fused image. We noticed that the wavelet transform produced coefficients of large amplitude and due to the non-perfect reconstruction imperfection of the transform the fused image attained large values too, outside the range [0 256] where an 8-bit image is normally valued. So, rescaling of the values needed to take place.

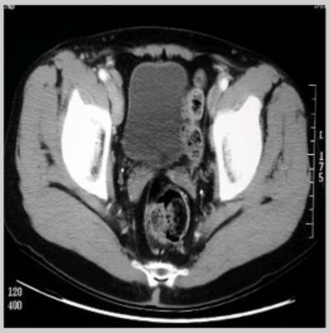
The qualitative (images) and quantitative (metric Qf) results of the fusion process are presented in the following pages.

Test 2: evaluate the analysis-fusion-synthesis process

Number of levels=4



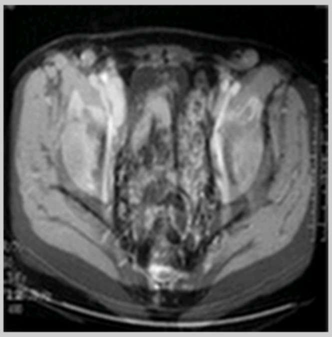
CT image



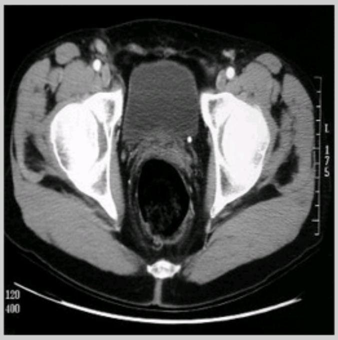
MRI image

Fused image using the MZDWT

Fused image using the dyadic transform

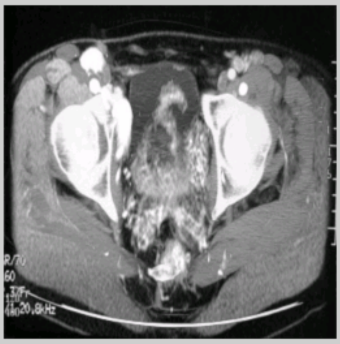


CT image

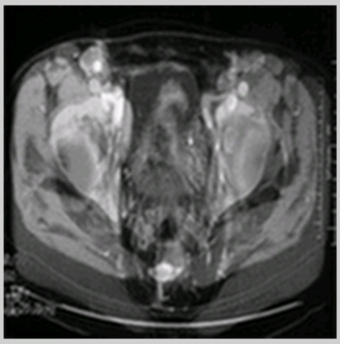


MRI image

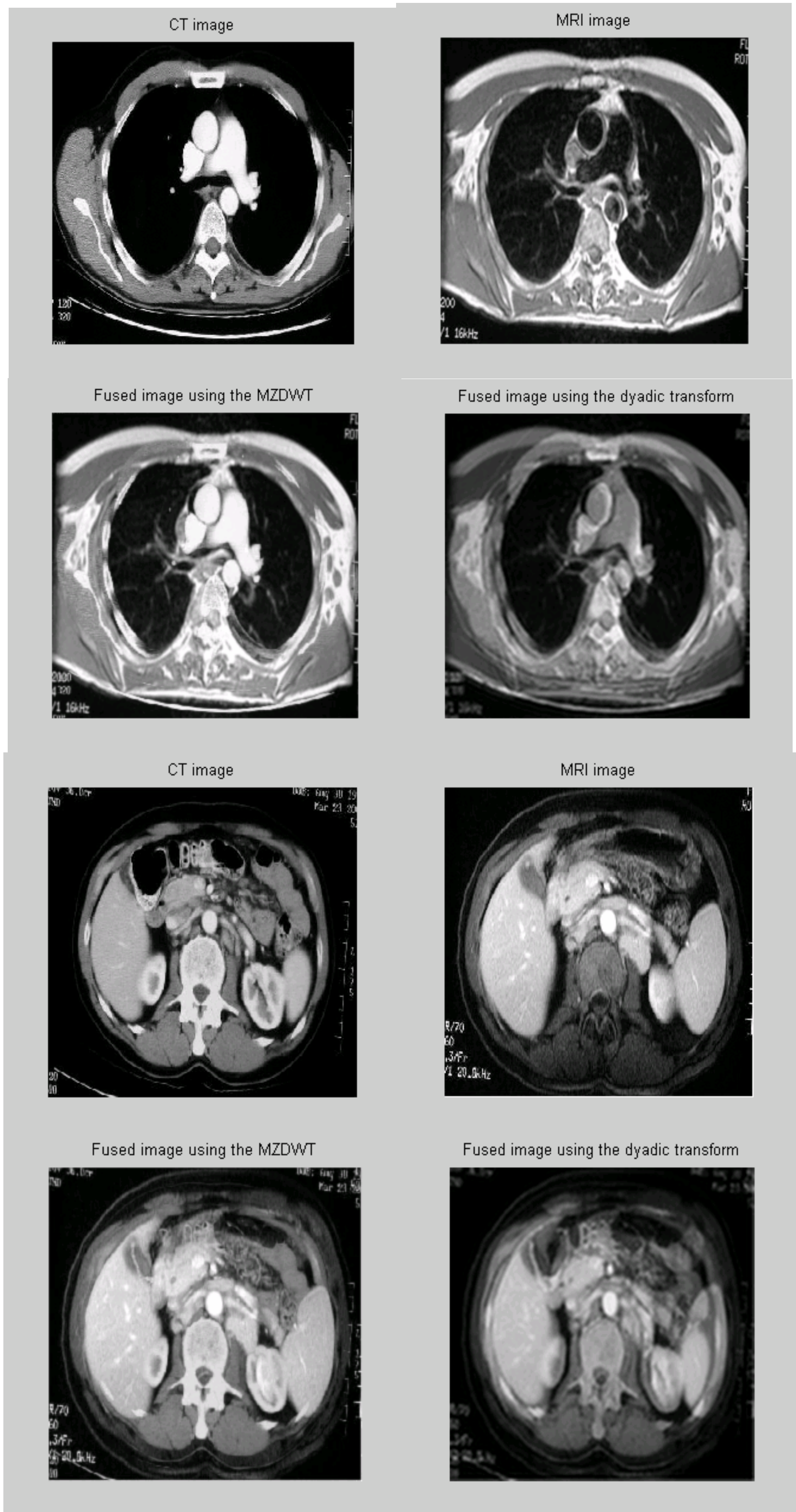
Fused image using the MZDWT

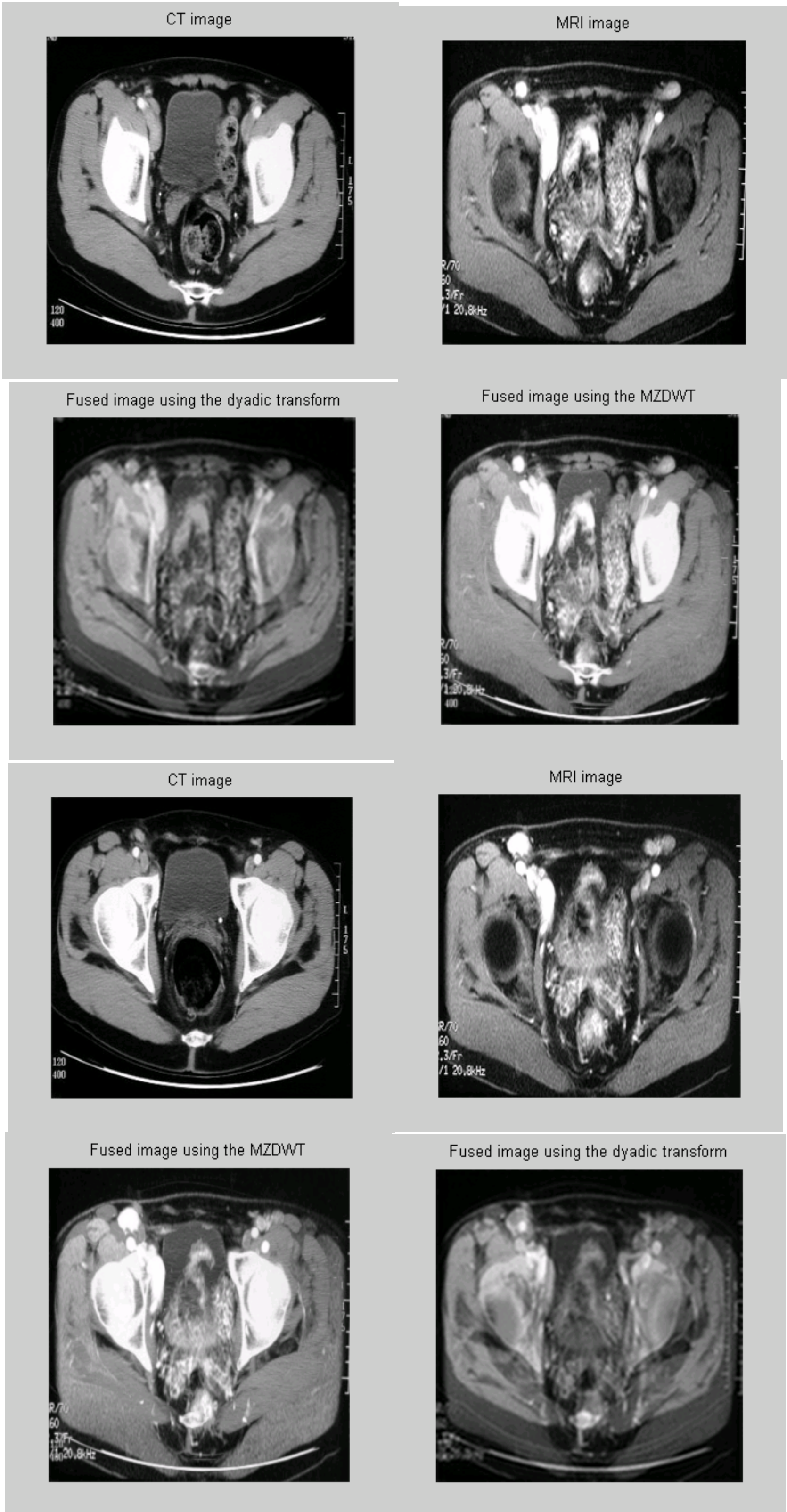


Fused image using the dyadic transform



Number of levels=6

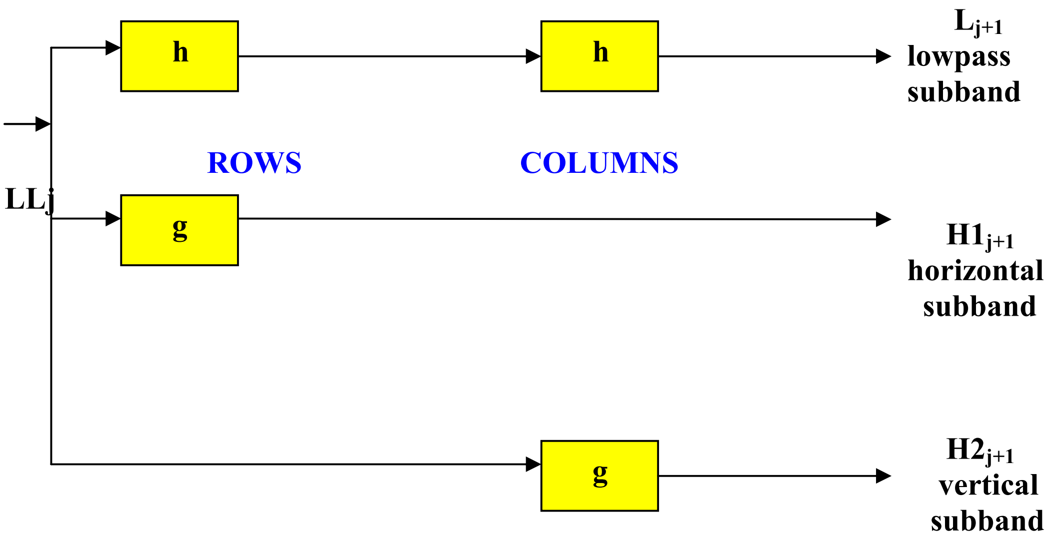



CT image
MRI image
Fused image using the dyadic transform
Fused image using the MZDWT
CT image
MRI image
Fused image using the MZDWT
Fused image using the dyadic transform

The evaluation results are summarized in the following table:

	Quality metric Qf		
	Image pair	**Dyadic WT version1**	**Mallat-Zwong WT**
Number of levels = 4	1	0.4805	0.8138
	2	0.4426	0.7400
	3	0.4595	0.7371
	4	0.4350	0.7387
Number of levels =6	1	0.4727	0.8138
	2	0.4361	0.7401
	3	0.4520	0.7372
	4	0. 4275	0.7384

Conclusion: From the analysis above we conclude that the Mallat-Zhong wavelet transform performs in a much better way than the simple version of the Dyadic Discrete Wavelet Transform. This is due to upsampling. The “simple” version of the Dyadic Transform tends to blur the input image and the lowpass subband seems to overwhelm the two highpass subbands in the final contribution of the coefficients because lowpass filtering is used in both dimensions (rows and columns) but highpass filtering is applied in rows, once, and the columns, next.



Another important conclusion is that the MZ-DWT is very stable when different number of stages are used, performs extremely equally in either case! Further tests, not presented here proved so**: the performance of the MZ-DWT in the fusion process is extremely stable along the different number of decomposition levels**. In the next sections, this version of the Dyadic Wavelet Transform will be adopted.

3.1.2 Multiscale edge detection via modulus maxima calculation

We can expand the Dyadic Discrete Wavelet Transform for the needs of multiscale edge detection. Again, edge detection can be implemented using two approaches: a) the analytical, mathematical one, which uses mathematical functions for wavelet representation and b) the practical one, which "simulates" the mathematical calculations via application of filters associated with wavelets. In both implementations, the mathematical basis comes from the Dyadic Wavelet Transform.

ANALYTICAL IMPLEMENTATION:

As already presented, the two dimensional Discrete Dyadic Wavelet Transform of an image I(x, y) at scale 2^j and in orientation k is defined as

$$W_{2^j}^{k} I(x,y) = I(x,y) * \psi_{2^j}^{\kappa}(x,y)$$

with k=1, 2(horizontal and vertical dimension) and j=1, …, N (=maximum number of levels)

The orientated wavelets $\psi_{2^j}^{\kappa}(x,y)$ can be constructed by taking the partial derivatives of the scaling function

$$\psi^1(x,y) = \frac{\partial\theta(x,y)}{\partial x} \quad \text{and} \quad \psi^2(x,y) = \frac{\partial\theta(x,y)}{\partial y}$$

where θ(x, y) is a separable spline scaling function which plays the role of a smoothing filter.

The smoothing function θ(x, y) can be defined as the product of two one dimensional smoothing functions, e.g. the prototypical example, the Gaussian

$$\theta(x,y) = \theta_1(x)\theta_2(y)$$

with the Gaussian functions $\theta_1(x) = \frac{1}{\sqrt{\pi}} e^{-x^2}$ **,** $\theta_2(y) = \frac{1}{\sqrt{\pi}} e^{-y^2}$

We can consider that at each scale $s=2^j$ the two orientated wavelets are given by the relation:

$$\psi_s^1(x,y) = \frac{1}{s^2}\psi^1\left(\frac{x}{s},\frac{y}{s}\right) \quad \text{and} \quad \psi_s^2(x,y) = \frac{1}{s^2}\psi^2\left(\frac{x}{s},\frac{y}{s}\right)$$

At each orientation 1, 2 we consider the two wavelet transforms

$$W_{2^j}^1 I(x,y) = I(x,y) * \psi_{2^j}^1(x,y) \quad \text{and} \quad W_{2^j}^2 I(x,y) = I(x,y) * \psi_{2^j}^2(x,y)$$

If we want to locate the positions of rapid variation of an image I, the "edges", we should consider the local maxima of the gradient magnitude at various scales [9]. This is given by the following equation:

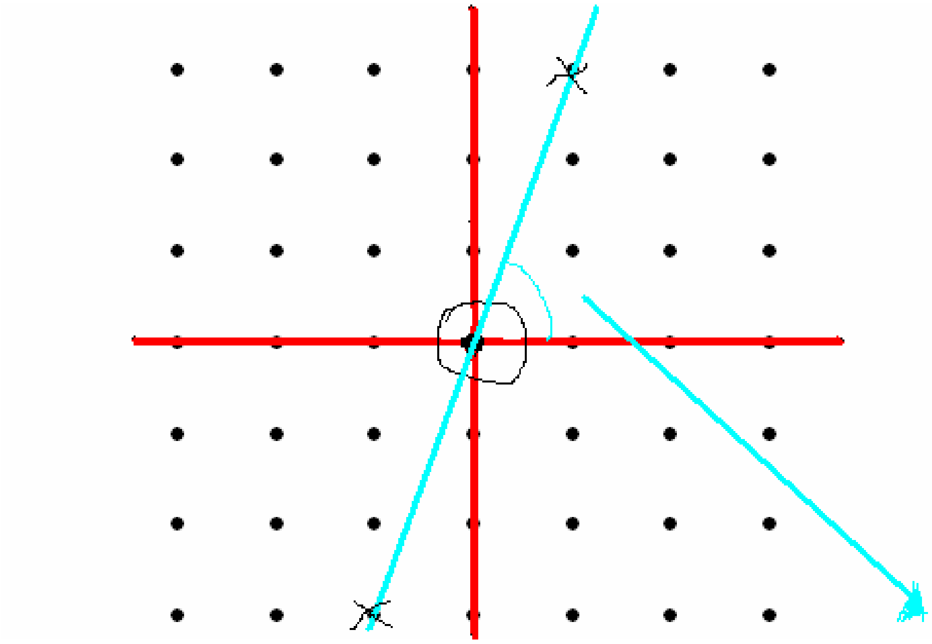
$$M_{2^j} I(x,y) \equiv \left\|\nabla_{2^j} I(x,y)\right\| = \sqrt{(W_{2^j}^1 I(x,y))^2 + (W_{2^j}^2 I(x,y))^2}$$

The gradient term $M_{2^j} I(x,y)$ is also known as the **Modulus** of the image I at scale j.

A point (x, y) of the image is a multiscale edge point at scale j if the magnitude of the modulus attains a local maximum along the gradient direction $A_{2^j} I(x,y)$, also known as the **Angle** of the image I, defined as

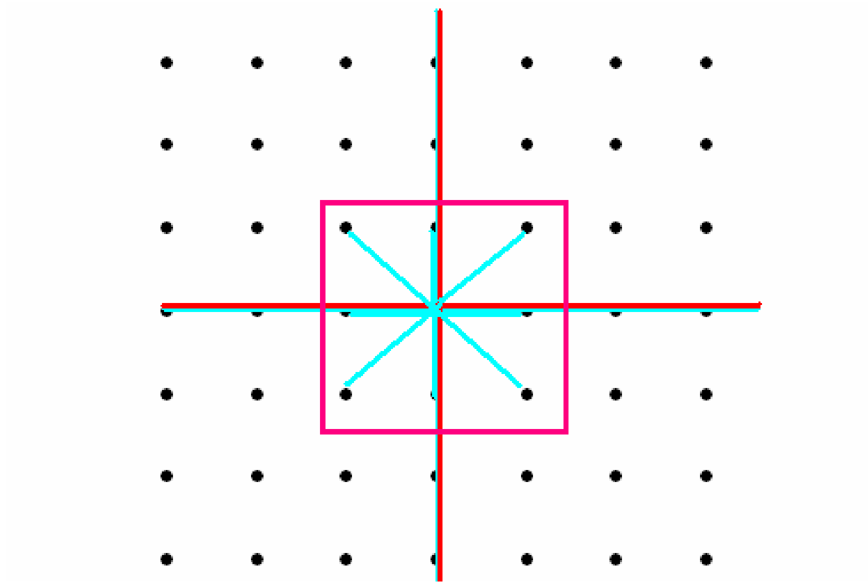
$$A_{2^j} I(x,y) \equiv \arctan\left[\frac{W_{2^j}^2 I(x,y)}{W_{2^j}^1 I(x,y)}\right]$$

This means that the point(x, y) must have the largest magnitude along its “neighbors” located at the same direction the Angle(x, y) defines, as illustrated in the scheme below:



angle

On the grounds that the value of the angle is not fragmented, there is a great number of possible “neighbors”, which means computational complexity and delay in order to find and compare the neighbored pixels. To avoid this inconvenience, we quantize the angle in the values [0° 45° 90° 135° 180°], so we have only a 8-elements neighborhood to check:



If a point (x, y) is found to be a local maxima (commonly named as modulus maxima) then it is assigned the value 1, otherwise it is assigned the value 0. As a result, at each scale, a new set (matrix) of values is obtained containing the edges of the image.

The set $\rho(I)=\left\{S_{2^j}I(x,y),[P_{2^j}(I)]_{1\le j\le J}\right\}$ where $S_{2^j}I(x,y)$ is the lowpass approximation of the image at scale j, $P_{2^j}(I)=\left\{p_{2^j,i}=(x_i,y_i);M_{2^j}I(x_i,y_i)\right\}$ is the set of points where the Modulus $M_{2^j}I(x_i,y_i)$ has local maximum at $p_{2^j,i}=(x_i,y_i)$ is called a ***multiscale edge representation of the image I(x, y)***, which a shift invariant form.

Two key points need to be clarified:

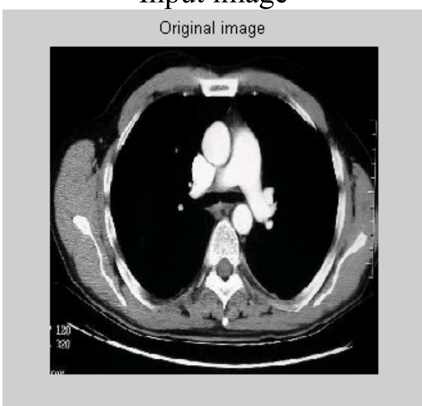
a) how the edge detection algorithm via modulus maxima calculation performs along the different number of levels/stages? On the grounds that this edge detection algorithm is mathematical and analytical it is proved much more computational expensive. Increasing the number of levels normally means that less edges will be detected and mainly, the delay in the extraction of the results will be increased.

b) what should be the size of the scaling function θ(x,y)? For the same reason as in the previous question, a scaling function of big size would delay the algorithm in a considerable measure and probably insert false edges or miss some edges.

The first results proved that the modulus maxima detection algorithm is very efficient but slow. The delay in the extraction of edges gets very slow as the size of the scaling function increases(over 32 elements) and mainly as the size of the input image is large(for example the classic "Lenna" 512x512 input image in combination with a 64 –point scaling function needed several minutes to extract the result!). So it is wise to carefully select and consider the parameters of the algorithm. There is no certain metric to evaluate the result of edge detection but our only "weapons", the observer's eye and the clock, immediately help us to make conclusions.

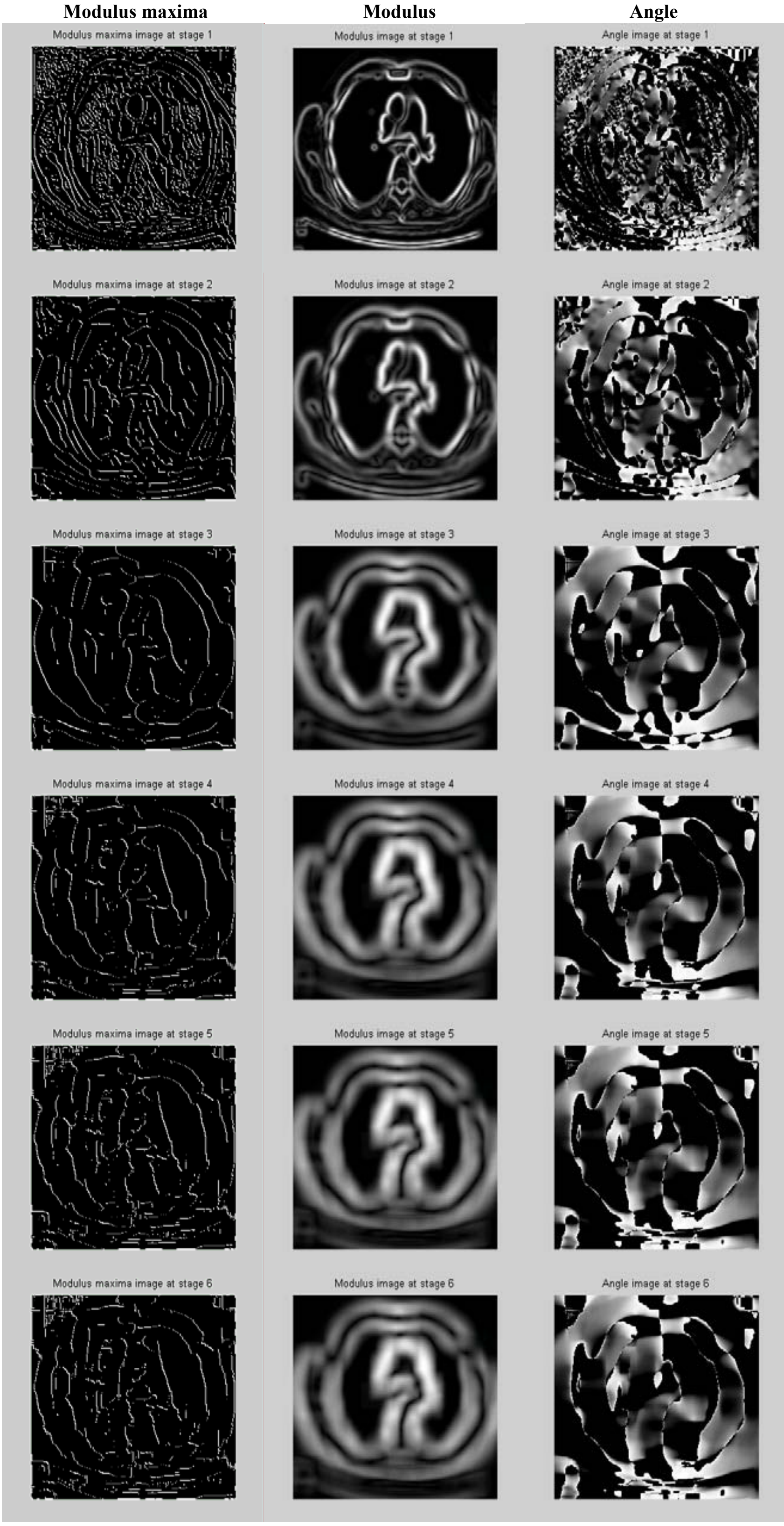
Examples of the use of the above process for the needs of edge detection are presented below:

Input image



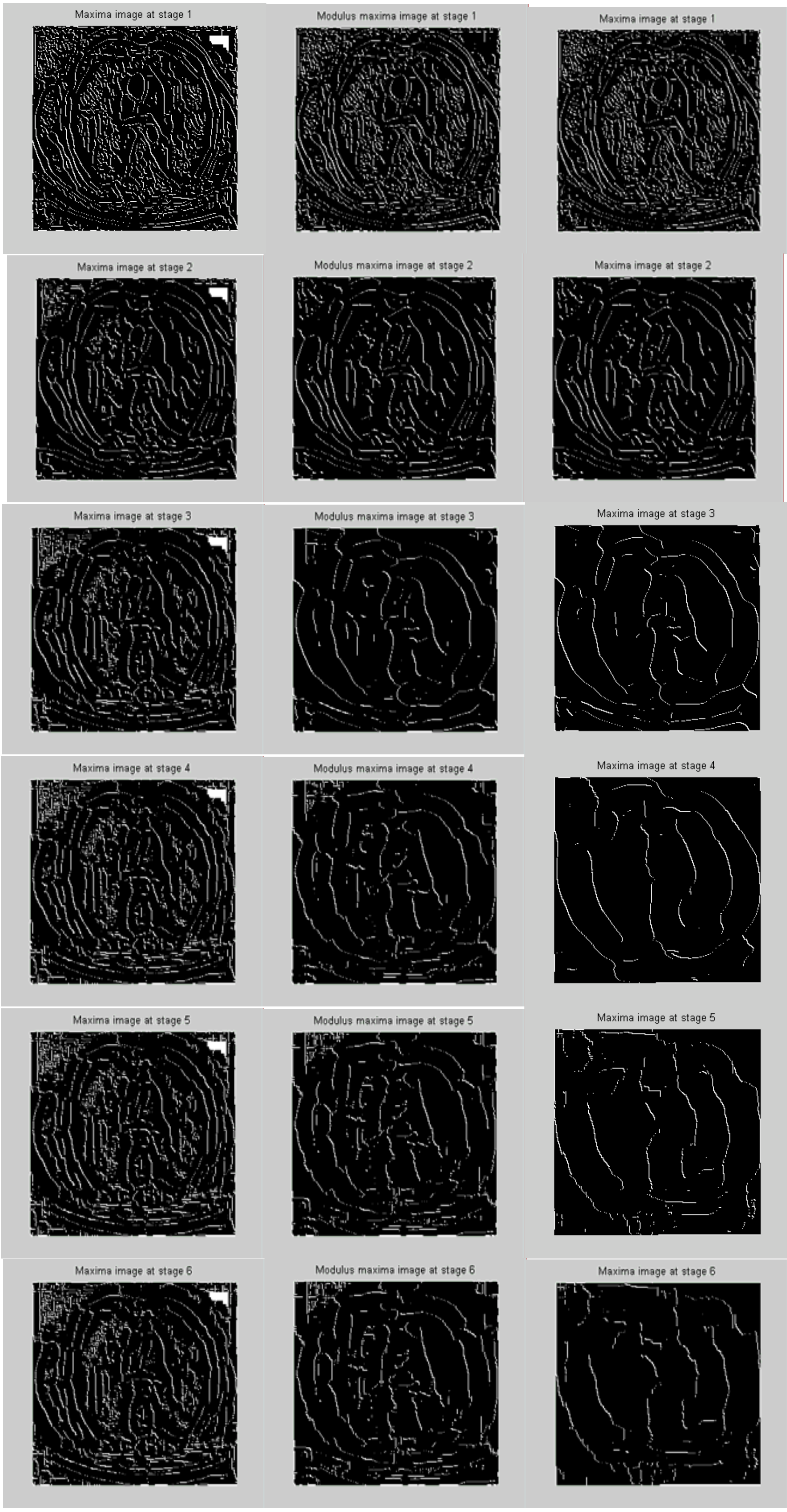
TEST 1: How the modulus maxima, modulus and the angle of an image look like?

Levels=1-6 **Modulus maxima** **Modulus** **Angle**



TEST 2: Modulus maxima using scaling function of different size

Levels=1-6 **size=16** **size=32** **size=64**



Conclusion: Observing the above, we notice that increasing the number of levels, as it was expected, the edge detection algorithm deteriorates; fewer edges are detected/preserved. As regards the size of the scaling function, we notice that a small size detects many information in some parts of the image and looses information in other (notice the right top hand corner) when the size is 16. In addition, bigger size of the scaling function means, firstly, computational delay and secondly less edges in the upper levels. For the needs of our work a Gaussian scaling function of size 32 is very efficient.

PRACTICAL IMPLEMENTATION:

Instead of using mathematical terms, we can follow the practical version presented in the previous section, applying low and high pass filters h and g to the input image, associated with the wavelets ψ^1 and ψ^2 of the theoretical approach. These filters have finite impulse response, which makes the transform fast and easy to implement:

$$L_{2^{j+1}}(x, y)=[H_{j,x} * [H_{j,y} * L_{2^j}]](x, y)$$

$$D^1_{2^{j+1}}(x, y)=[D_{j,x} * [G_{j,y} * L_{2^j}]](x, y)= H1_{j+1}$$

$$D^2_{2^{j+1}}(x, y)=[G_{j,x} * [D_{j,y} * L_{2^j}]](x, y)= H2_{j+1}$$

where L_1 is the input image, L_{2^j} is the lowpass subimage at scale j, D is the Dirac filter whose impulse response is equal to 1 at 0 and 0 otherwise and $^i\mathbf{D}_{2^j}$ i=1,2, j=1, …, N are the detail, or highpass, subbands at scale j. The orientation 1 refers to the rows and the orientation 2 refers to the columns.

If we want to locate the positions of rapid variation of an image I, the "edges", we should consider the local maxima of the gradient magnitude at various scales. This is given by the following equation:

$$M_{2^j} I(x,y) = \sqrt{(D^1_{2^j} I(x,y))^2 + (D^2_{2^j} I(x,y))^2}$$

The gradient term $M_{2^j} I(x,y)$ is also known as the **Modulus** of the image I at scale j.

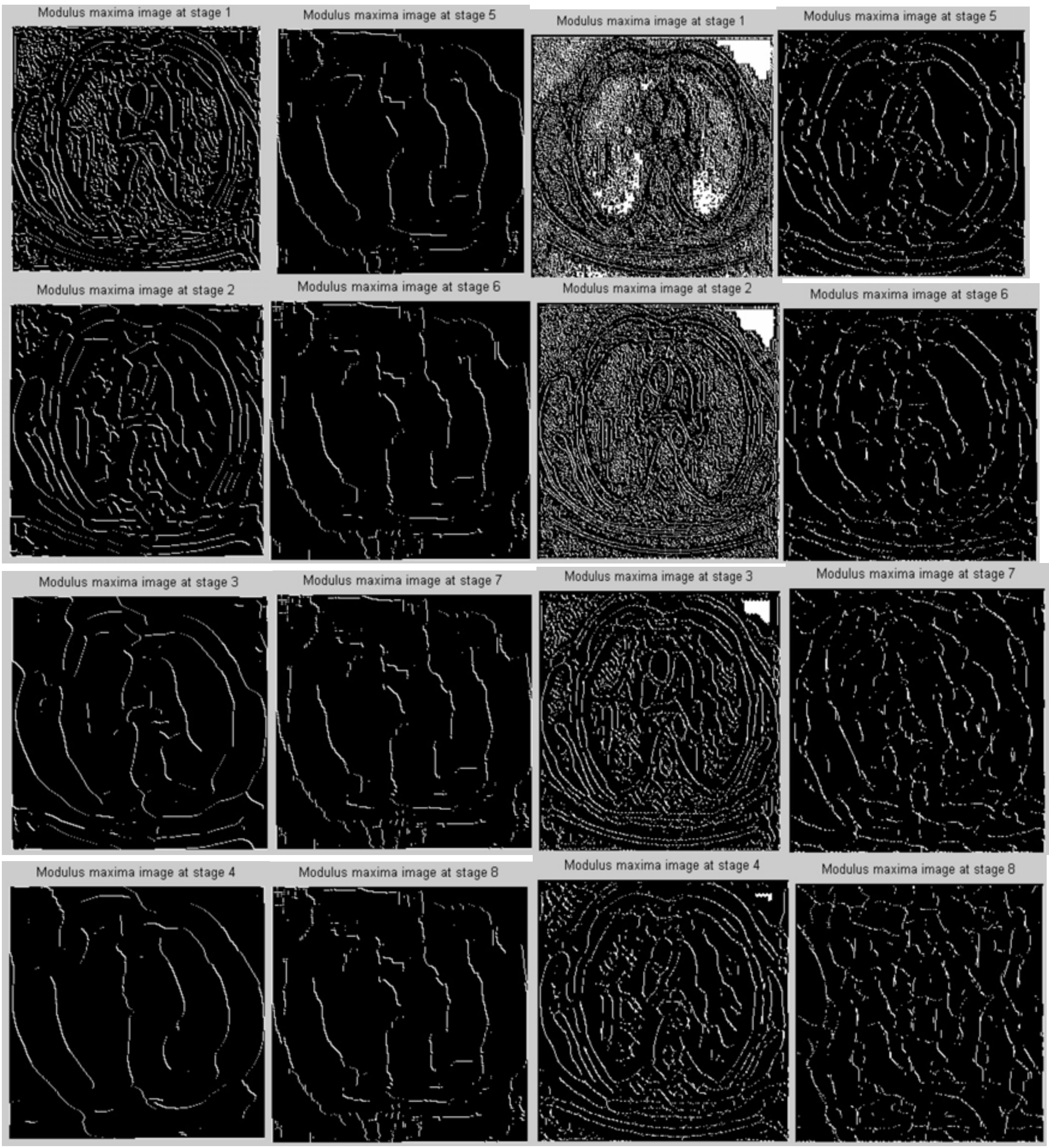
A point (x, y) of the image is a multiscale edge point at scale j if the magnitude of the modulus attains a local maximum along the gradient direction $A_{2^j} I(x,y)$, also known as the **Angle** of the image I, defined as

$$A_{2^j} I(x,y) \equiv \arctan\left[\frac{D^2_{2^j} I(x,y)}{D^1_{2^j} I(x,y)}\right]$$

This means that the point(x, y) must have the largest magnitude along its two "neighbors" located at the same direction the Angle(x, y) defines. Again, for simplicity, we quantize the angle in the interval $[0^\circ\ 45^\circ\ 90^\circ\ 135^\circ\ 180^\circ]$, so we have four possible directions, which means 8 possible neighbor pixels to check.

TEST 3: Analytical VS practical implementation

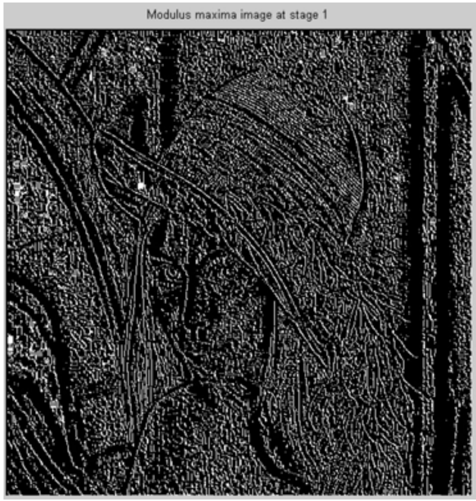
ANALYTICAL APPROACH PRACTICAL APPROACH



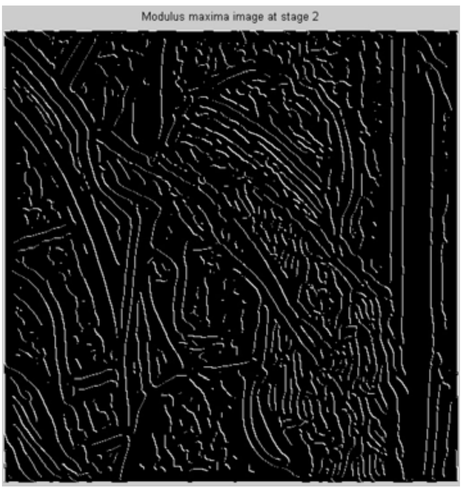
Levels=1-4 Levels=4-8 Levels=1-4 Levels=4-8

ANALYTICAL APPROACH **PRACTICAL APPROACH**

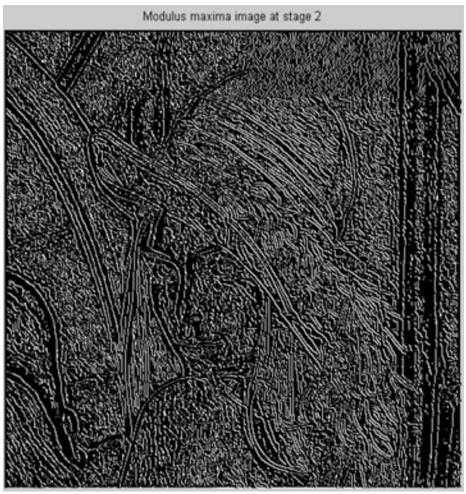
Level 1



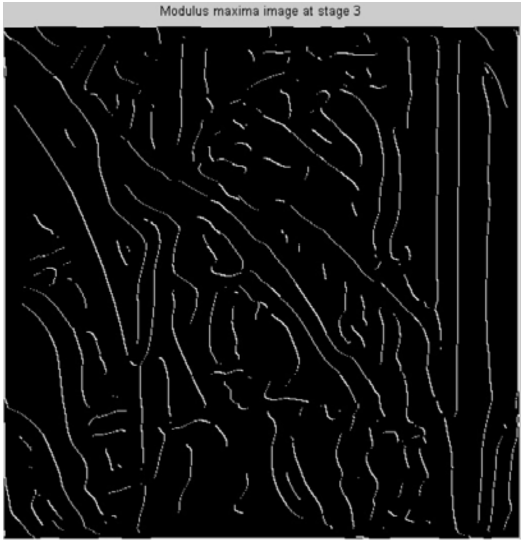
Level 1



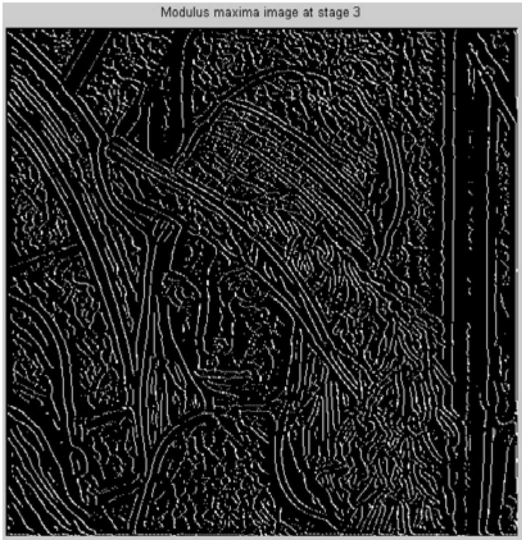
Level 2



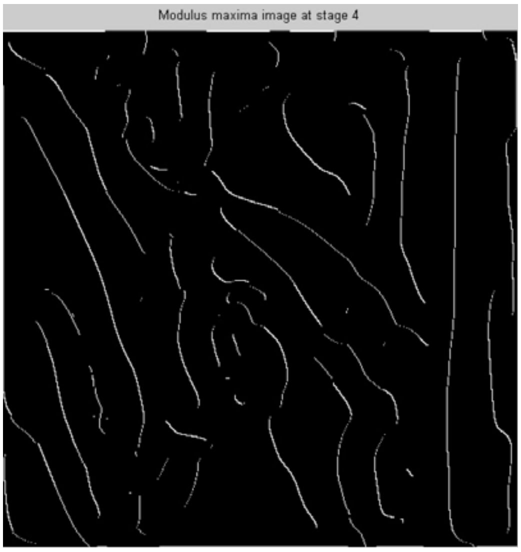
Level 2



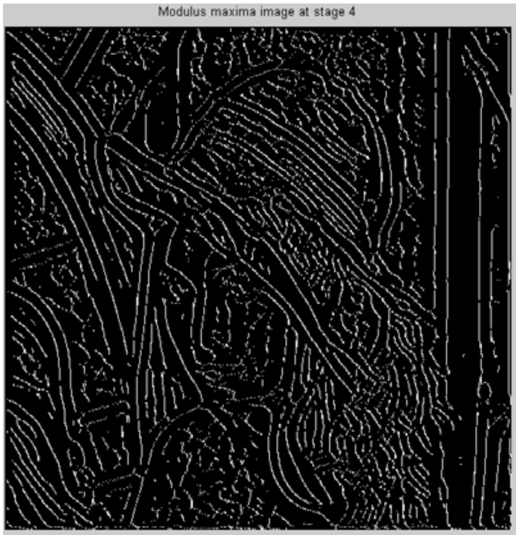
Level 3



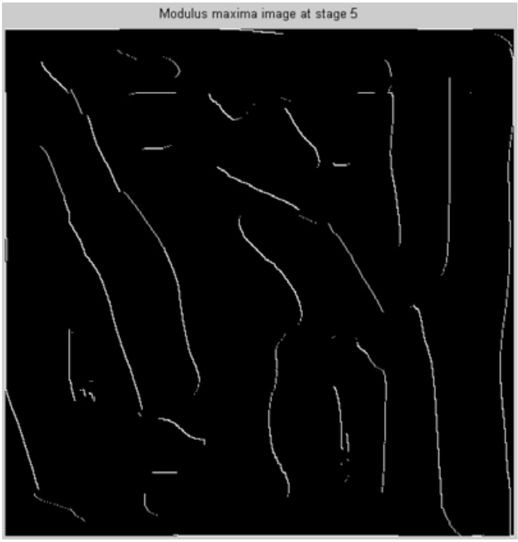
Level 3

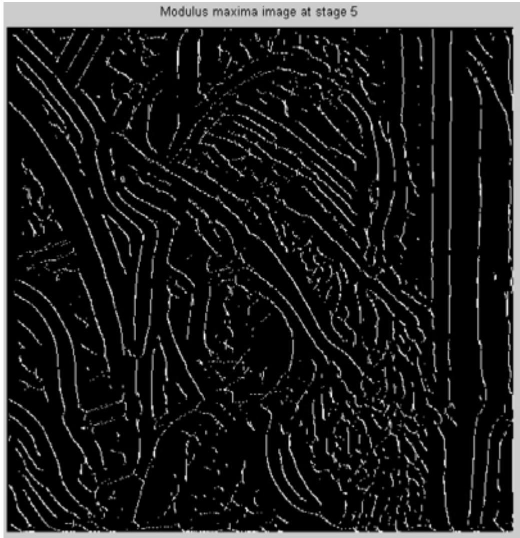


Level 4

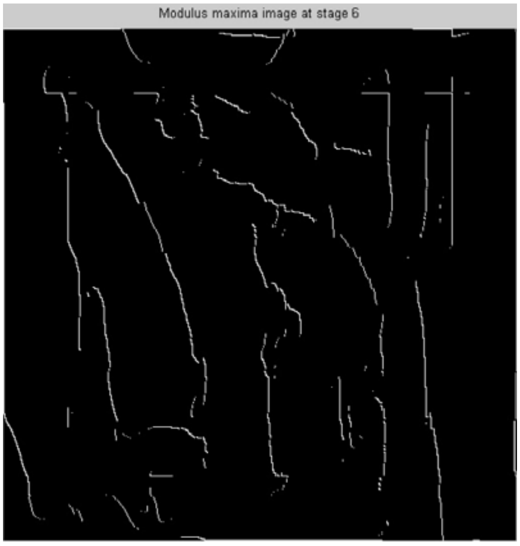


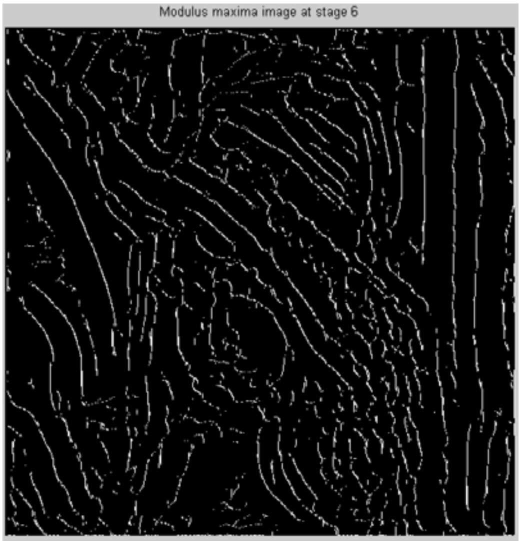
Level 4





Level 5 **Level 5**





Level 6 **Level 6**

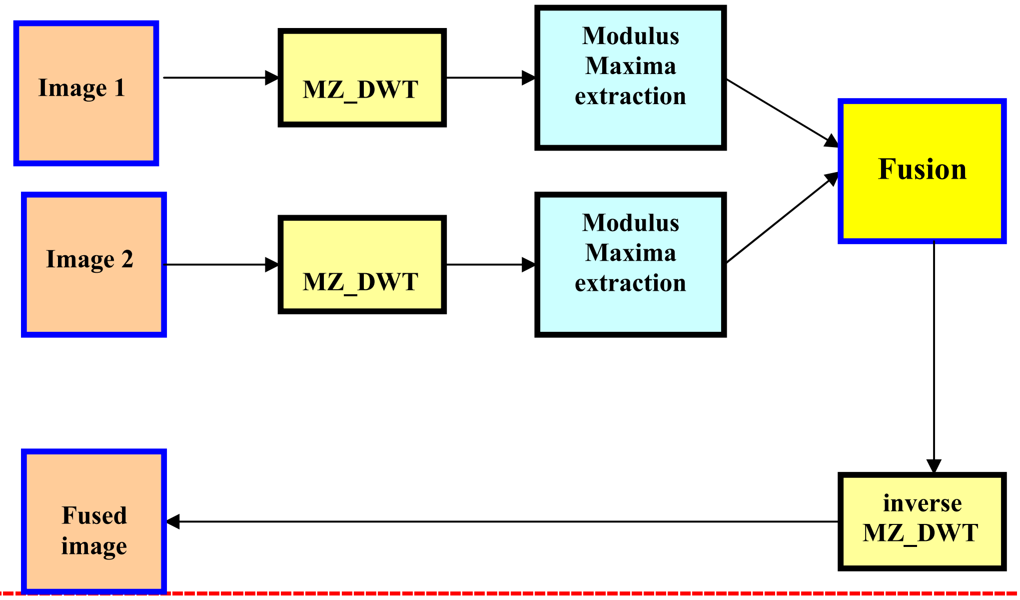
Conclusion: From the above, we notice that the analytical approach preserves less edges as the number of levels is increased but does not insert "blank" areas in the lower levels (notice the white areas in the images of the practical implementation in the first 4 levels). This is done because the filters in the "practical" version (the scaling function in the analytical one) have small size in the first levels (remember that at each stage they are upsampled; their size is doubled) and the convolution with the input cannot represent the whole information leading to loss of some data. Finally, it is worth mentioning that the computational complexity and delay was rapidly increased when the input was the 512x512 "Lenna" image and the result was extracted after several minutes of calculations.

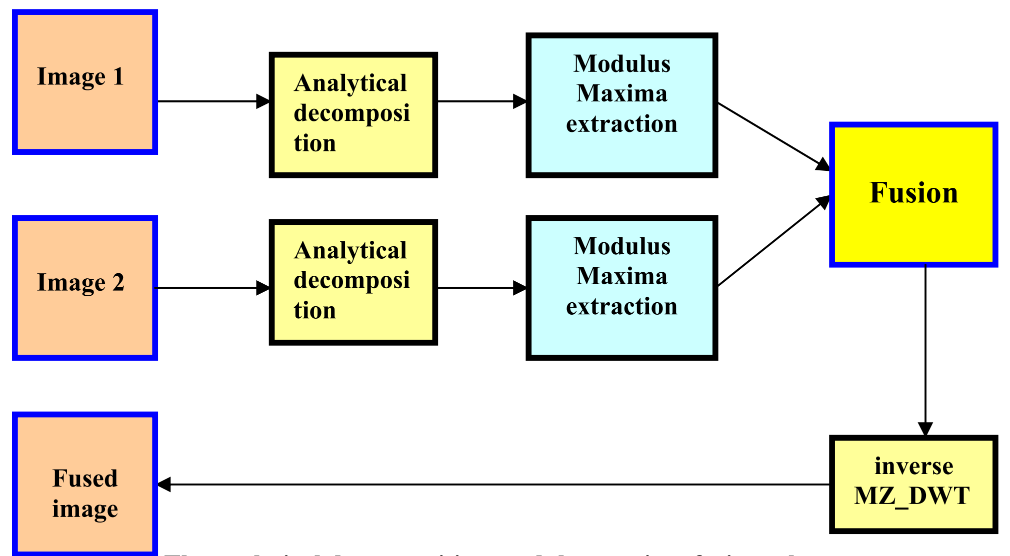
3.2. Fusion of images using their multiscale edges

3.2.1. Point Representation fusion

Having extracted the multiscale edge representation of the input images we can fuse them and produce the final fused image. Reconstruction from edge representation is not an easy issue; on the contrary, it is a very complicated problem. Mallat and Zhong proposed an algorithm at [Mallat and Zhong, 1992] that reconstructs a very close and visually indistinguishable approximation of the input image from its multiscale edge representation. Thus, this representation of the image is complete. The algorithm makes use of complicated mathematical models and algebraic equations and is not very easy implemented. In this report, it is attempted to reconstruct the final image using the inverse MZ-DWT, using application of filters as presented in the previous sections. The general fusion scheme is the following:

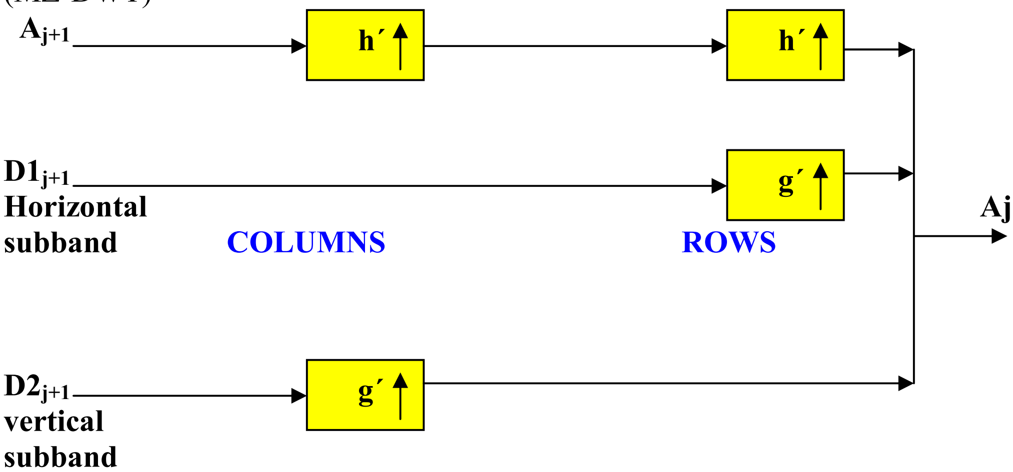
The MZ_DWT-modulus maxima fusion scheme





The analytical decomposition modulus maxima fusion scheme

In this report we will try to reconstruct the fused image using the inverse Dyadic Discrete Wavelet Transform, mentioned as Mallat-Zhong Discrete Wavelet Transform (MZ-DWT)

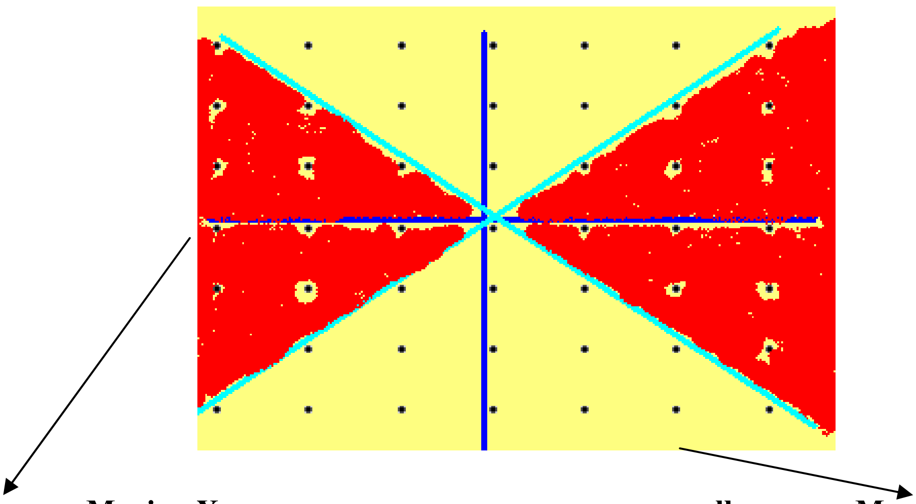


Reconstruction procedure at level j

The above scheme will be slightly modified and adjusted to the needs of edge representation, as follows:

- the coefficients $\mathbf{A_{j+1}}$ remain the same**:** they are the lowpass approximations of the input image at scale j+1, computed by cascaded filtering with the lowpass filter h, along both dimensions
- the coefficients $\mathbf{D1_{j+1}}$ are replaced by the modulus maxima $\mathbf{MaximaX_{j+1}}$ at the horizontal direction
- the coefficients $\mathbf{D2_{j+1}}$ are replaced by the modulus maxima $\mathbf{MaximaY_{j+1}}$at the vertical direction

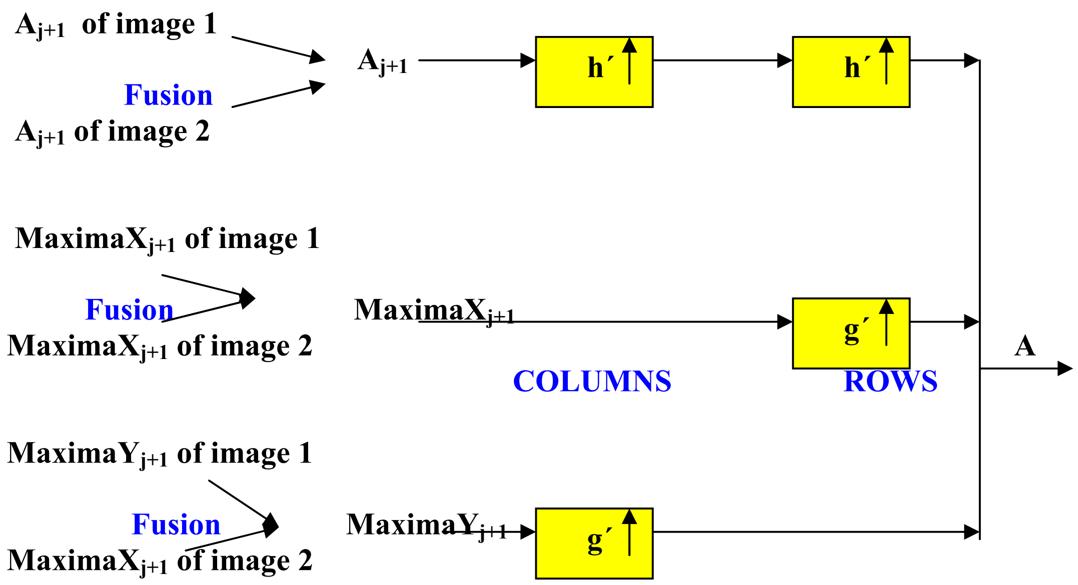
How the coefficients **$\mathbf{MaximaX_{j+1}}$, $\mathbf{MaximaY_{j+1}}$ are computed**? Simply by splitting the angle range in two intervals, one for the vertical and one for the horizontal direction, as illustrated below:



red area=MaximaX **yellow area=MaximaY**

The points that are located within the red neighborhood, as indicated by the angle at point (x,y) and are found to be maxima will be referred as MaximaX(modulus maxima at x-direction) while those found within the yellow region will be considered as MaximaY.

The fusion scheme takes its final form:

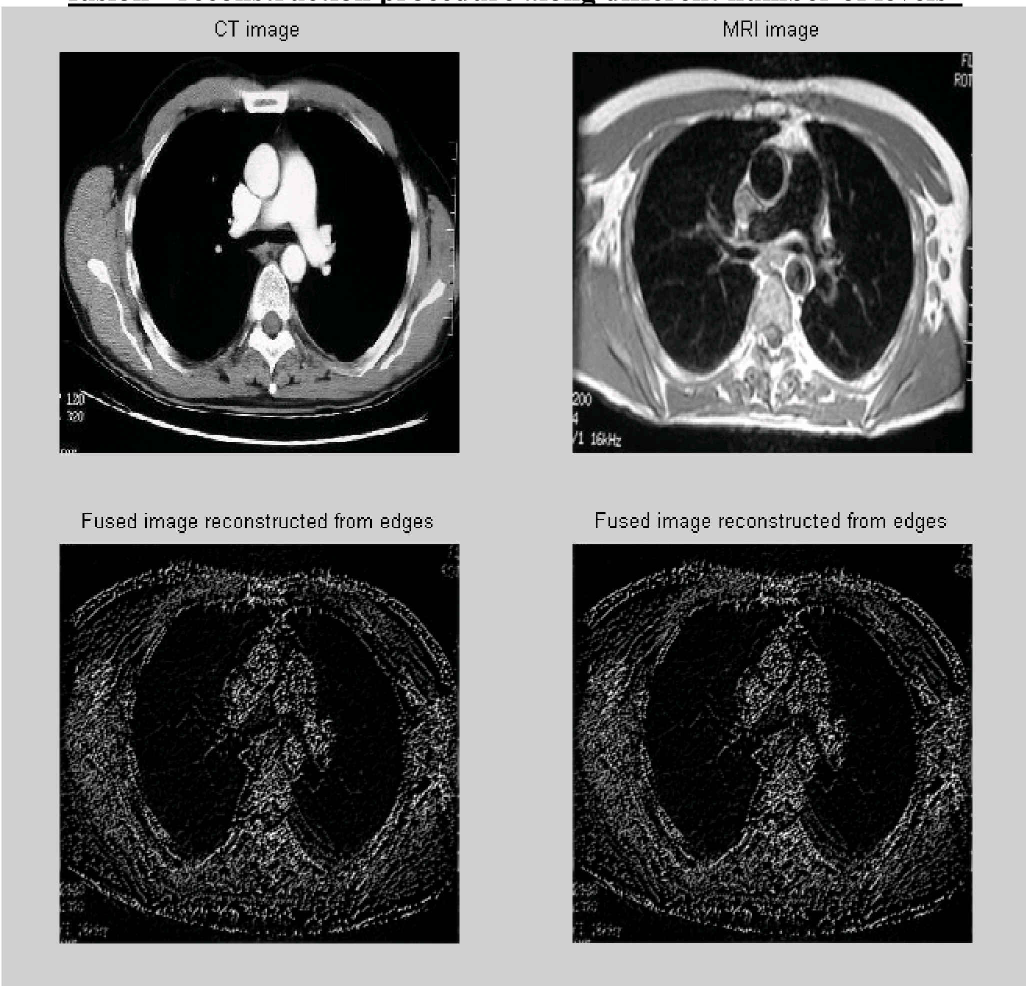


Reconstruction procedure at level j

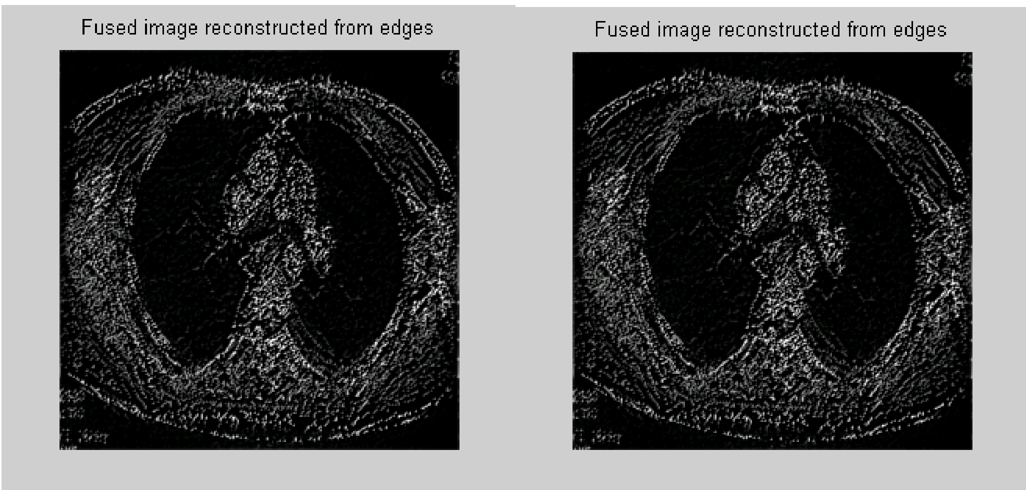
Following the above approach it is attempted to reconstruct the fused image, fusing the multiscale edges of the input images. Instead of combining all wavelet coefficients, we can fuse the two edge representations of the two source images. The fusion rule is simple: average the lowpass approximations and take the union of the edges. If an edge point (x, y) is found in both images then the one edge point will overwrite the other.

A sequence of results is presented in the following pages in order to test the algorithm. At first we attempt to evaluate the reconstruction result along different number of levels.

“The MZ-DWT decomposition->modulus maxima calculation-> fusion->reconstruction procedure along different number of levels”

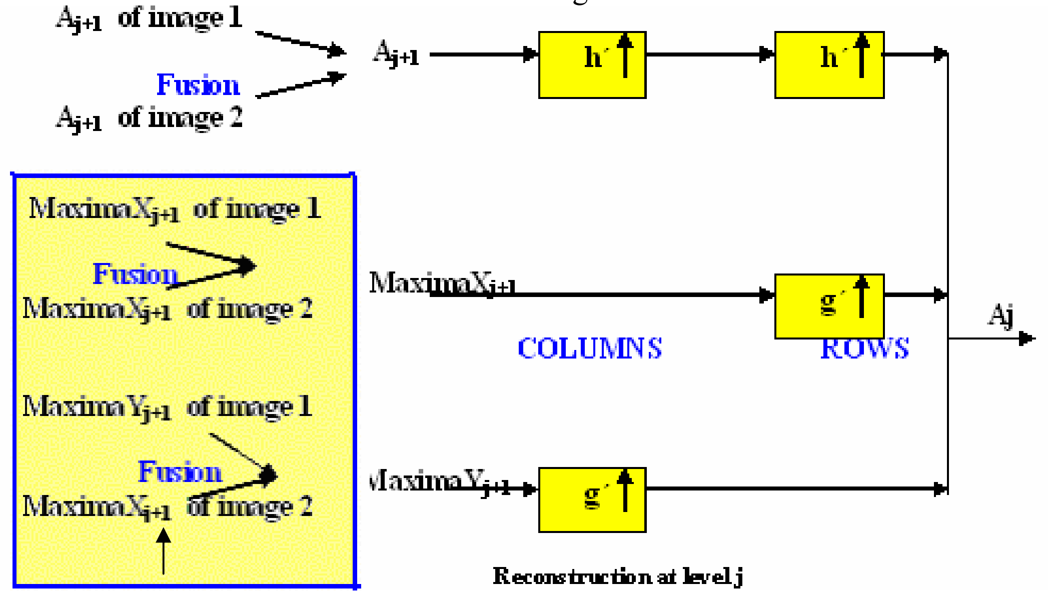


Levels=1 **Levels=2**

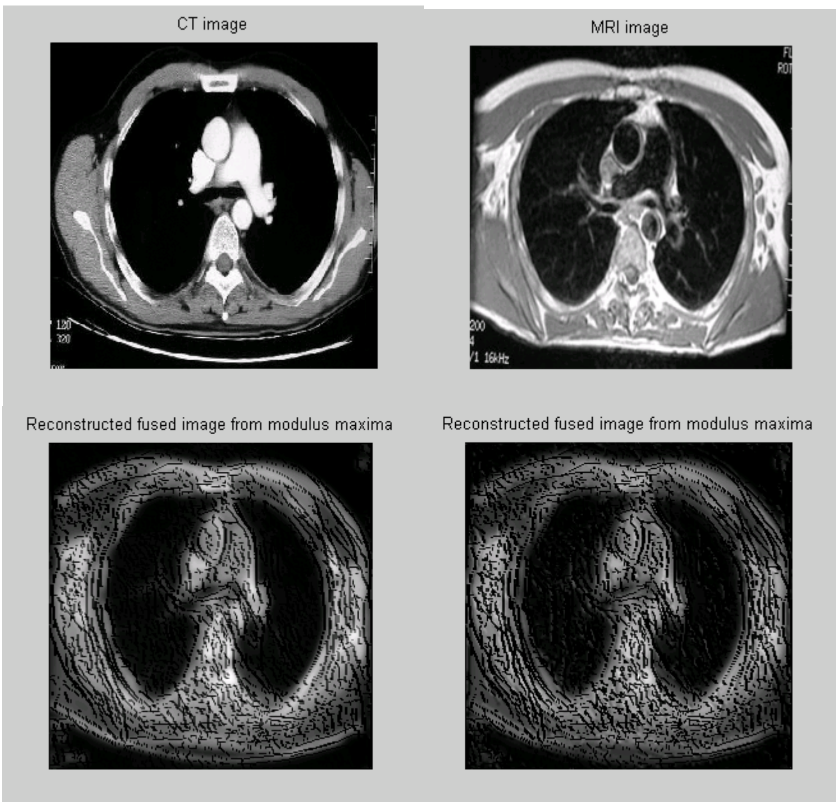


Levels=3 **Levels=4**

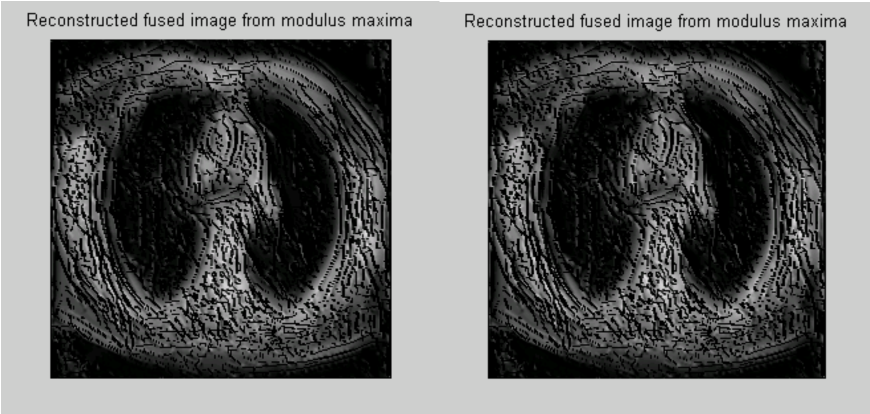
We try another version: calculate the modulus maxima using the analytical approach using mathematical forms, the Gaussian scaling function and use the MZ-DWT reconstruction scheme to obtain the fused image.



ANALUTICAL COMPUTATION

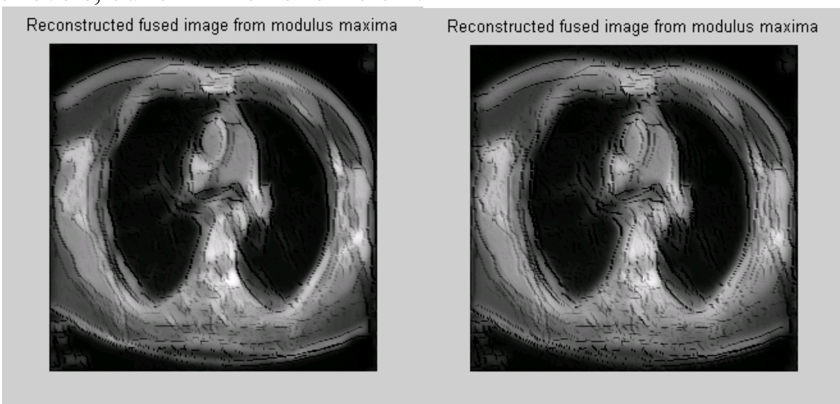


Levels=1 **Levels=2**

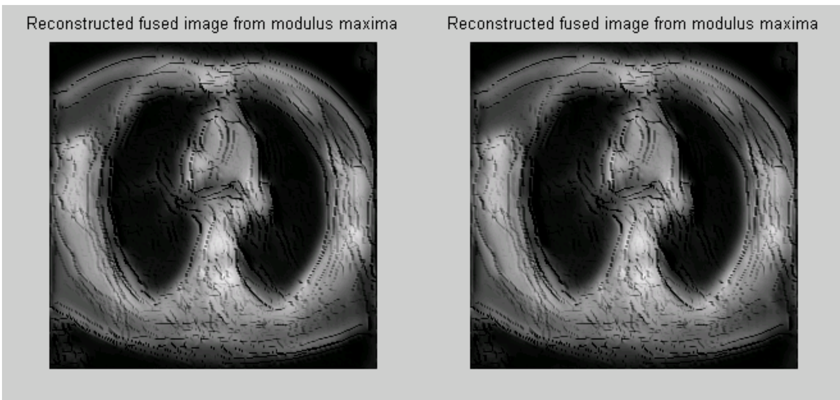


Levels=3 **Levels=4**

<u>Conclusion</u>: As we can observe, the reconstruction from the edge representation is not perfect. On the contrary, it is very poor, as the final, reconstructed image still preserves a respectable amount of edges, which, of course, is not desirable. The result deteriorates as the number of levels is increased. As a solution, we can replace the value of the modulus maxima points (1 if the point is a local maximum, 0 otherwise) with the value of the Modulus of the image at this point. The result is much more encouraging, especially in the first two levels, but still it is not efficient.



Levels=1 **Levels=2**



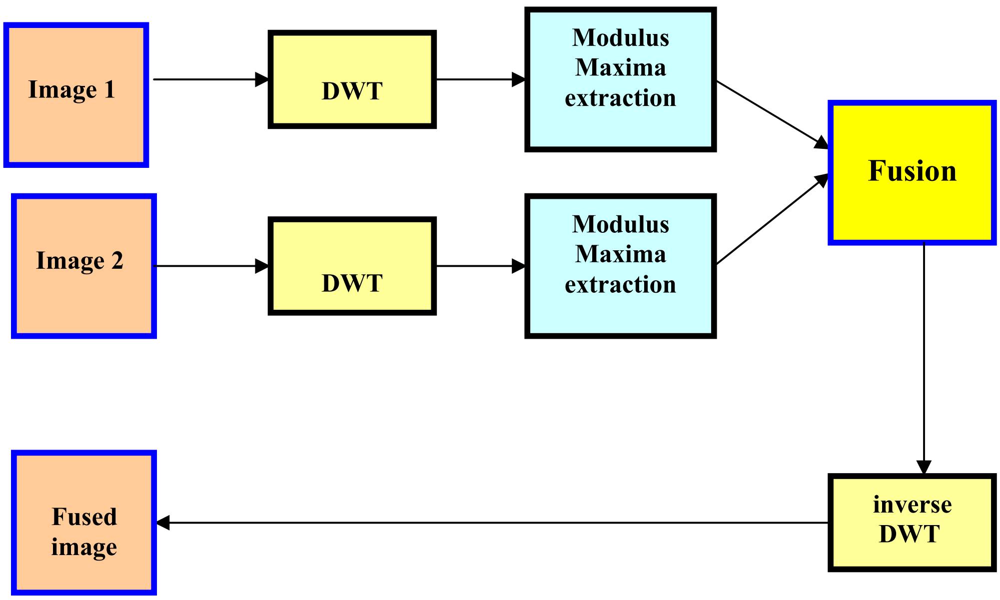
Levels=3 **Levels=4**

On the grounds that the MZ-DWT suffers from poor directionality, as only the two of the three orientations are represented in the coefficients, another solution needs to be found. As a final presentation, a novel method for multimodality medical image fusion proposed by Guilong Qu, Dall Zhang and Pingfan Yan is analyzed. It is based on the wavelet transformation modulus maxima and emphasizes the extraction of the edge and margin information [22]. The advantages of the algorithm are:

- It offers better preservation of both edge features and component information in the fused image
- Fusion can be performed at different levels

Wavelet transform modulus maxima has been used to extract explicit important features of images. However, it could become a mean for image fusion performance. Mallat and Zhong in [20, 21] introduced a projection algorithm that can compute an efficient, visually identical to the original image, approximation from their edges. However, due to complexity, the following scheme is adopted:

The DWT-modulus maxima fusion scheme



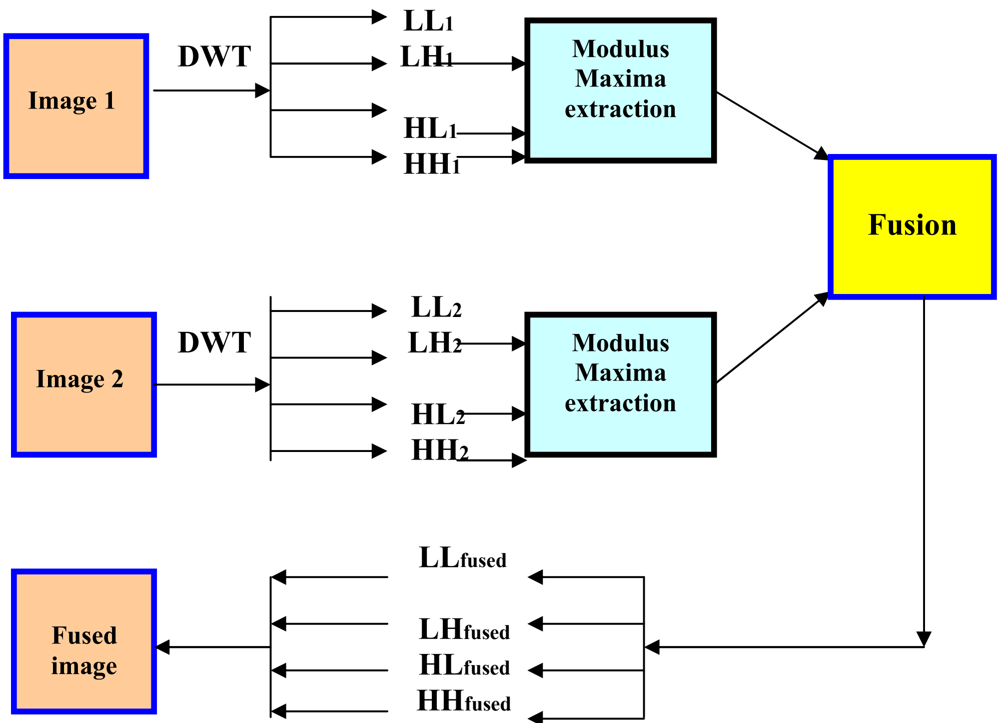
At first, the input images are decomposed in the low and high pass subbands by using the Discrete Wavelet Transform, which has already been widely analyzed in the section 2.1.1. As already mentioned, this transform decomposes each image into one low-frequency subband image and three high-frequency subband images (one along each orientation) at each level.

The wavelet transform modulus maxima for each high-frequency subband are calculated using the analytical approach of section 2.2.2.

The information from each image is then combined using a certain fusion rule; average the lowpass subbands of the last level and select the highpass subbands of the greatest magnitude at each level of decomposition, as the larger absolute transform values correspond to sharper intensity change such as edges, lines and region boundaries [22].

To make things clearer the general procedure is illustrated graphically in the following figure:

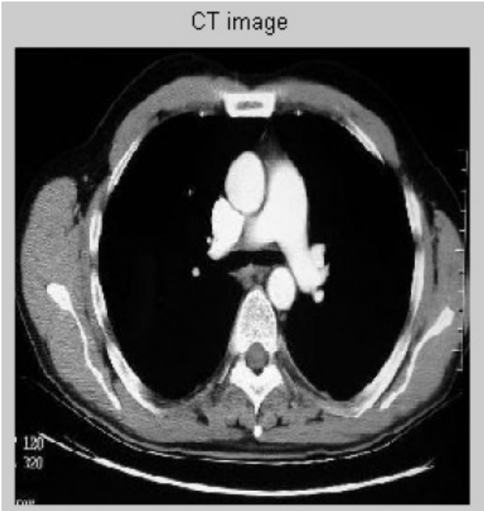
The DWT-modulus maxima fusion scheme for one level of decomposition



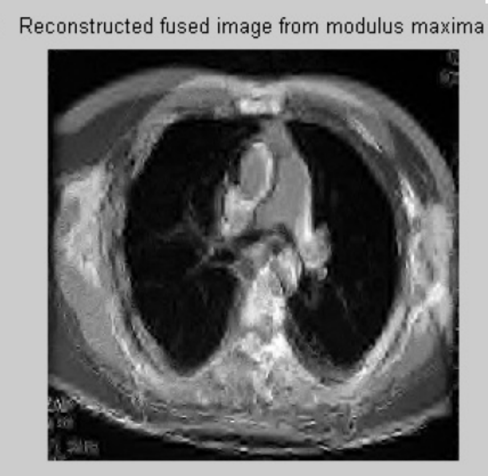
One thing to take into consideration is the number of decomposition levels to be used and the number of stages in which the modulus maxima will be calculated. These two numbers are not necessarily the same. For example, we can use four levels of decomposition and compute the modulus maxima only at the first level, which seems to be the optimum suggestion, as the first level preserves grater number of edges(=useful information).

Finally, it is worth making a remarkable notice: in order to achieve the best reconstruction result, when a point (x, y) is found to be a local maximum its value will be replaced by the value of the corresponding wavelet decomposition coefficient at this point.

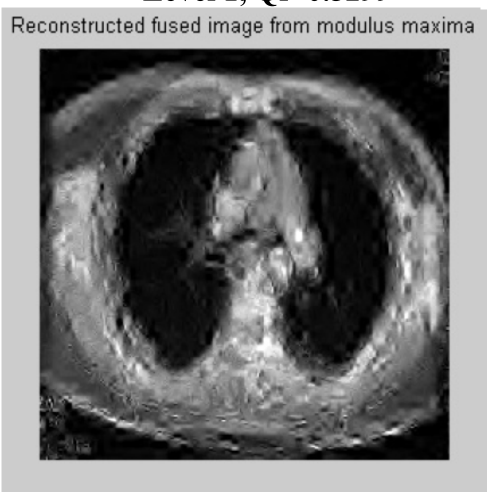
The results of this approach are presented below:



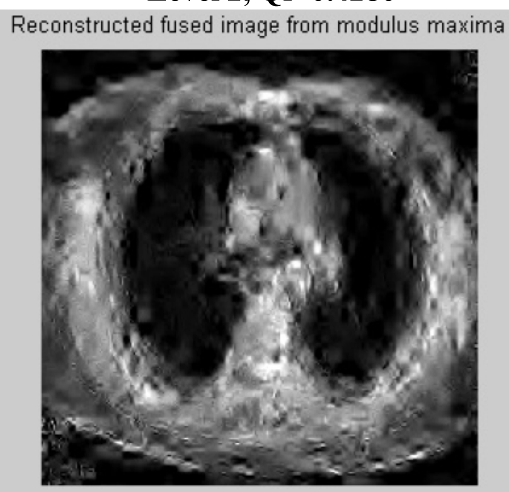
Level 1, Qf=0.5299



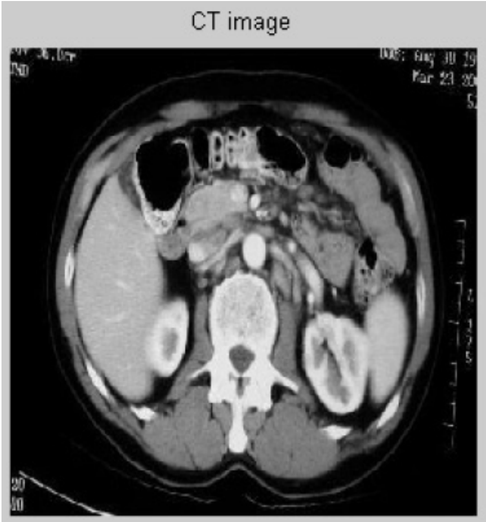
Level 2, Qf=0.4280



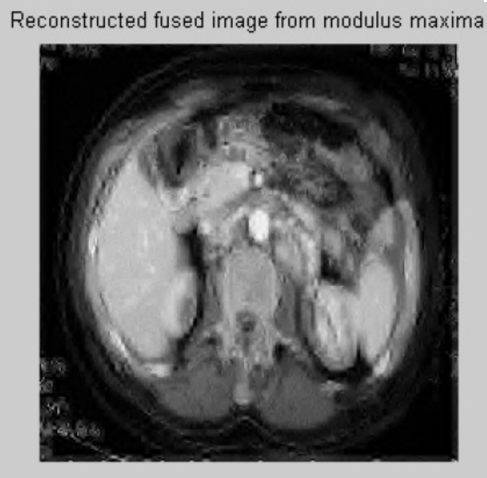
Level 3, Qf=0.3056



Level 4, Qf=0.2234

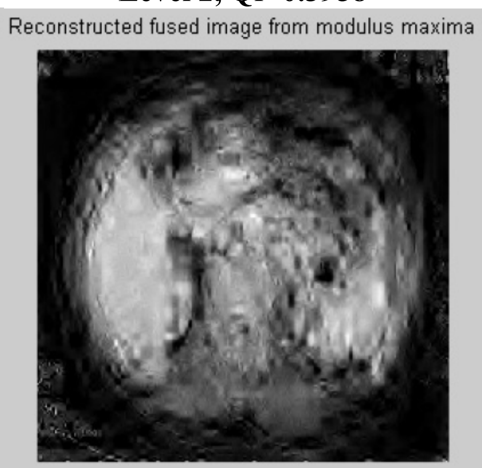


Level 1, Qf=0.4877

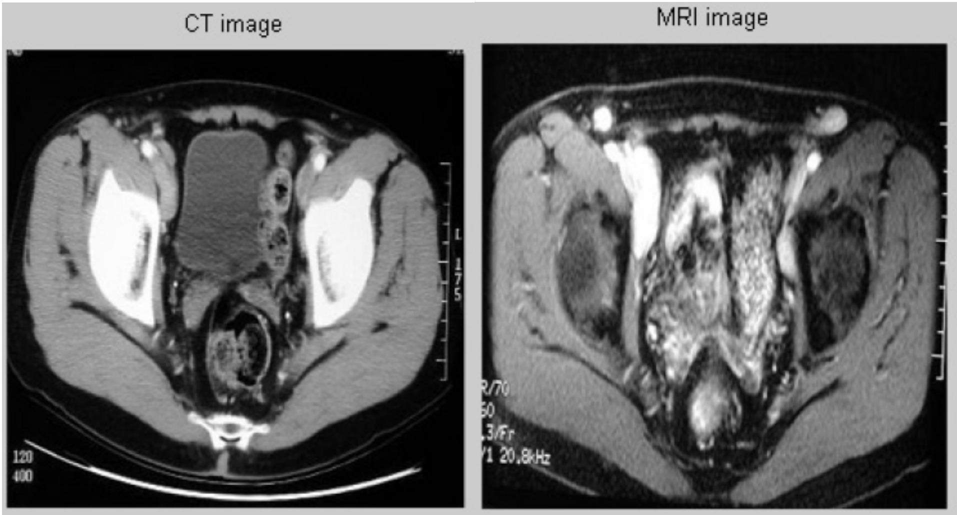


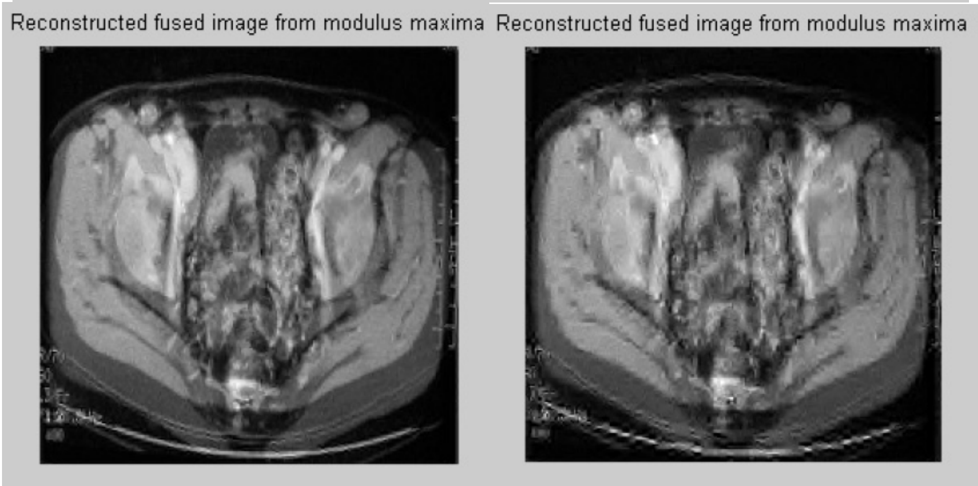
Level 2, Qf=0.3958

Level 3, Qf=0.2971

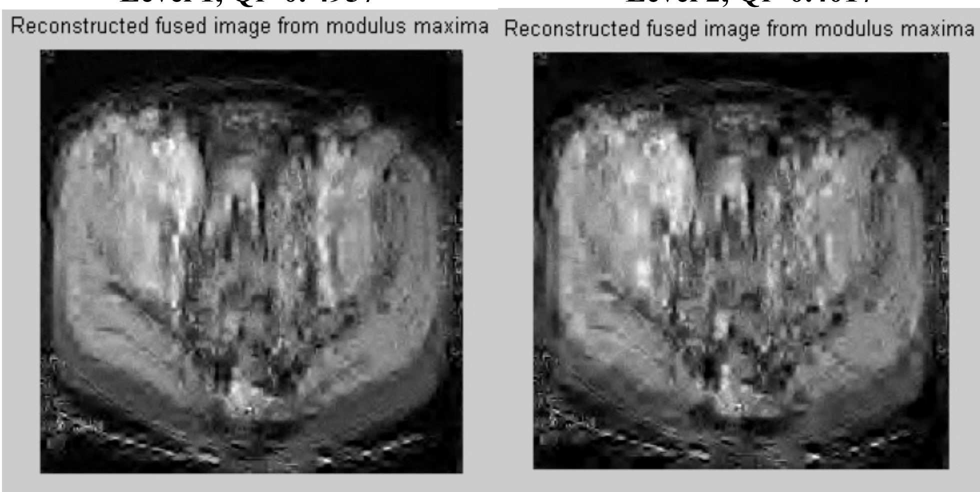


Level 4, Qf=0.2416

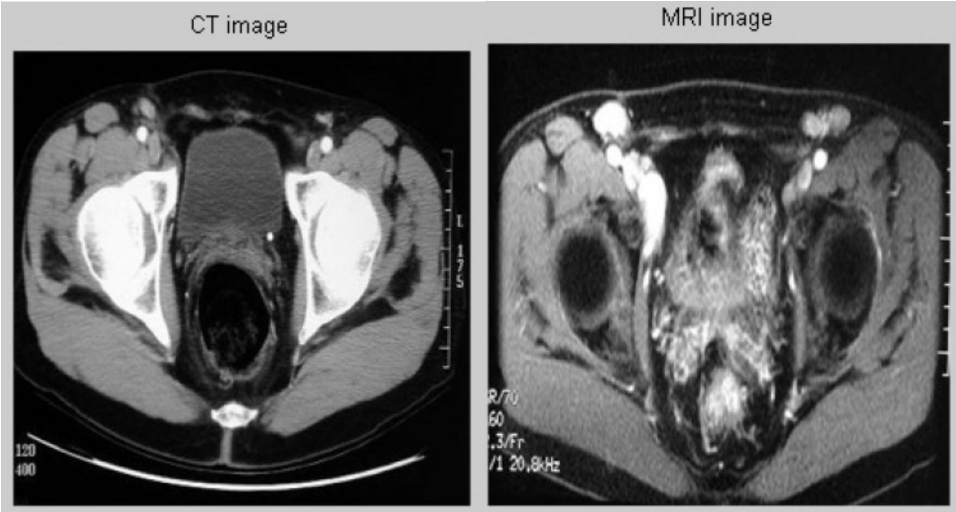




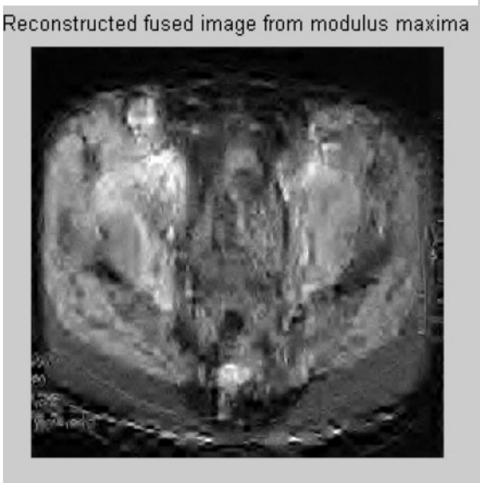
Level 1, Qf=0. 4957 **Level 2, Qf=0.4017**

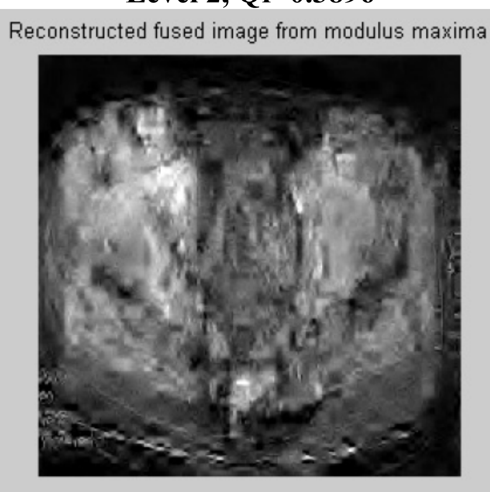


Level 3, Qf=0.2952 **Level 4, Qf=0.2270**



Level 1, Qf=0.4831 **Level 2, Qf=0.3896**





Level 3, Qf=0.2894 **Level 4, Qf=0.2176**

The evaluation results are summarized in the following table:

Quality metric Qf "DWT maxima calculation and reconstruction " Image pair	Level 1	Level 2	Level 3	Level 4
1	0.5299	0.4280	0.3056	0.2234
2	0.4877	0.3958	0.2971	0.2416
3	0.4957	0.4017	0.2952	0.2270
4	0.4831	0.3896	0.2894	0.2176

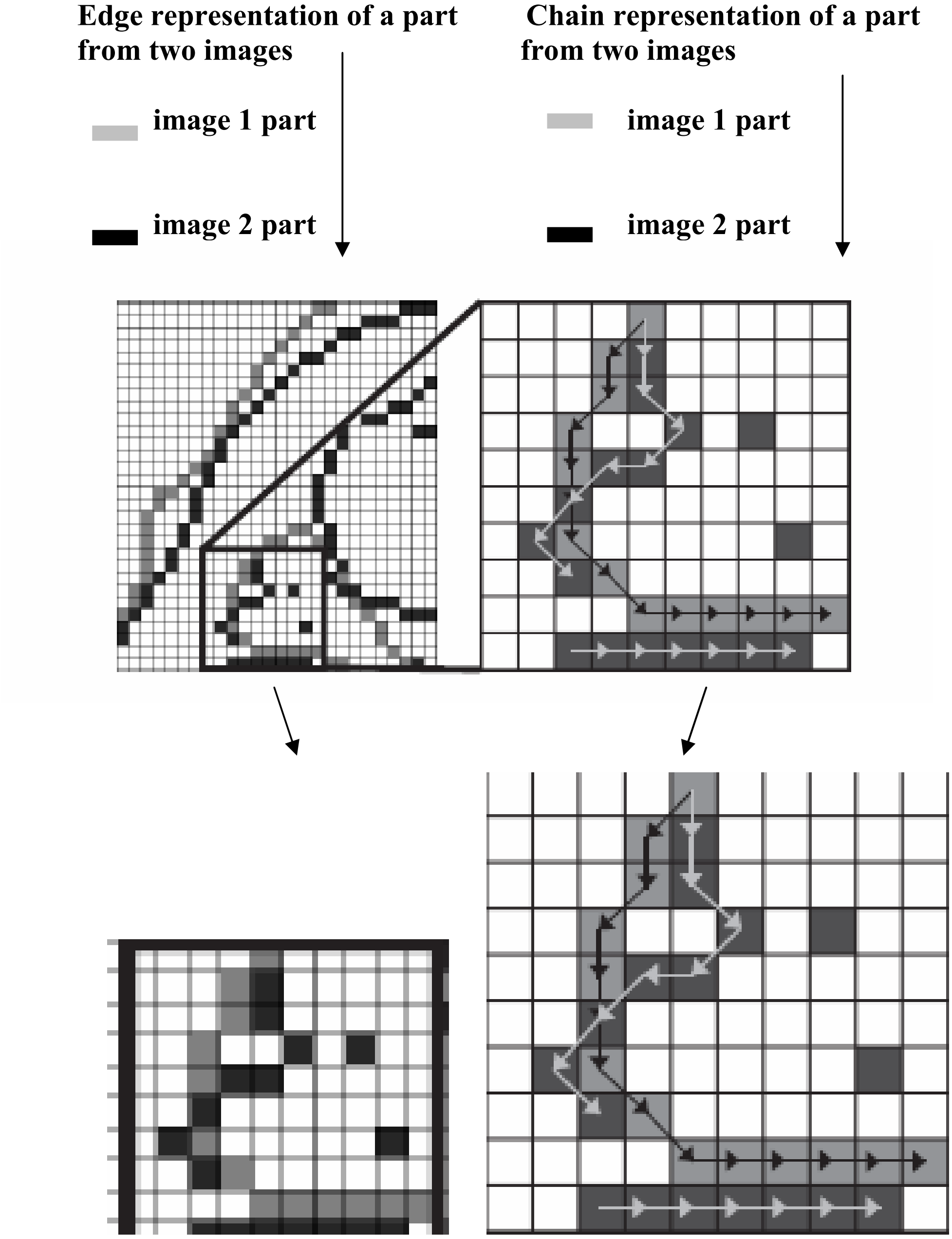
**The computation of the modulus maxima was performed taking only one stage, while the decomposition levels had values in the range [1-4].

Conclusion: The final algorithm proved efficient enough to fuse and reconstruct the two images from their multiscale edges, especially in the first two decomposition levels, where the result is impressive contrary to the methods presented before. It is worth mentioning that computing the modulus maxima of each highpass wavelet coefficient in a level higher than one produced exactly the same results, so our choice for only one stage of maxima calculation was proved efficient, simple and mainly fast!

3.2.2 Chain representation fusion

Except for the point edge representation of the input images described in the previous section, there is another approach: the chain edge representation. Due to its complexity it will not be implemented and further analyzed in this report but we simply mention its main characteristics for academic purposes only.

Sharp variations of images often belong to curves in the image plane [Mallat and Zhong, 1992]. Along these curves the image density can be singular in one direction and varying smoothly in the vertical direction. These curves are sometimes more meaningful than edge points. Mallat and Zhong created a high-level structure, a set of chains of edge points in order to represent the image. Two neighbored local maxima points belong to the same chain if their respective position is perpendicular to the direction indicated by the Angle of the image at each point (x, y). The idea is to recover the curves along which the image profile varies smoothly, so maxima points are chained together only where the Modulus(x,y) of the image has close values [9]. The chain image representation is proved very useful in the image fusion process.



Enlarged selected part of edges

The corresponding part represented by chains

Following this procedure one oriented graph for image is constructed. Fusion will take place in this form this time, so much more complicated fusion rules need to be developed. The goal now is to combine the two oriented graphs from the two images to a new one, from which the final, fused image will be reconstructed. This is not an easy part!! Distance measures between the nodes of the graphs, merging or deleting nodes needs to take place and so on. Graphs handling is a very complex and computationally insufficient problem. However the fusion result is satisfactory.

Chapter 4

Performance evaluation and comparison

Having widely analyzed the fusion algorithms based on wavelet analysis in the previous chapter, we attempt to evaluate and compare their performance via a well defined and objective, as possible, metric. This is not an easy part as meaningful comparison of image fusion techniques is often application dependant. We can evaluate the results through two approaches:

- The **qualitative** one, based on the human vision system and observe the fused images, which needs experience and can be misleading
- The **quantitative** one, based on a well defined mathematical quality metric

For the needs of the second approach, the quality metric Qf (mentioned in the previous section) is introduced, based on the presentation of Gemma Piella and Henk Heijmans [12]. Its advantage is that it is computed by "raw" data (the two input images and the fused one) without demanding the existence of a model-ideal (commonly known ground-truth) reference image. Its disadvantage is that, in our implementation, the computational complexity and delay was remarkable, as the results needed a couple of minutes to be extracted and its performance deteriorates as the system equipment gets old and the size of the images increases. The fusion index Qf is an expansion of the quality index Q that was proposed by Wang and Bovik in [10].

Given two real-valued sequences x = [x1, x2, …, xn] and y = [y1, y2, …, yn] the following terms are defined

Variance of x and y. $$\sigma_x^{\ 2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \ , \ \sigma_y^{\ 2} = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

Covariance of x,y $$\sigma_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})\ (Y_i - \overline{Y})$$

We compute the term Qo as follows

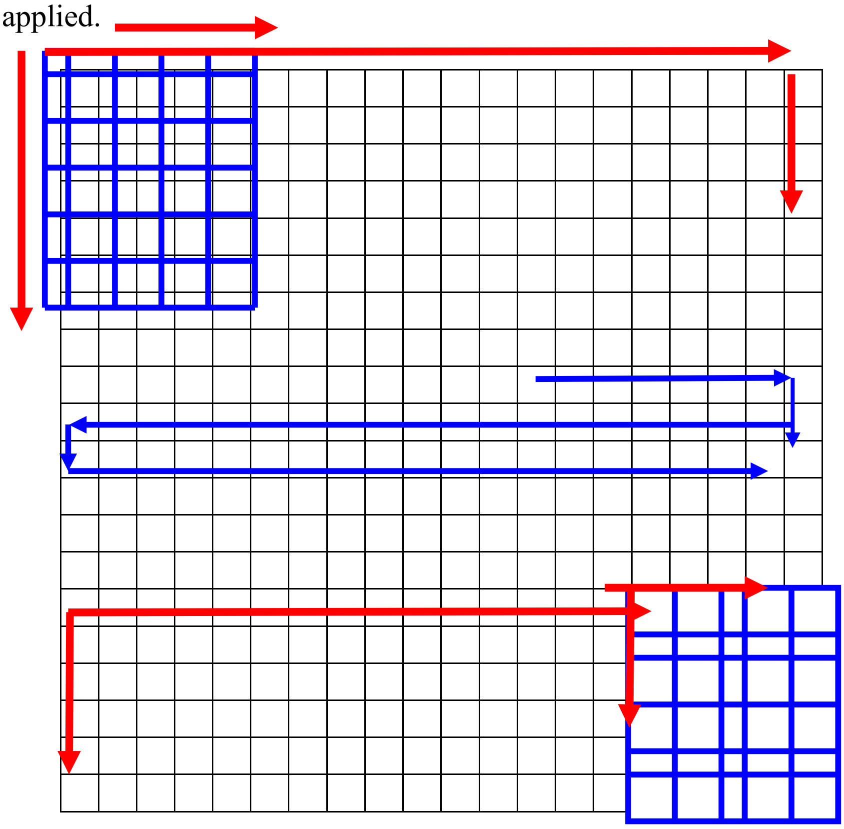
$$Q_0 = \frac{4\sigma_{XY}\overline{X}\overline{Y}}{(\overline{X}^2 + \overline{Y}^2)(\sigma_X^2 + \sigma_Y^2)} \quad ,$$

which is a measure for the similarity of the vectors x and y and takes values between -1 and 1. The maximum value Qo=1 is achieved when x and y are identical.

This measure can be expanded to images too.Instead of computing the above index point-by-point it was suggested to compute the quality index Qo over local regions and then combine the different results into a single measure. In [10], Wang and Bovik proposed the "sliding window approach': starting from the top-left corner of the images a and b, a sliding window w of fixed size (with nxn pixels) moves pixel-to-pixel over the entire image until the bottom-right corner is reached. For each window w, the local quality index Qo (a, b|w) is computed for the values a (i, j), b (i, j) where pixels (i, j) are within the window. The overall image quality index is computed by averaging all local quality indices, according to the equation:

$$Q_0(a,b) = \frac{1}{|W|}\sum_{w\in W} Q_0(a,b\,|\,w) \quad ,$$

where W is the family of all windows and |W| is the cardinality of W, the total number of windows applied.



If the window falls outside the borders of the image then no action is taken and the window moves to the next pixel.

The above procedure can be expanded for the needs of image fusion performance evaluation. We define the fusion index Q (a, b, f), where a, b are the input images and f is the fused image. Again the fusion index is computed within a sliding window. This time we denote a saliency measure (e.g. the variance, the contrast, the entropy) s (a|w) in the window w. Given the local saliencies s(a|w) and s(b|w) of the two input images a and b, we compute a local weight λ(w) between 0 and 1, indicating the relative importance of image a compared to image b [11].

$$\lambda(w) = \frac{s(a\,|\,w)}{s(a\,|\,w)+s(b\,|\,w)}$$

and the fusion index is computed from the following equation

$$Q(a,b,f) = \frac{1}{|W|}\sum_{w\in W}\big(\lambda(w)Q_0(a,f\,|\,w)+(1-\lambda(w))Q_0(b,f\,|\,w)\big)$$

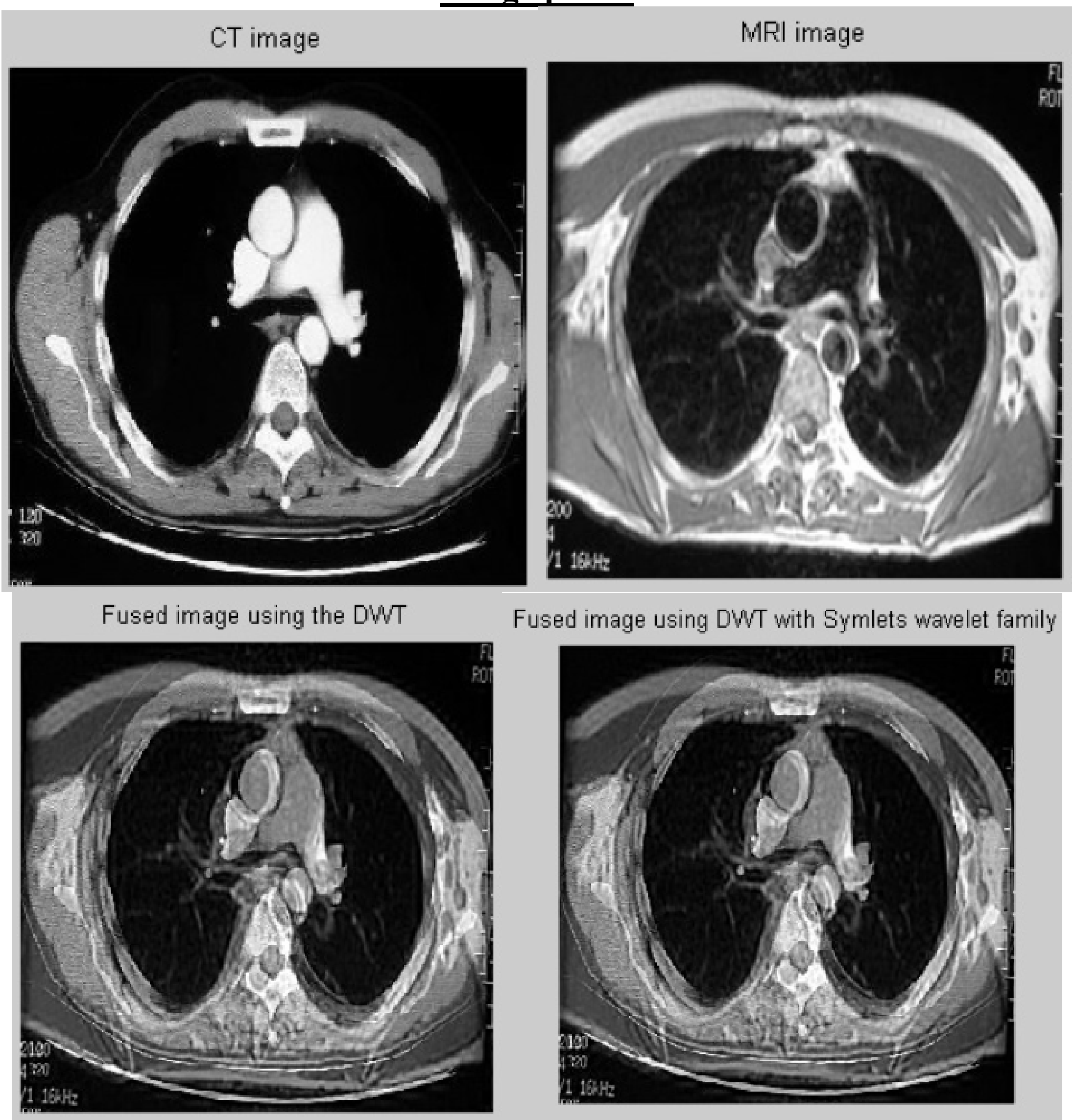
Thus in regions where image *a* has a large saliency compared to *b*, the quality index Qf (or Q (a, b, f)) is mainly determined by the input image a. Vice versa, in regions where the saliency of image b is larger than that of *a*, the input image *b* contributes to the computation of the quality index. In the following examples, the saliency measure to be adopted is the variance of the images *a* , *b*. As already mentioned, the computational complexity of the algorithm is high, since a great number of mathematical calculations along rows and columns are needed. A key point is the size of the window to be applied. In our implementation, the default window size was 8x8. It was proved efficient enough as it both extracted relatively quick results and preserved enough information from the input images. A window of larger size would delay the computation very much, especially when the input images were of big size (e.g over 256x256).

The algorithms presented in this report will be compared using this metric. A list of results follows. As a notice, we mention that the relatively low values of the quality index Q (a, b, f), from referred as Qf, are due to the form of the input images: medical, CT and MRI, images. On the grounds that some algorithms have poor performance as the number of levels is increased and others are more stable at the higher levels, the comparison takes place using two numbers of levels, 2 and 4. Some conclusions can be extracted immediately observing carefully the fused images, other conclusions are extracted through the quantitative results using the quality index Qf. The algorithms to be compared have the following properties:

1. **Discrete Wavelet Transform** using symmetric orthogonal filters of size 10
2. **Discrete Wavelet** Transform using filters of size 10 derived from the Symlets wavelet family
3. **Shift Invariant Discrete Wavelet Transform** starting with the Haar wavelet family filters
4. **Dual Tree Complex Wavelet Transform**, using both its Real and Complex version
5. **Modulus maxima fusion method** with Discrete Wavelet Transform decomposition method using orthogonal filters of size ten, analytical modulus maxima calculation using the Gaussian scaling function of size 32 points at only the first stage

Test 1 : Number of levels = 2

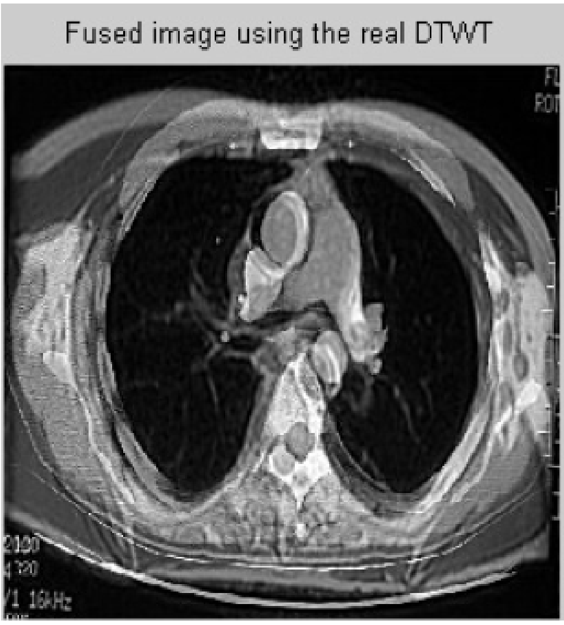
Image pair 1



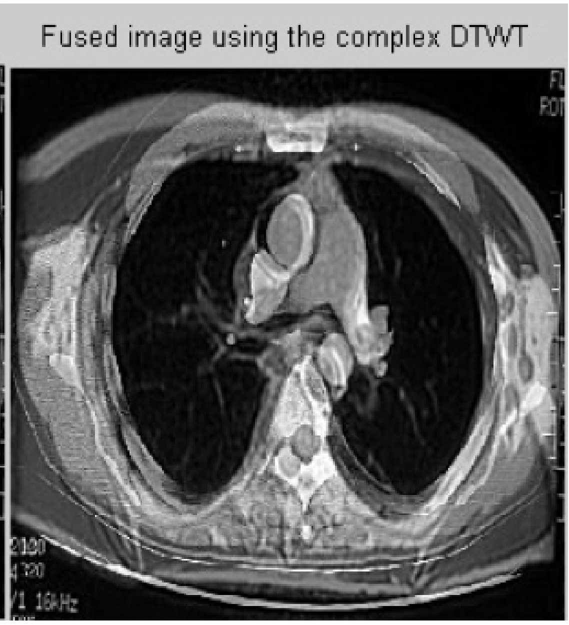
Qf=0.5813 Qf=0.6325

Qf=0.6714

Qf=0.4280

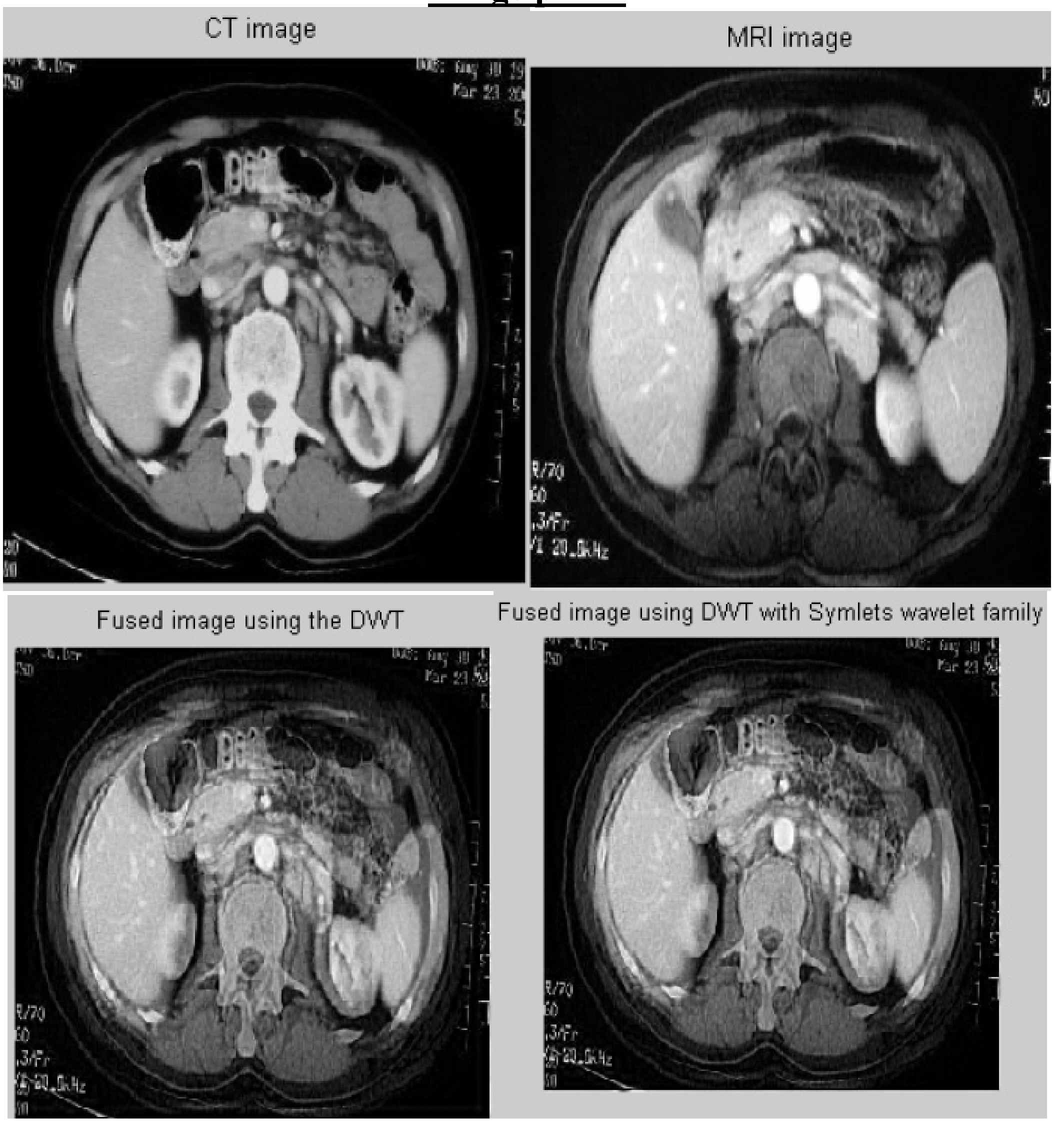


Qf=0.6466



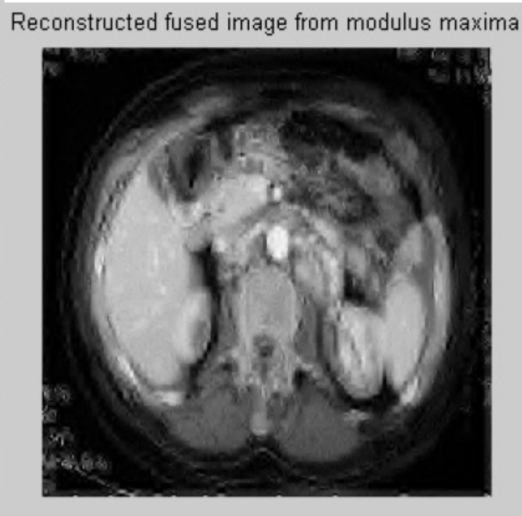
Qf=0.6490

Image pair 2



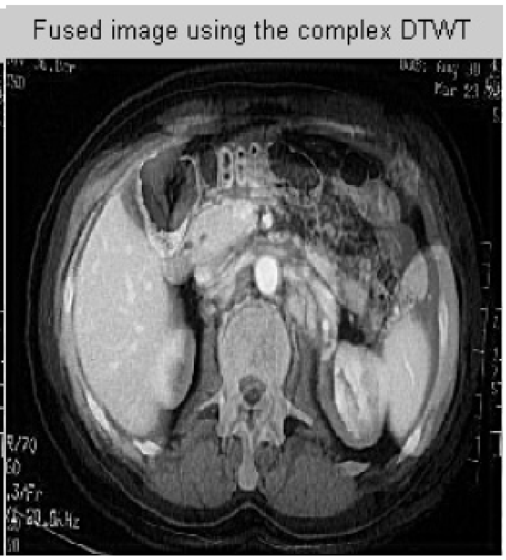
Qf=0.5740 Qf=0.5955

Qf=0.6287



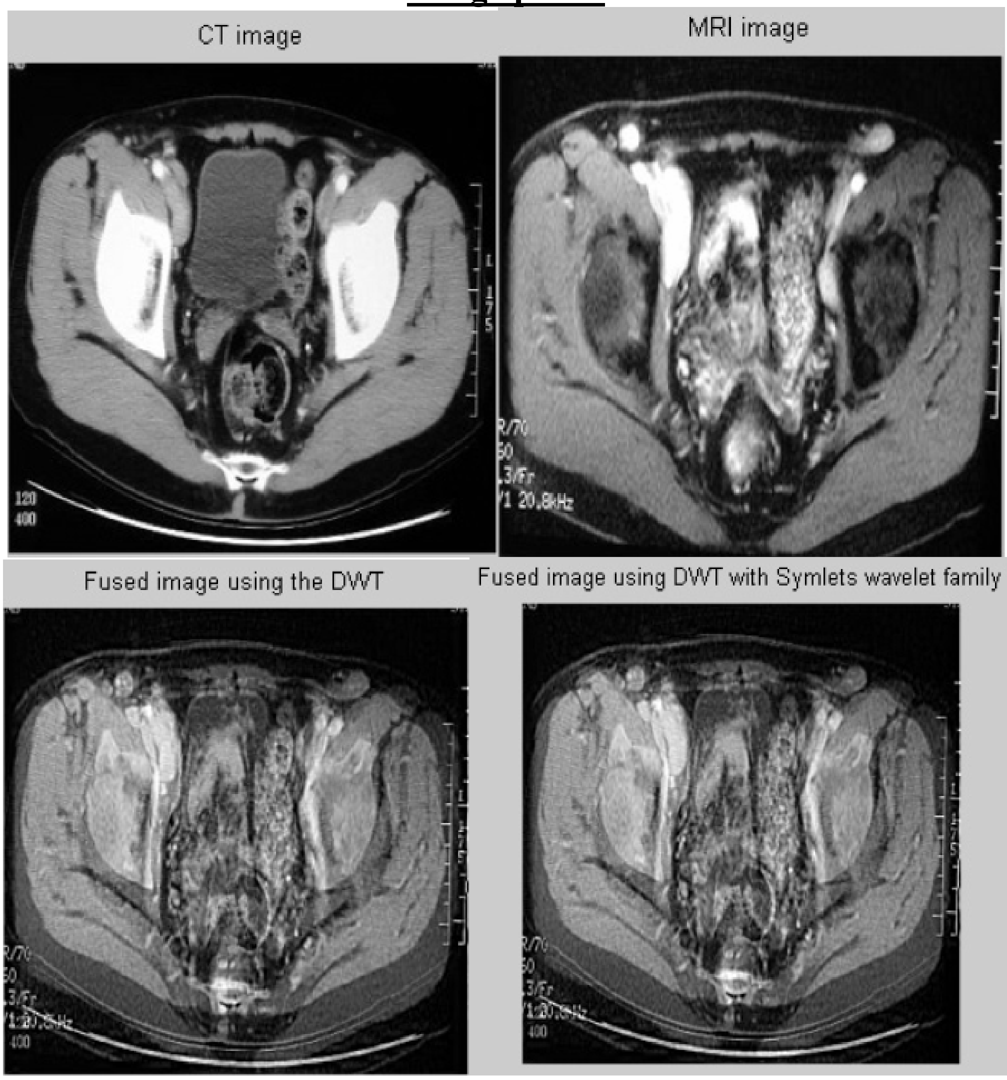
Qf=0.3958

Qf=0.6056

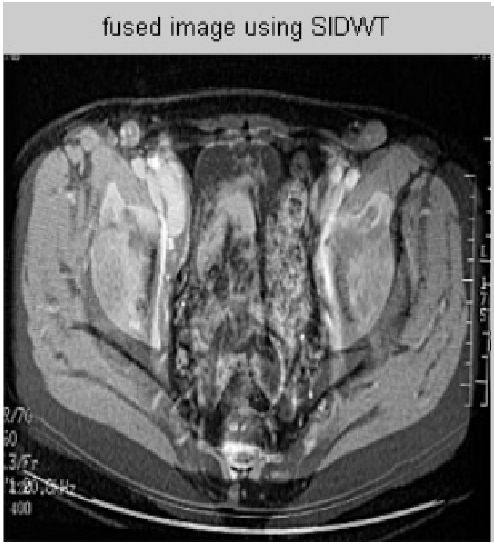


Qf=0.6097

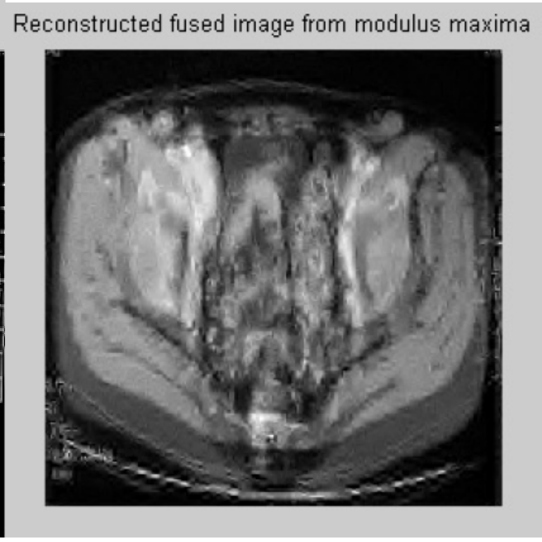
Image pair 3



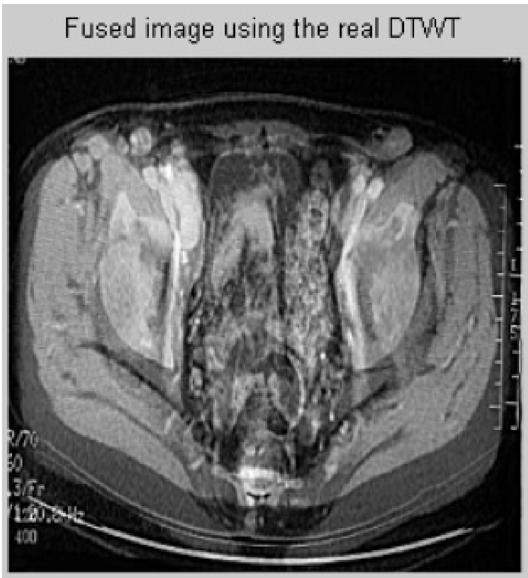
Qf=0.5549 Qf=0.6621



Qf=0.7022



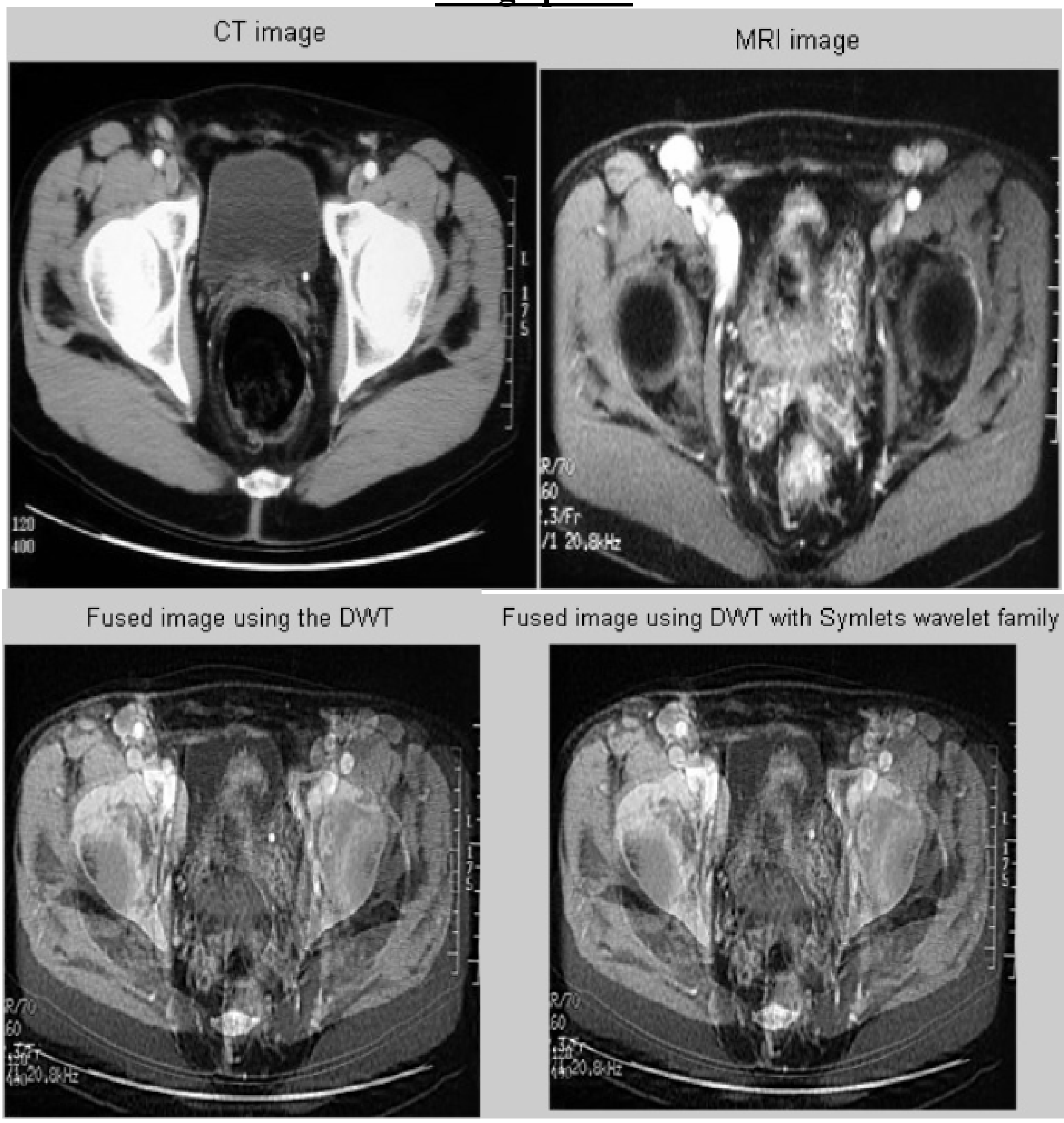
Qf=0.4017



Qf=0.6742

Qf=0.6773

Image pair 4



Qf=0.5504 Qf=0. 6465

Qf=0.6870

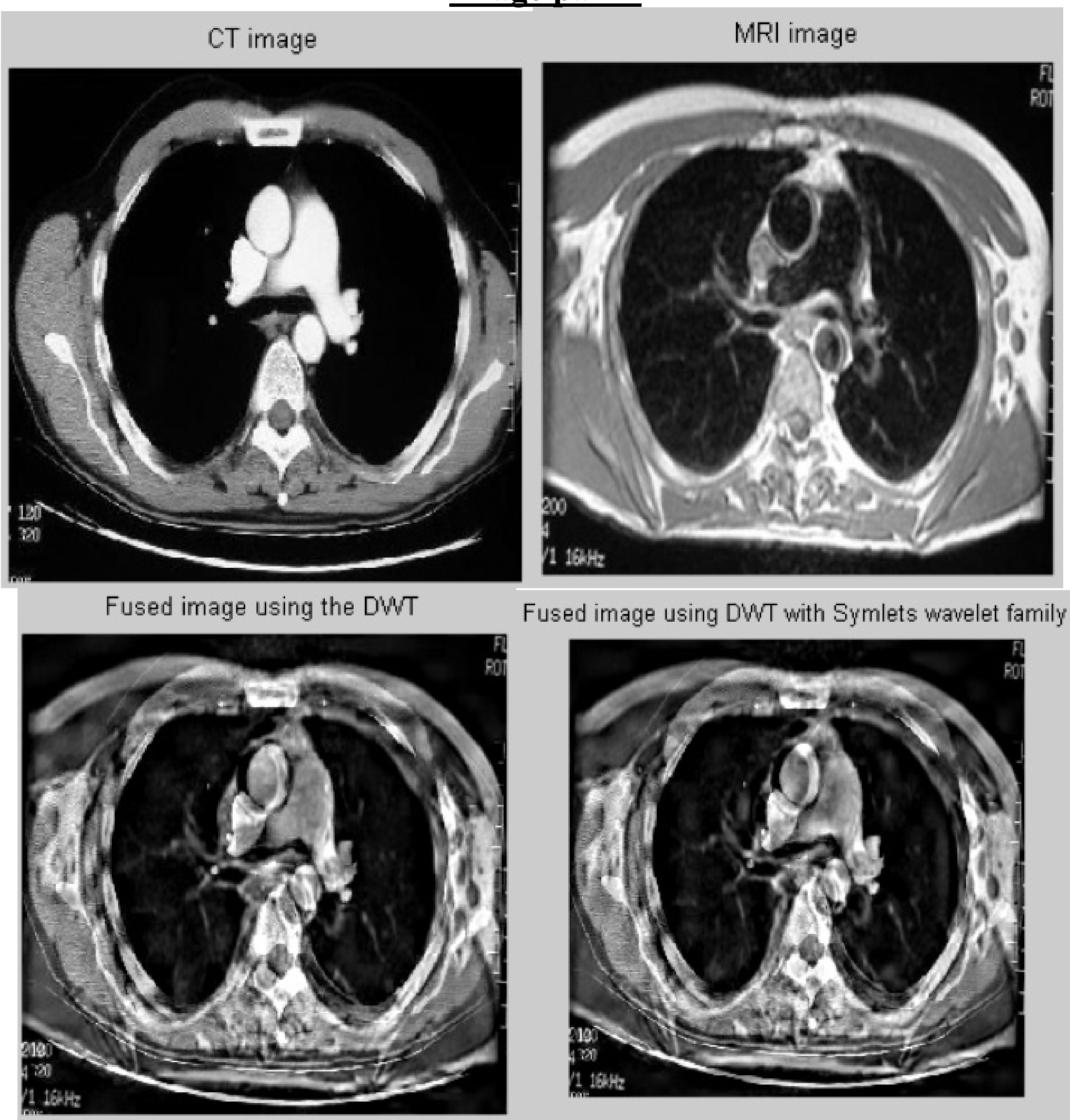
Qf=0.3896

Qf=0.6593

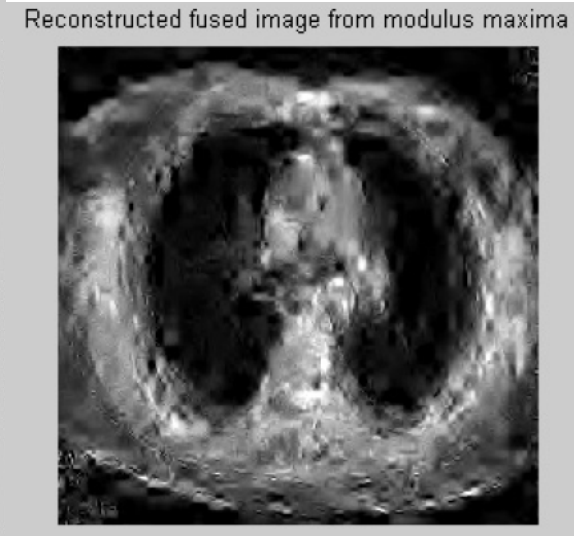
Qf=0.6620

Test 2 : Number of levels = 4

Image pair 1

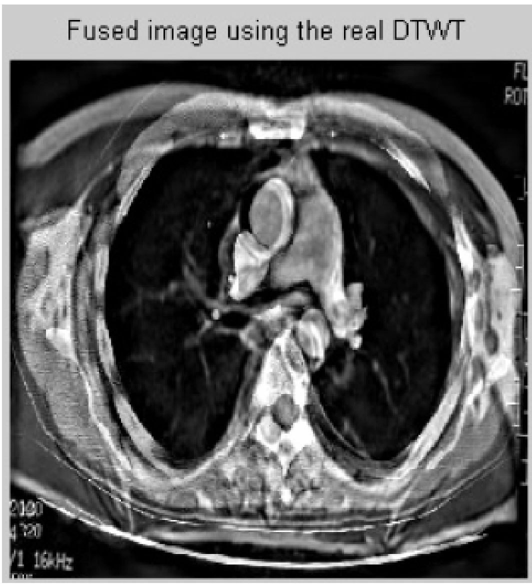


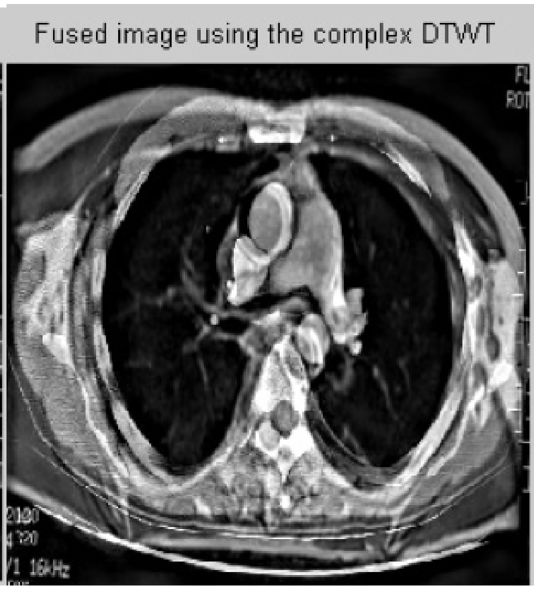
Qf=0.5282 Qf=0.5631



Qf=0.6270

Qf=0.2234

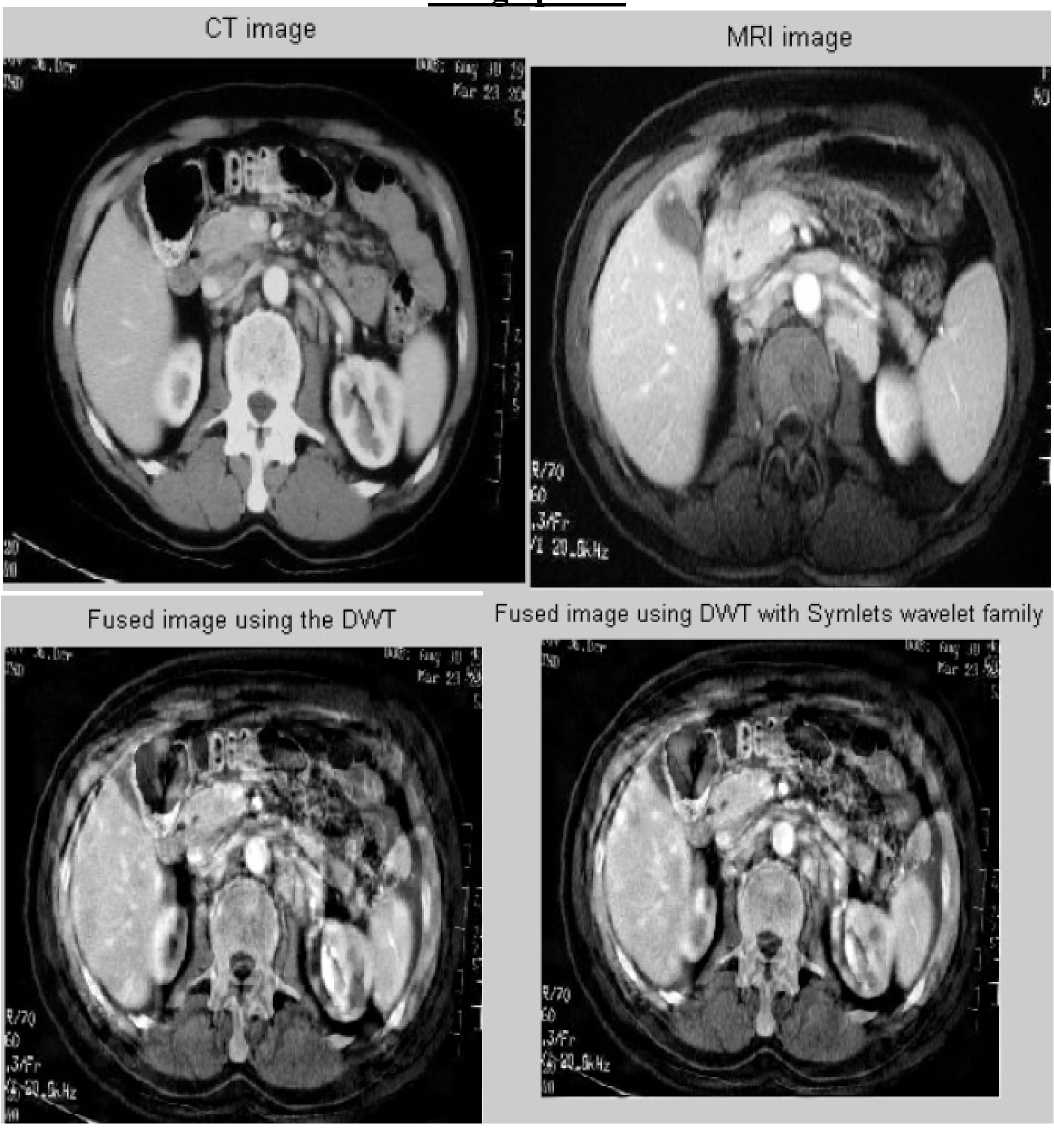




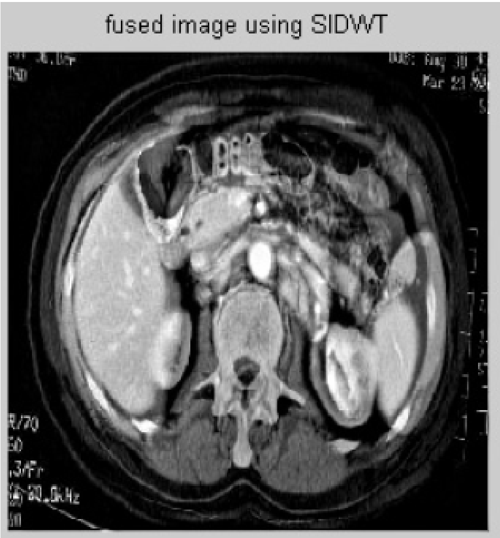
Qf=0.6033

Qf=0.6178

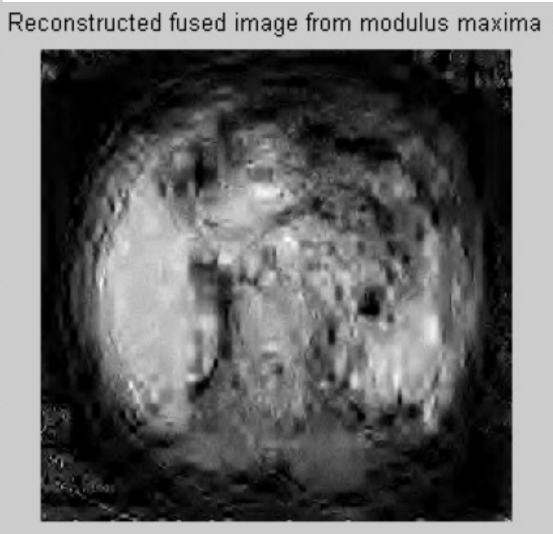
Image pair 2



Qf=0.5325 Qf=0.5786

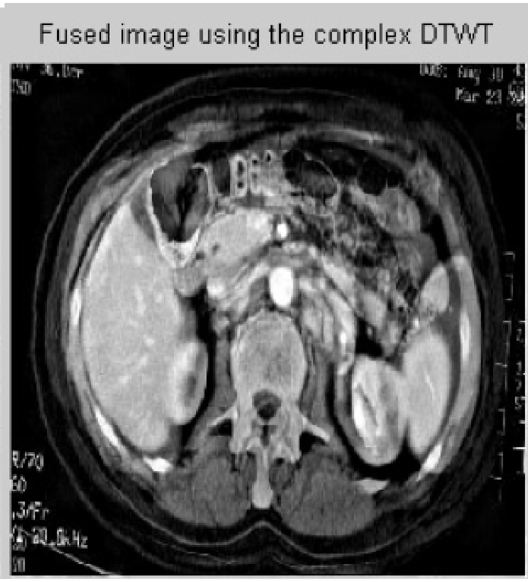


Qf=0.6343



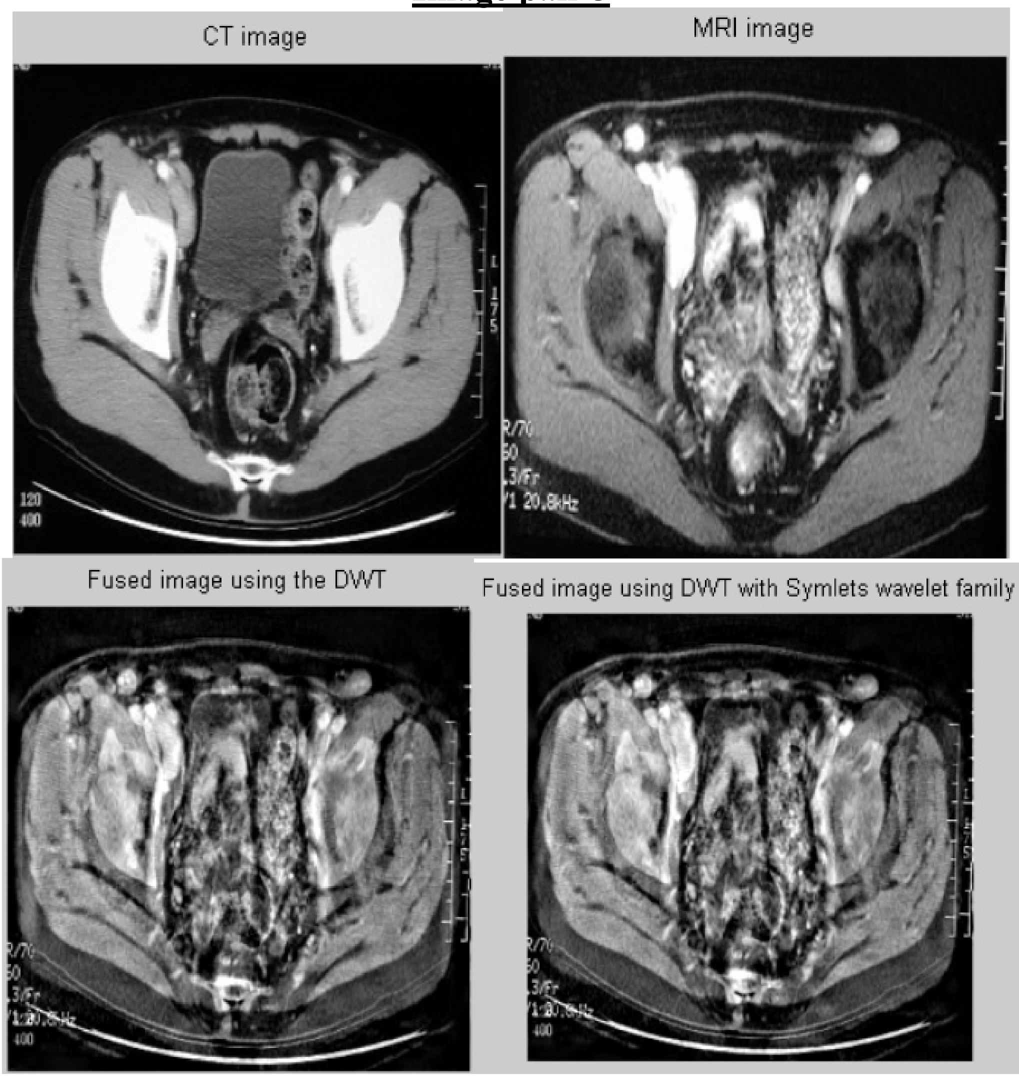
Qf=0.2416

Qf=0.6097

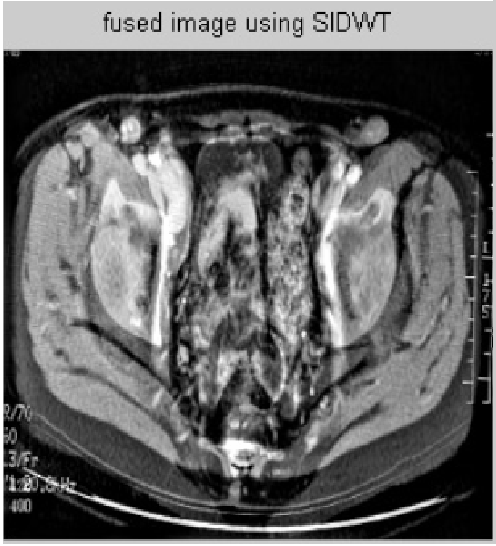


Qf=0.6205

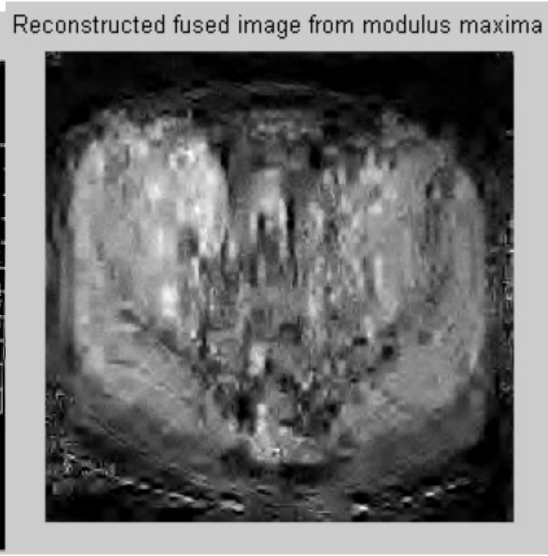
Image pair 3



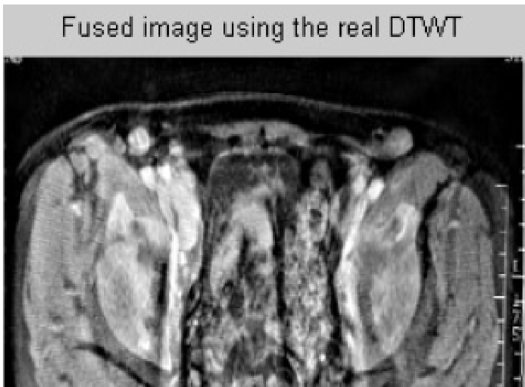
Qf=0.5114 Qf=0.6288



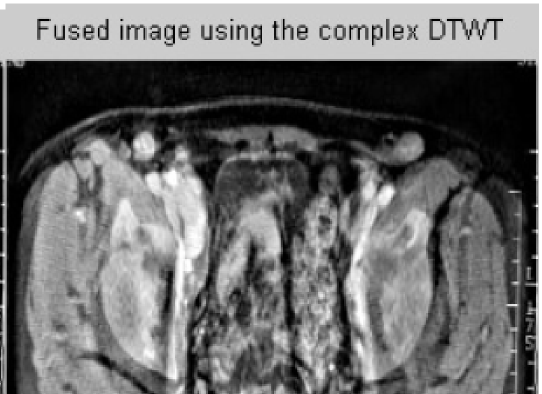
Qf=0.6868



Qf=0.2270

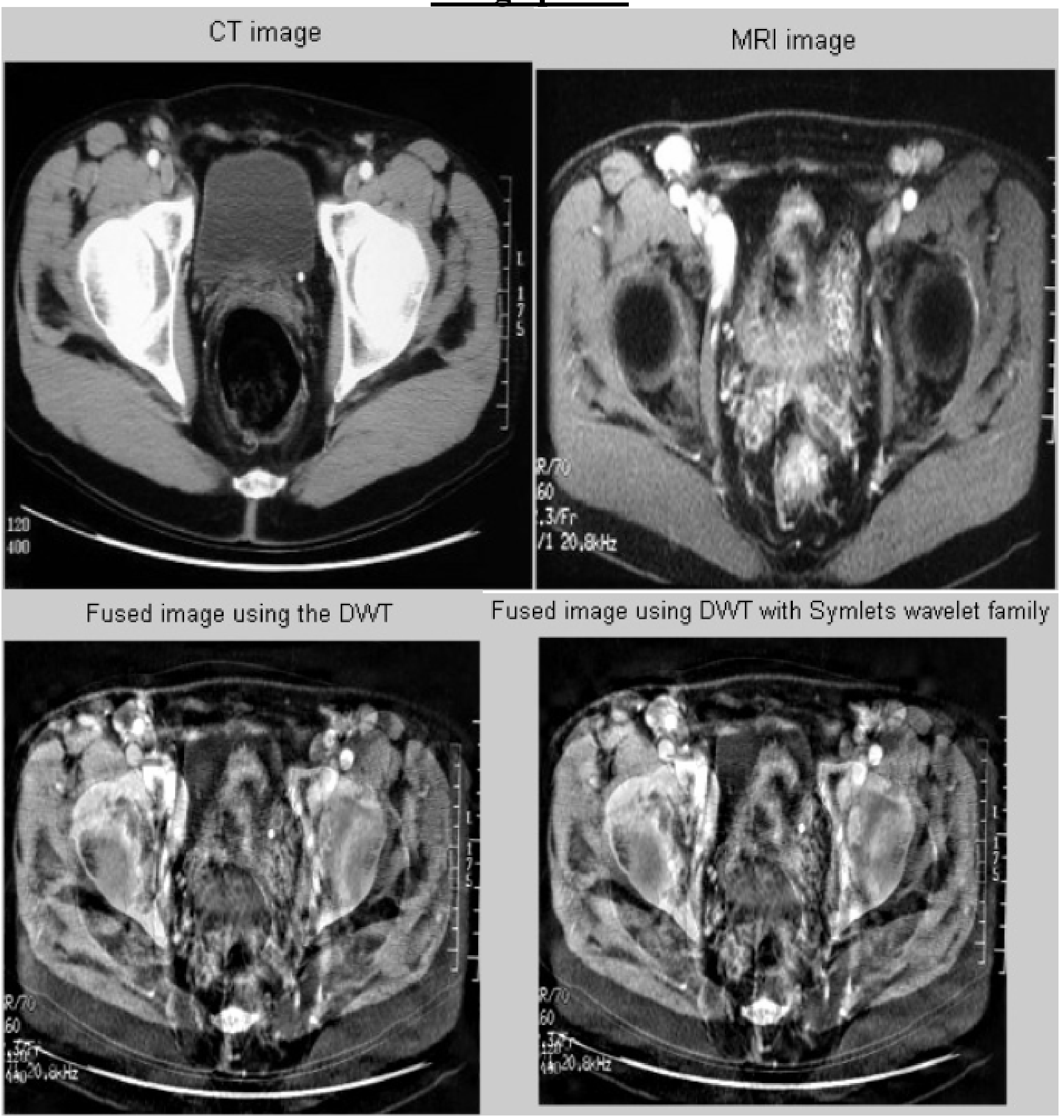


Qf=0.6616



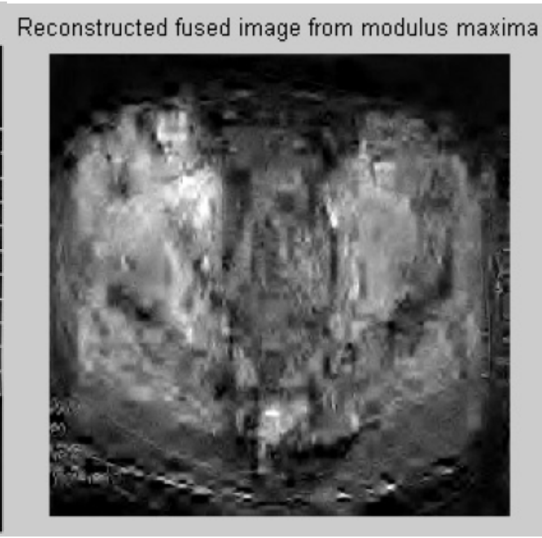
Qf=0.6754

Image pair 4



Qf=0.5114 Qf=0.6124

Qf=0.6888



Qf=0.2176

Qf=0.6545

Qf=0.6700

Although some conclusions could have been visually extracted, the summarization of the quantitative results will give us a clearer view of the image fusion evaluation problem

“ Overall comparison ”

Method	Number of levels	Quality index Q(a,b,f) Image pair Pair 1	Pair 2	Pair 3	Pair 4
DWT using symmetric orthogonal filters	**2**	**0.5813**	**0.5740**	**0.5549**	**0.5504**
	4	**0.5282**	**0.5325**	**0.5114**	**0.5114**
DWT using filters derived from the Symlets wavelet family	**2**	**0.6325**	**0.5955**	**0.6621**	**0.6465**
	4	**0.5631**	**0.5786**	**0.6288**	**0.6124**
SIDWT	**2**	**0.6714**	**0.6287**	**0.7022**	**0.6870**
	4	**0.6270**	**0.6343**	**0.6868**	**0.6888**
Real DTWT	**2**	**0.6466**	**0.6056**	**0.6742**	**0.6593**
	4	**0.6033**	**0.6097**	**0.6616**	**0.6545**
Complex DTWT	**2**	**0.6490**	**0.6085**	**0.6773**	**0.6620**
	4	**0.6178**	**0.6205**	**0.6754**	**0.6700**
Modulus Maxima reconstruction	**2**	**0.4280**	**0.3958**	**0.4017**	**0.3896**
	4	**0.2234**	**0.2416**	**0.2270**	**0.2176**

Chapter 5

Conclusion

Taking a look at the table above we notice that the SIDWT outperforms the other image fusion algorithms based on wavelet analysis. This was not surprising as it evolved as an improved version of the Discrete Wavelet Transform, which constitutes the basis for wavelet analysis techniques. On the contrary, the technique based on image edge representation through the modulus maxima calculation was proved poor enough over the third level of decomposition. These two conclusions were confirmed from all the input image pairs. Another very important issue is that the Complex Dual Tree Wavelet Transform is the only technique that has increased performance as the number of decomposition levels gets higher! In general, the performance of fusion process deteriorates rapidly as the number of levels increases.

The DWT fusion method provides a computationally efficient image fusion technique. It is simple and fast enough and can be easily expanded to different “versions”. Its performance is efficient, qualitatively, as a visual result and quantitatively, as a quality measure result. Its main disadvantage is the presence of shift-variance and the poor directionality.

The SIDWT fusion method is definitely an improved expansion of the DWT offering better performance results with the cost of inserting more computational complexity, as the amount of data to be processed is not decreased between the different decomposition levels (subsampling is discarded). Comparing to the DWT method the complexity is of factor O ($\log N^2$) while the regular Discrete Wavelet Transform has a complexity of factor O ($\log N$). The performance is improved contrary to the DWT; the new transform is now shift-invariant but still suffers from poor directionality. However, as mentioned, it suffers from excessive redundancy.

The Dual-Tree Discrete Wavelet Transform comes to overcome the limitations of the two previous techniques. It is shift-invariant, offers good directionality as the two diagonal orientations are now present in the fused result, it is non-redundant as the data to be processed are subsampled but has a more complex structure. As the number of decomposition levels is increased this transform outperforms all the other, so it is highly recommended for a large number of analysis stages.

Finally, the fusion method using the edge representation of the input images proved to be the poorest image fusion technique, especially after the third decomposition level. The main difference, comparing to the other methods presented, is that this is a feature-level based wavelet fusion method, which combines the high-level sparse representations of the input images, in the form of multiscale edges (wavelet transform modulus maxima) or chains of such edge points in order to fuse the images. The advantage of this technique is that the sharp characteristics of the initial images are preserved in the final image in a better way. In addition, by suitably thresholding the edges of the source images it is possible to control the edge information to be retained in the fused image, especially in cases where there is some pre-knowledge about the importance of certain types of edges [9].

However, it is not an easy part to completely determine which algorithm is optimum. The evaluation comparison is deeply application dependant and not totally objective. The same algorithm can perform in a different way when applied to a different input or when its parameters are altered. For example, the use of different filter structure (e.g. the size of the filters or the wavelet family from which they are derived), or the increment of the decomposition levels may deteriorate or improve the performance of a method contrary to one other. For example, the DT-CWT overcomes the other transforms after the fifth decomposition level.

Taking all these into consideration, we conclude that the image fusion problem is a very complicated and "thin" issue. In this report, most of the basic wavelet transforms were presented and deeply analyzed and the important issues on image fusion were discussed. Having introduced the theoretical background it is then up to the researcher to decide the most efficient way to apply theory in practice according to the needs of his/her application.

Appendix

A. Categorization of fusion methods

Depending on the level of description in which information is fused, image fusion takes place in the following levels:

- **pixel-level,** also known as picture or data or signal level. Fusion of images in this level is the process of combining two or more spatially registered pictures in an "enriched" one. Images are described in the spatial or frequency domain. The characteristics in each separate initial source image should be preserved or enriched in the fused image and any artifact should be avoided. Multiresolution techniques have a particular application in this category.
- **feature level** .In this level the initial images are described by edge or region maps, shape feature values, fuzzy measures, probabilities e.t.c. It requires algorithms in order to recognize objects based mostly on the statistical characteristics of dimension, shape, edges, and regions. Segmentation algorithms have been proved useful.
- **decision level,** also known as symbol level. It separately processes the entry images in order to derive information and applies decision rules to achieve common interpretation and remove the differences.

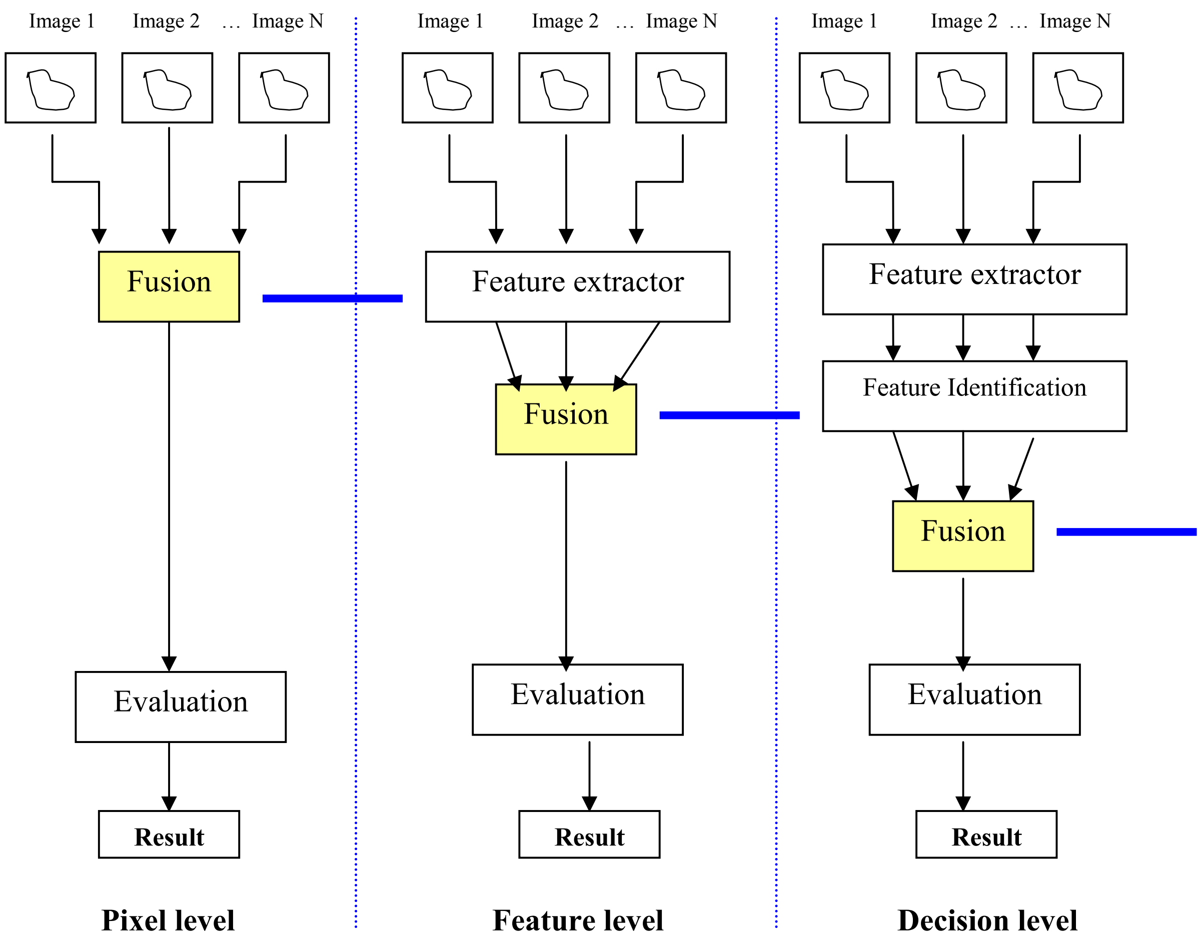
Image fusion levels

<table>
<tr><th colspan="2">Level</th><th>Description</th></tr>
<tr><td colspan="2">Pixel level
(Data or image or signal level)</td><td>The lowest of the three levels, description of the information in its raw form.The aim is the representation of the visual information provided by the input images in a fused image without distortion or loss of the original information.</td></tr>
<tr><td rowspan="2">Feature</td><td>intermediate</td><td rowspan="2">Demands algorithms capable to recognize objects in the various,different sources of information,based on statistical features of dimension,shape,edges,or regions.Segmentation methods have been proved useful for this category</td></tr>
<tr><td>feature</td></tr>
<tr><td colspan="2">Decision level
(Symbol level)</td><td>Processes the input images seperately in order to derive information and applies decision rules to achieve common representation and remove differences and inconsistencies</td></tr>
</table>

The basic categorization of methods of image fusion that is followed in this text is done depending on the three levels of description in which the information is fused, that is

- Pixel level
- Feature level
- Decision level

A schematic representation of the fusion process in the three levels above is the following:



We can notice the difference of the levels. They have the same structure but as the level gets higher one more abstraction-processing segment is added.

A detailed correspondence of most known methods to one of the levels above and their general description follows, without focusing so much in the mathematic background and the implementation algorithm

A.1 Pixel-level Image Fusion (also known as Image Merging)

This category is the most popular and most widely adopted by the researchers and the implementators. Fusion of images in pixel-level refers to the process of combination of two or more spatially registered images in an enriched, fused image. The combination/fusion takes place in the level of pixel description .The real time description is a new image that is represented by a spatial array, mostly containing values by applying nymerical or statistical transformations to pixel densities of the source images. When fusion takes place in the frequency domain an inverse transformation is required. Fusion operators are usually averaging or selective while the fusion rule may be different [3].

Up to day, fusion of images in pixel level is used mainly in order to create enriched, fused images with characteristics for human observers.

The general requirements of this category are summarized in the following:

- **Pattern conservation:** The relevant, basic information in the initial input image should be preserved in the final output image too. No loss of "crucial" data is allowed.
- **Minimal inconsistencies and artifacts:** The implementation schema should not insert additional, non-existent or inconsistent elements that will cause confusion in the observer or in the next processing stages.
- **Shift and rotational invariance:** The fusion algorithm should be invariable to the change of position and rotation of the image or some object, always giving the same results.
- **Robustness to registration errors:** The algorithm should not be sensitive to errors in the position and the one-to-one matching of the pixels of the input images.

The most common drawbacks of this category are the facts that:

1. The output image suffers from blurring effects due to filtering and artifacts.
2. The pixel-level based techniques are very sensitive to noise.
3. There is sensitivity to imperfections in the source images and the sensors that provide the information.

Most image fusion methods are based on multiresolution analysis. Other methods include IHS fusion and fusion based on probabilities. In the following table the categories of fusion methods in pixel level and their general explanation are summarized.

A more analytic approach and description of multiresolution fusion and multispectral techniques follows.

Categorization of pixel-level image fusion methods

Category	Methods	Description
Linear superposition	PCA analysis,IHS ,RGB method,averaging	Probably the most direct way to fuse images.Implements fusion as a linear combination weighted by all input images.
Non linear methods	Bayesian model	One more simple approach by using a simple non-linear operator e.g. min,max or a morphological operator or,to another aspect,image algebra
Optimization approaches	Hidden Markov trees methods	Image fusion is expressed as a Bayesian optimization problem, using a-priory model of the fusion result in order to find the image that optimizes the a-posteriori possibility
Artificial neural networks	Newman and Hartline approach	Inspired by fusion of signals taken by sensors in biological systems, researchers established neural networks to pixel-level image fusion
Image pyramids	Gradient, difference, morphological, averaging, contrast, ratio pyramids	Image pyramids are a sequence of images where each image is constructed by lowpass filtering and downsampling of its "ancestor" image
Wavelet transform	MWD 2DWT	Relevant to the category above with the difference that this transform leads to non-redundant representation of the image having as a disadvantage its shift dependency
Multiresolution fusion scheme	**Includes the 2 categories above,** Toet's method, Burt's method, Wilson's method, Yocky's method, Region-based method	The basic idea lies to the fact that the human vision system is especially sensitive to local contrast changes e.g.edges.The fused image arises as a combination of multiscale edge representation

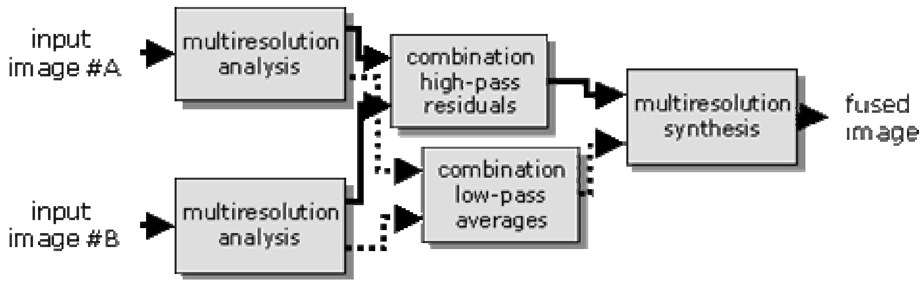
A.1.1Multiresolution image fusion

A.1.1.1 Introduction to multiresolution analysis

Undeniably the most important method for pixel-level image fusion is the last of those mentioned before, multiresolution image fusion, a hyper-generalized technique that includes, among the others, the method of pyramids and its “relative” approach, the wavelet analysis. The hierarchical fusion model based on multiresolution decomposition and reconstruction of images, has become a subject of research for the last 20 years.

The basic idea lies in the fact that the human visual system is especially sensitive to local contrast changes e.g. edges. Motivated by this conscience and having in mind that both image pyramids and wavelet analysis lead to a multiresolution edge representation we are immediately lead to the generation of the final, fused image as a combined/fused multiscale edge representation. The “building”-“construction” process is summarized below:

In the beginning, the input images are decomposed into the multiscale edge representation using either a pyramid technique or a wavelet transform. Two kind of information are generated: the high-pass information, which contains the detailed parts of the source image and the low-pass information. The real fusion process takes place in the relevant domain, where the final multiscale representation is created by a pixel-to-pixel selection of the factors, both of the high-pass and low-pass coefficients, having the maximum value. Ultimately, the fused image is calculated by application of the suitable reconstruction schema [23]. The general multiresolution fusion scheme is the following



General multiresolution image fusion scheme

A subset of the approach above is the fusion of images that is based on various pyramid types, which differ in the type and in the rules for the selection of the characteristics-coefficients. The decomposition of image pyramids can be implemented by using Gaussian, morphological filters and the 2-D discrete Wavelet Transform (2DWT). The pyramidical approach is described in the following pages.

The image fusion schema that is based on pyramid techniques can be considered as an approach of pattern selection since an image can be described via a sequence of levels. The characteristics are presented only in a given range of levels. Combining the characteristics of these different levels obtained by each input image using concrete criteria, the most important characteristics are presented in the final, unified image. This process can be valued in the three following stages [24]:

- **stage of decomposition**, in which the pyramid description is constructed from each of the spatially registered source images
- **stage of fusion**, where the description for the fused image is constructed by selecting values from the corresponding nodes in the component pyramids. The selection rule depends on the reliability of sensors, the performance measure and the values of intermediary nodes and the kind of application.
- **stage of reconstruction**, where the final picture is recovered from its pyramid representation through the corresponding reconstruction procedure.

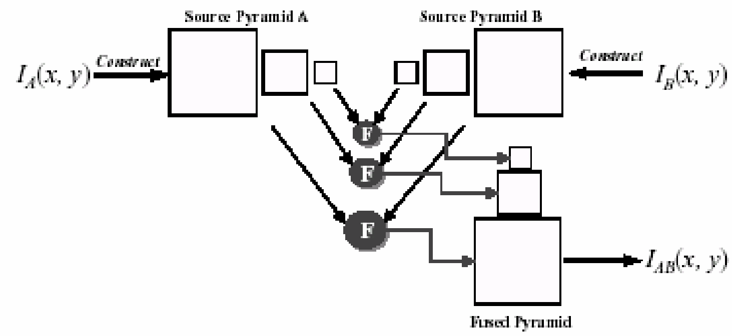


Image fusion based on image pyramids

The construction of multiresolution description initially requires that the source images acquire the same size, using a certain method, if they haven't already done so. Then the image representation as pyramid in levels is created. The initial picture constitutes the minimum level of pyramid, the "beginning point". Each node of pyramid is acquired by sampling the filtered "version" of the previous level. Thus for each of the N levels the following equation stands

$$1 \le k \le N$$
$$P_k = \text{REDUCE}\ (P_{k-1})$$

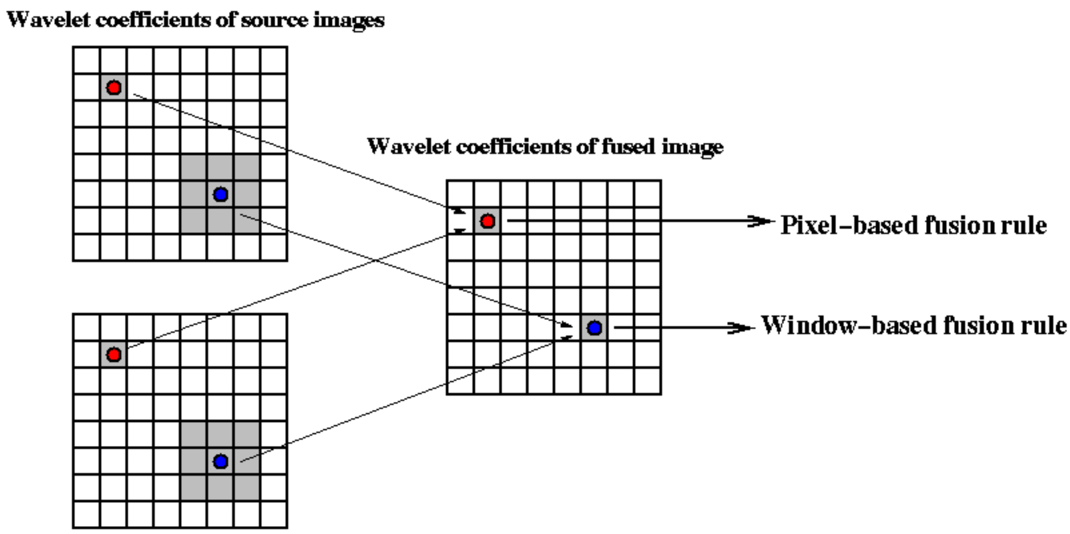
where the REDUCE operation is a filter of reduction of the pixel density and the resolution-size of the image pyramid (usually by a factor of 2).

The pyramids convert the local characteristics of the images to general ones. Each level maintains the characteristics of the initial image in a concrete degree. The initial image can be recovered by inverting the procedure above. For the construction and reconstruction of the pyramid levels sampling (downsampling/upsampling respectively) is required so that redundancy is removed, having the cost however to insert undesirable high frequency noise due to aliasing. Reducing the number of decompositions levels eliminates the processing redundancy, which means less complexity. Having fewer levels also means that less sub-band images are created, which reduces the possibility of errors in the reconstruction process, as fewer discontinuities are inserted during fusion. However, fewer levels also mean fewer detail data extracted from the sub-images.

The construction of the fused image pyramid is based on the following three basic fusion/selection rules:

- **sample-based**, where the comparison of coefficients in the corresponding positions takes place without taking into consideration the coefficients in the neighbor pixels, with the cost of the sensitivity to noise. Only one pixel, the corresponding pixel to the particular position, from each image is compared and participates in the claculation of the corresponding pixel value in the fused image.
- **window-based**, where the comparison of coefficients in the corresponding position takes place with comparison of 3x3, 5x5 windows centered in these positions.
- **region-based**, where decisions are taken based on the object in which the pixel belongs to.It is the most general set of fusion rules, as each pixel is corresponded to a large set of pixels that form an object in the source image.

The fusion rules above are more clearly explained in the following figure



Schematic explanation of the two basic fusion rules

A.1.1.2 Description of multiresolution fusion methods

The most common multiresolution techniques found in bibliography are the following:

1. Maximum Contrast Selection or Toet's method (contrast pyramids)

This method is based on the fact that the human vision system is sensitive to local luminance contrast. According to Toet's analysis the ratio of lowpass images in successive levels of the Gaussian pyramid is calculated. As a result a luminance contrast pyramid is created having the property that each of its level is the ratio of two successive levels of the Gaussian pyramid of decreased size according to the relation

$$R_k = P_k \,/\, EXPAND(P_{k+1}),\ \ l = 0,1,\cdots,N-1,\ \text{and } R_N = P_N . \tag{1}$$

where the EXPAND (P_{k+1}) function interpolates the image-subimage to the higher (k+1) level by filtering with a suitable "construction" filter and upsampling and R_k, Pk are the contrast and the Gaussian pyramid respectively.

The contrast pyramid is a complete representation of the initial image, which is reconstructed immediately by inversing the steps of pyramid building

$$P_N = R_N,\ \text{and } P_k = R_k \cdot EXPAND(P_{k+1}),\ \text{for } k = N-1, N-2,\cdots,1,0. \tag{2}$$

When two images A and B are fused, the ratio of their pyramids is constructed and the final, fused pyramid is built by applying a selection criterion based on the maximum absolute value of contrast

$$R_k^F(x,y) = \begin{cases} R_k^A(x,y), & \text{if } \left|R_k^A(x,y)-1\right| > \left|R_k^B(x,y)-1\right| \\ R_k^B(x,y) & \text{otherwise} \end{cases} \tag{3}$$

where Rk(x, y)s are the fused pyramids, the pyramid of image A and B respectively in level k

The final image is reconstructed by its pyramid according to the second relation.

Conclusion: This method manages to offer more information for observation by selecting details of the maximum contrast but does not take into consideration the fact that these contrasts appear in noisy images. As a result the method chooses the points that have been affected more by noise; hence it loses information about the corresponding desirable sections.

2. Salient pattern selection or Burt's method

Contrary to the method above Burt used other selection metrics in order to make the details more obvious and distinct. Thus, at the points where the images differ obviously the most clear-sighted patterns are selected by the pyramids while the less obvious are rejected. At the same time, at the points where the images are similar the patterns are calculated on average, which contributes to the removing of noise and to the existence of stability at the points where the initial images have similar characteristics. The degree in which we determine how clear-sighted a pattern is depends on its contribution to the representation of scene information. Thus, a pattern has a great distinctness degree if it is relatively noticeable by the eye in the initial image," it seperates". Various mathematic models are used for this determination.

Conclusion: This method manages to decrease the effect of noise and preserve the details of low contrast if these are the "distinct characteristics". The main problem however is that the metrics are calculated in absolute values for pyramids. Moreover the determination of the size of the window to be applied for the neighboring contour is ambiguous, since windows of small size do not eliminate the noise and windows of big size tend to blur the patterns.

3. Perceptually Salient pattern selection or Wilson's method

This method constitutes an improved extension of the previous method taking into consideration the frequency response of the human optical system. It avoids the case where the energy of images that are not perceptual to the human eye becomes dominant in the image fusion rule. Thus, which part of the input images will be calculated by averaging or not, in order to eliminate noise, is determined by the sensitivity in the analyst's contrast. The metrics for the determination of distictness are redesigned including the perceptible energy in a given set of the information-detail.

4. Yocky's Mwd-based image fusion

This method creates a general background for MWD (Multiresolution Wavelet Decomposition) image fusion. According to Yocky, each level of an N-dimensional pyramid built by MWD consists of three detailed images in this level of resolution with the highest level being the N-th approach in the wavelet decomposition. In each level of the resolution pyramid the images can be evaluated and manipulated how they are being reconstructed. Each stage of reconstruction can include various techniques such as conservation of energy, edge enhancement, contrast stretching e.t.c. and the wavelet coefficients can be arithmetically or logically manipulated as it happens in the simple case of wavelet substitution and reconstruction.

The most common case of application of this method is reported for image fusion of two different sensors: one that provides data of high spatial resolution, having the cost of high spectrum bandwidth and a second one that provides low-resolution multi-spectra data, having the cost of the spatial resolution. Thus we have a combination of panchromatic data and low-resolution multi-spectral data. The MWD is used in order to decompose the two categories of images above in three different levels: 1/2, 1/4 and 1/8. Then new pyramids are manufactured for each color band importing its 1/8 color approach in the 1/8 panchromatic pyramid. Inverse transformation is applied in each pyramid in order to receive the final, fused image.

5. Additive MWD image fusion or Nunez method

Nunez proposed that instead of fusing each colored element we should convert the multi-spectra image into the hue-intensity-saturation (HIS) space and fuse the intensity component only with the panchromatic image. The decomposition of the input images in wavelet level is achieved using the "a trous" Discrete Wavelet Transform. Each pyramid level is computed as the difference between two consecutive approximations. All levels of the pyramid should be summed up as well as the residual in order to recover the image. It should be marked that the pyramid built by this way has the same size at each level since this algorithm is non-orthogonal oversampled transform. The new intensity values are converted again into the RGB model with" old" elements H and S.

Conclusion: Nunez supports that by this additive method the detailed information from both images is used and since the difference pyramids have zero mean, the total

flux of the multi-spectral image is preserved. However the fact that the components are added leads to the effect that intensity exceeds the range of gray-levels and the result should be re-scaled with the cost of color distortion.
(For more information about the methods above refer to [3])

6. Region-based MWD fusion

This method is based on the fact that each pixel belongs to an object. Thus the decision-making on each coefficient should also take into consideration the wider region they represent. Features such as edges and regions are guiding elements for image fusion that takes place in the following stages:

1. Edge detection using an operator leading to the generation of edge images that provide information about the location and the intensity of edges in the initial images.
2. Segmentation of regions based on the information above.
3. The activity of regions is calculated by averaging the high-resolution wavelet coefficients. Bigger values of activity mean that the region contains more information [5].

7. Gradient-based image fusion

The information is presented in the gradient maps domain. Contrary to the coefficients of wavelet pyramids, whose size is only a clue of the distinctness of features within a pixel neighborhood, the absolute size of the elements of gradient maps is a precise, direct, spatial indication of the feature contrast. Moreover the elements of maps contain information from all spectra which adds reliability to the selection and the fusion of elements. That's why this method achieves important restriction to the distortion of the fused visual information [25].

8. Image pyramids

Generally, an image pyramid is a set of simplified versions of the initial source image in which its size is decreased. As a result different levels-pyramid parts are created. Each level inludes the **low-pass or approximation** pyramid and **the high-pass or detail** pyramid. In the low-pass pyramid, the lowest level (also mentioned as zero level) image is the original source image X and each of the following level images $X_{(k)}$, $(k>0)$, is built by filtering and sub-sampling of the previous level image $X_{(k-1)}$. Then, the high-pass pyramid at each level is built by interpolation of image $X_{(k)}$ and the resulting image $Y_{(k)}$ is created by substraction of $X_{(k)}$ from its predecessor $X_{(k-1)}$. The resulting image $Y_{(k)}$ at each level contains only the information lost from one level to the other, does not keep useless data.

Usually, this method is implemented as a process of features selection that transfers the most important coefficients of the input pyramids to the pyramid of the final image, which is converted to the fused image via the reconstruction procedure [26].The advantage of this method is the fact that:

1. it can provide information on the sharp contrast changes of the images, which is directly noticeable to hyman eye
2. it can provide both spatial and frequency dimain localization

The most common image pyramid schemes are shortly inttroduced below:

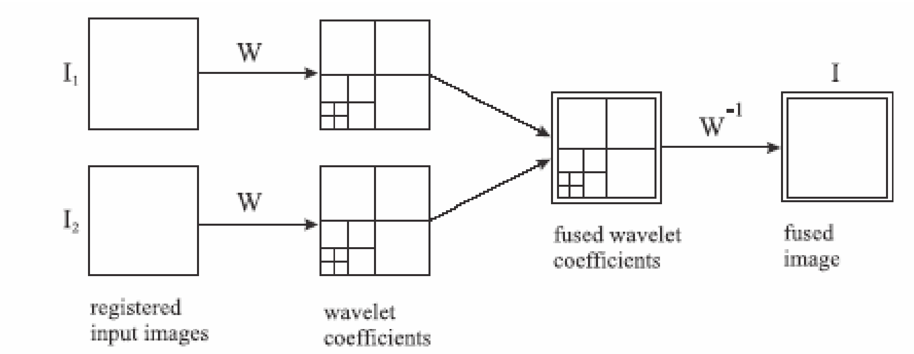
- **Gaussian and Laplacian pyramids:** a Gaussian pyramid is a sequence of images in which each image is a filtered and downsampled copy of its predecessor one. Due to sampling, the size of the image is subdoubled in both spatial dimensions in each stage of the decomposition procedure, leading to a multiresolution signal representation. The difference between the input image and the filtered one is essential so as a precise reconstruction by the representation of pyramid to be allowed. This approach leads to a representation of signal with two pyramids: the smoothing pyramid that contains the average pixel values and the difference pyramid that contains the differences in pixels e.g. edges. The difference pyramid can be considered as a multiresolution edge representation of the entry images. The procedure of the decomposition of the initial image to pyramid structure is named REDUCE and the process of interpolation of an image of higher level in the pyramid is named EXPAND. The successive application of the REDUCE operation in an image leads to the smoothed pyramid (Gaussian pyramid) and the application of Gaussian filtering followed by an abstraction process leads to the difference pyramid(laplacian pyramid).The image is reconstructed by the Laplacian pyramid inversing the steps above

- .**Gradient pyramid:** the process above is followed with the difference that one other mathematic background (different values of the filters applied) is used. The reconstruction is implemented by converting the Gradient pyramid to Laplacian and the Laplacian to Gaussian pyramid.
- **Morphological pyramid:** it differs as regards the mathematic calculations for the pyramid building. Morphological filters are used where the intensities of objects are preserved, thus the appreciated position of the contours remains unchanged. This filters abstract details from the image (foreground and background) which are less relevant to each structural element and are suitable for shape extraction
- .**Contrast pyramid:** it is built by dividing each level of the Gaussian lowpass pyramid with the extended "version" of the next level. Each level of the pyramid contains information only about the corresponding level of the Gaussian pyramid. Moreover, the size of the pyramid coefficients is a measure of the contrast and the fusion of the final pyramid is achieved by the selection of the coefficients that correspond to the values of maximum contrast in each position of the pyramid.

9. Wavelets

It is a mathematic tool that was initially developed for signal processing, however it can be used in the frames of MRA (multi-resolution analysis). The wavelet transform generates a sequence of elementary functions (wavelets) from arbitrary functions of finite energy. The weights assigned to wavelets are the wavelet coefficients, which play an important role in the determination of the structural features in a given scale and position [9].

The basic schema is the following:



- **Discrete Wavelet Transform:** it is successfully applied to image fusion applications based on DWT decomposition using Mallat's implementation algorithm. The DWT pyramids are fused using metrics of energy calculated for each pixel such as the maximum absolute coefficient value in a surrounding region of neighboring pixels or simply using the sample-based rule. It produces a non-redundant image representation, a fact that offers better spatial and spectral localization of information in relation to other multiresolution representations. However it introduces errors to reconstruction and ringing artifacts when changes and discontinuities are imported in the subband coefficients values.
- **Dual-Tree Complex Wavelet Transform (DT-CWT):**it applies repetitive different spatial filters in order to produce frequency subbands in a similar way as the classic discrete wavelet transform.The number of subbands is twofold in this case.
- **Discrete Dyadic Wavelet Transform (DDWT)**
- **UDWT:**

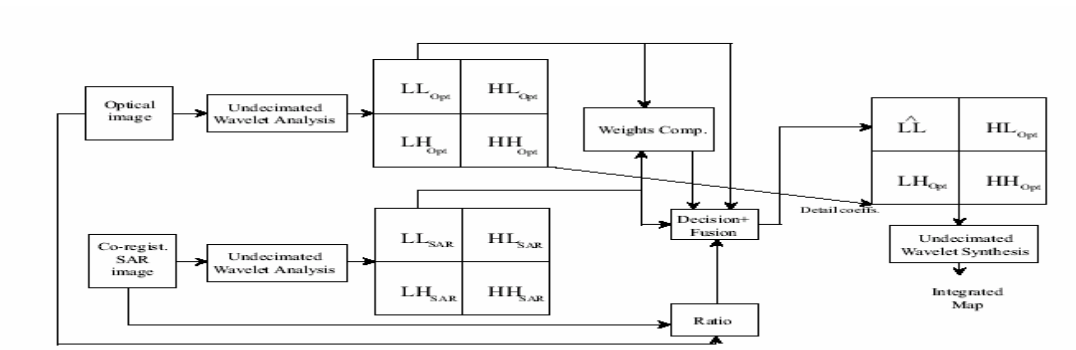
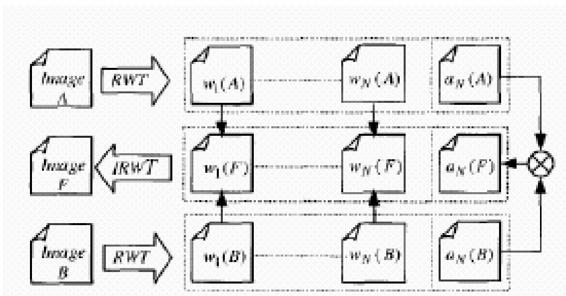


Figure 2. Scheme of M-UDWT.

- **Redundant Wavelet Transform:**



A.1.1.3.Conclusions for multiresolution image fusion

Summary:

Certain questions immediately originated about the multiresolution techniques are:

- Which pyramid technique is better for the image fusion problem?
- To what resolution level should images be decomposed?
- Which fusion rule is preferred?
- How the algorithm performance is evaluated? Is there any objective criterion for all applications?

The answer cannot be absolute! Everything depends on the application and the attributes of the entry images. Concepts such as stability of techniques to imperfections of the entry images (such as misregistration) or artifacts in reconstruction where redundant, non-relevant information is imported "from nowhere" during reconstruction should be seriously taken into consideration.

Advantages of multiresolution scheme:

- **Flexibility:** the information(the "scene") is described on different resolution scales. The number of levels is selected depending on the resolution requirement and the computation load.
- **Compactness:** the redundant information is removed decreasing the pixel intensity as long as the level is increased, preserving the memory requirements in reasonable bounds.
- **Feature conservation:** features both in fine and coarse scale are preserved in the final image.
- **Feature enhancement:** with suitable criteria the features of the entry images can be enhanced and become more recognizable by the human eye-observer.

Restrictions of multiresolution scheme:

- **Introduction of noise:** sampling, used for the pyramid "building" and the reconstruction of the final image has as a result the introduction of high frequency noise and interpolation due to aliasing.
- **Restriction on images size:** pyramid algorithms require certain image size. As a result zero-padding is needed, which introduces noise on the edge.
- **Preprocessing of source images:** Before the application of algorithms registration of the source images, sampling or histogram matching are required. Except for the increased computational activity, errors during the preprocessing stage can introduce noise and feature distortion on the fused image.

The above multiresolution techniques are summarized in the following table:

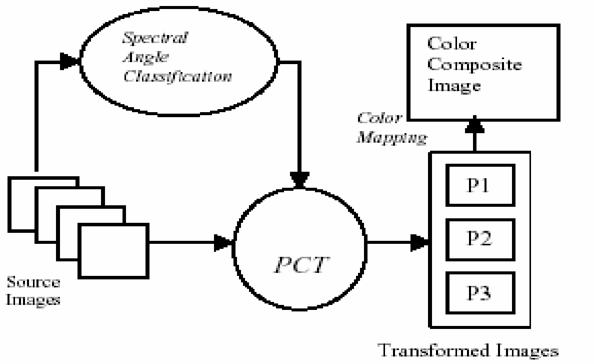
Multiresolution Analysis	**Image pyramids**	**-Gaussian** **-Laplacian** **-Contrast** **-Gradient** **-Morphological**
	Wavelets	**-DT-CWT** **-DDWT** **-UDWT** **-RWT**
	Other	**-Maximum Contrast Selection** **-Salient Patern Selection** **-Perceptually SPS** **-MWD-based** **-Additive MWD fusion** **-Region-based MWD fusion** **-Gradient-based fusion**

A.1.2. Other methods (Multispectral techniques)

Linear superposition

1. Principal Component Transform (PCT) or Principal Component analysis (PCA)

It summarizes and decorrelates the images by removing noise and redundancy and packing the residual information in small sets of images, termed principal components. These components are rank ordered by the magnitude of their variances (eigen values) [27].Therefore, most of the spectral contrast is pushed forward to the first few components with an increase in the signal-to-noise ratio of these components. The flow chart of the method is the following:



PCT-based fusion algorithm

2. HIS transformation

Another pixel-level fusion method is HIS fusion. It is used when a low resolution color image and a high resolution monochrome one are joined in order to create a color high resolution image. It is based on the transformation of the source images intensity components from the RGB model to the HIS and the replacement of intensity components with the monochrome image and the retransformation to the RGB color image as seen in the figure below:

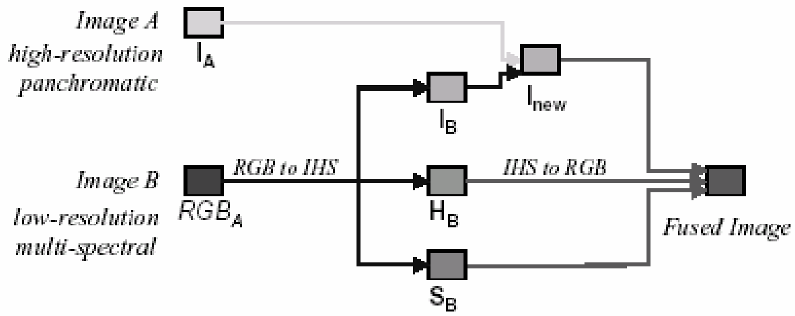


Figure 6. IHS merger.

Probabilistic image fusion

In the previous methods selection or averaging was mostly used to reconstruct the final image. In the probabilistic method the mapping from the true scene to a sensor image is modeled as a noisy, local affine transformation whose parameters are allowed to vary across the image[1].Taking into consideration the parameters of transformation probabilistic models are used, such as the Bayesian MAP and ML rules, in order to evaluate “the true scene” value. This model produces cleaner images when the source ones have been affected by noise. The restrictions of this method are summarized below:

- the sensor parameters are estimated based on local pixel values
- the image formation model is assumed to be linear and the noise is normal
- the evaluation of parameters is performed locally, thus for each new patch the fusion model should be recomputed, which increases computational complexity.

Performance evaluation for pixel-level algorithms

As already reported, the determination of metrics for the performance of the various algorithms is not absolute and objective. It depends on the applications, the source images, the sensors, the algorithms. The most acceptable method is the root mean square (RMS) error, which is defined as

$$\varepsilon_{RMS} = \sqrt{< (c-\hat{c})^2 >}$$

where c is the initial value and c^ is the value of the fused image. This quantity is calculated for each pixel and afterwards the mean value is selected.

This metrics offers facility in calculation and provides the average deviation from the ideal image. However, since the objective of pixel-level algorithms is to create images with enhanced features for human observer another statistical metrics is used, the Mutual Information, in order to compare the methods based on their faculty to preserve and enhance those features that are sensitive to be observed by human eye.

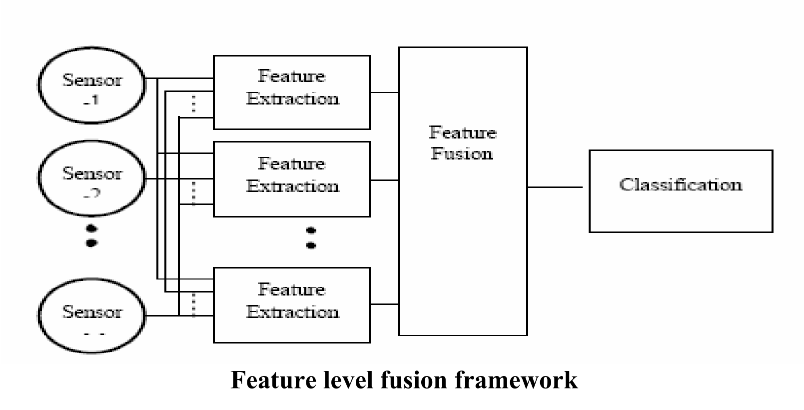
A.2. Feature-level image fusion

Feature level was created in order to overcome the drawbacks of pixel-level fusion methods such as blurring, sensitivity to noise and imperfection of the source images.The theoritical background under this level is that instead of fusing images based on their raw data form, pixel densities,we should be based on a more robust and clear data form, image features.They consitute a more abstract and less complex structure that is stable, capable of fully representing the source image and less possible to produce errors in the output image

A feature can be considered as an elementary or basic attribute in an image or in a region of an image. The first type of features is associated with the objects in a scene. Such features may describe the shape of the object, the texture or a combination of them. The second type of features is associated with distortions or abnormalities in the generation of image. The techniques of this category are relational to the field of pattern recognition and image segmentation. Raw image data will be transformed in the output into a suitable representation, which depends on the application and the method followed, such as image segments or signal amplitude or as shape, length, or orientation of objects in an image. Raw image data are transformed into features via certain mechanisms named as "feature extractors".The output of the extractors is the input of the feature level scheme, usually in the form of vectors containing values that uniquely describe features according to the method adopted and the application in which fusion takes place. .

Feature level fusion techniques require the extraction of objects from the image, either manually or automatically. These features can be detected using geometric or structural characteristics (e.g. edges), statistics or spectral properties. Features are extracted from the images using proportional techniques and are transmitted in classifiers for class recognition, which is the ultimate goal of most object recognition systems. The classifiers determine the status and the kind of the feature; they simply "decide what it is". "Is it a region, an edge or something else?" Features can be edge maps, lines, moments, Fourier descriptors, height-width ratios or any compact characterization that is suitable for the description of the problem.

The segmentation features include points of interest, edges and regions. Even if these two significances are considered as "dual", they refer to two different image properties: local differences and global homogeneity [28]. Their behavior is different on certain parts of the image due to the complexity of the real image and does not always coincide with the ideal model of edges-regions. For most compact features the classification can be performed on each set of features independently or on the set of all features.



Feature level fusion framework

Features are extracted from each of the sensor data, followed by a registration step, usually performed at the level of regions of interest or image segments containing more than one pixel. **Such a co-registration of features from individual sensors is often easier to achieve than pixel level fusion.** A detection/classification algorithm can then be applied on the combined feature vector characterising a region of a certain spatial extent. Feature level fusion involves extracting feature vector information and creating a single feature vector for identifying an object.

From all the above it becomes obvious that the feature extraction mechanism plays a very important role in the fusion process. The creation of a robust feature selection scheme and a sophiticated feature extraction technique becomes critical.

One method of achieving feature-level fusion is with a region-based fusion scheme. An image is initially segmented in some way to produce a set of regions. Various properties of these regions can be calculated and used to determine which features from which images are included in the fused image. **This has advantages over pixel-based methods as more intelligent fusion rules can be considered, based on actual features in the image, rather than on single or arbitrary groups of pixels.**

Advantages-objectives:

- Fusion at the feature level has many advantages, especially for medical application such as multi-modality tomography. In particular, there is no need to put reconstruction procedures or image format in a specific form; processing is facilitated at each modality unit; the amount of information to be transmitted to a central processor unit can be reduced. It is less sensitive to noise, less computationally complex and more intelligent fusion rules can be adopted.
- Feature level fusion, combines various features. Those features may come from several raw data sources (several sensors, different moments, etc.) or from the same raw data. In the latter case, the objective is to find relevant features amongs available features that might come from several feature extraction methods. The objective is to obtain a limited number of relevant features.

A synoptic presentation of feature-level image fusion methods follows. It should be marked that the existent bibliography for the techniques of this level is too much limited.

The following methods have been reported [27]:

- **parametric templates**
- **hierarchical clusters**
- **neural networks:** the most effective of the rest methods, especially in medical applications. It performs a nonlinear transformation between an input vector and an output feature. It is fed by multiple images taken on different planar views and produces an output image of clearer view and higher quality. By using a suitable training set the network produces the output from its associated input within seconds, after it has generated association function between the input and the output computing random interconnection weights. These functions will be used later for any input to produce the appropriate output. However high training level is required in order to achieve the transformation function.
- **knowledge based approaches:** an alternative neural network method that emulates the cognitive processes used by human. It emphasizes the use of certain production rules and computational logic. Unfortunately it requires a considerable high level of training

The following four methods have been more extensively reported:

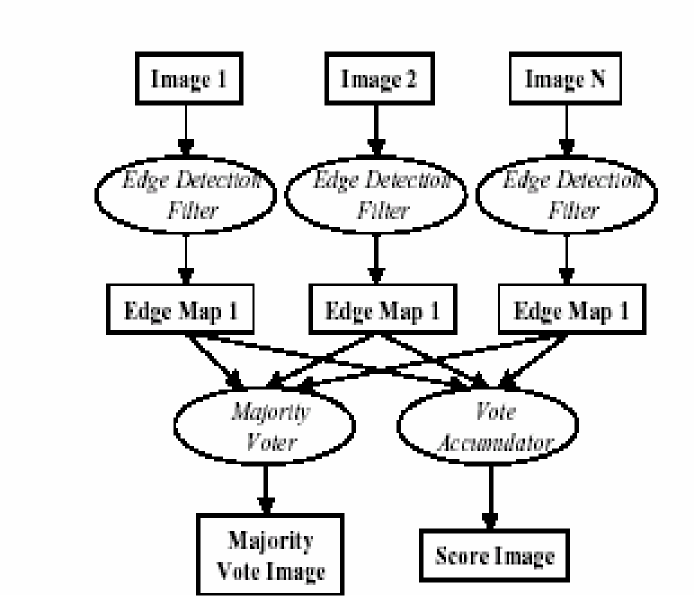
1. Multi-Spectral Cooperative Segmentation

It is used for edge point fusion according to which an edge-thinning algorithm is used in order to preserve a simple pixel in the orthogonal direction of the edge. A majority or a maximum-weighted-sum rule is used for the selection of edge points and the determination of their direction.

The multi-spectral region is extracted by the different mono-spectral images which are processed in parallel but with explicit information exchanges and synchronization. An area is considered a multi-spectral region if it is multi-spectral homogeneous enough. The computation of the homogeneity criterion should be applicable in the area of each mono-spectral image [3].

2. Binary Edge Map Fusion by Majority Voting Rule

Binary edge maps obtained from multi-spectral images are fused pixel to pixel using the majority voting rule. Edge maps are obtained with application of edge detection algorithms from all the different spectral channels. The fused edge information is achieved creating the “majority vote" and "score" images from the maps. Each pixel value in the score image is simply the number of channels the pixel detected as edge point. The majority vote image is simply the score image thresholded at half the number of multi-spectral channels. If the pixel has been considered as edge pixel by at least half of the classifiers it is assigned the value 1, otherwise it is assigned the value 0.The flow chart of the algorithm is the following:



Multi-spectral edge detection algorithm combining spectral edge maps using majority voting rule.

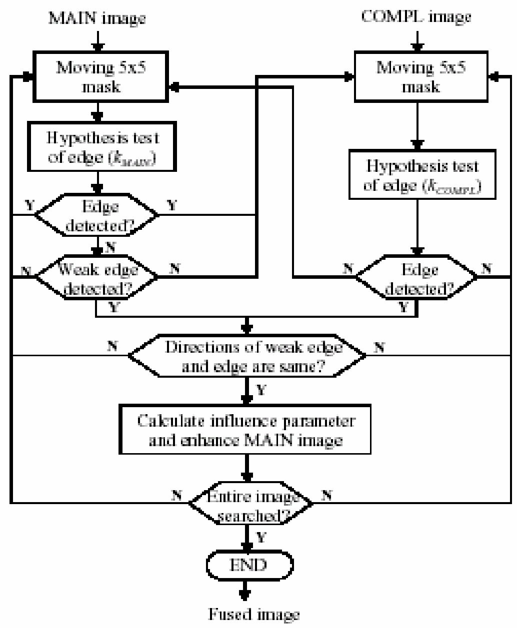
3. Influence Factor Modification Fusion (IFMF)

In contrast to the previous method, the IFMF [29] method performs fusion before each edge map is derived independently. The local edge pixels are calculated using the analysis of variance (ANOVA=analytical tool for image processing) contrast function edge detector via an influence factor [3]. Successful ANOVA models exist for line, edge detection and image segmentation. The edge line detected by the ANOVA algorithm is presented by several parallel lines. At strong lines, where sharp gray-level change exists, thin regions are presented while at weak lines thin and discontinuous regions are found. The size of line is smaller than 5 pixels. Various, mathematically founded, computations take place for the creation of the new, fused image from the addition of the located edges, lines, regions. The rules that condition these computations are based on the principle that the final edge map is generally selected by the basic information where the main image presents a strong line feature. The system takes two images, MAIN and COMPL (ementary), as an input. The first (main) image contains more recognizable information while the second image is used in order to modify the initial and create the image with the combined information. The complementary image is used only when the main image has unilateral edge features, in order to enhance the features of the initial one, if it contains the same feature, or in order to reject it in the opposite case.

The method however meets the following restrictions:

- the refinement of the edge map of the fused image is not an easy task
- It is not necessary to modify the main image and repeat the ANOVA procedure for edge detection. Instead we can simply modify the edge map of the main image and proportionally receive the fusion map.

The flow chart of the method is summarized in the following form:

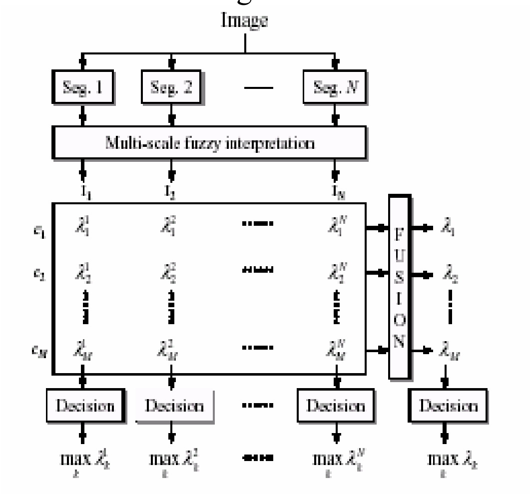


The IFMF method.

4. Fusion of Segmented Images

Instead of selecting a segmentation method it is preferable to receive the best of each method, performing "competition of methods" in a way. The N different algorithms are applied to an image giving N different, segmented images as a result. The regions acquired are interpreted using a multi-scale fuzzy classifier and afterwards the "interpreted" image is merged using various fusion operators. Using a mathematic model in this method too, that will not be analyzed here, certain quantities are also computed based on some selection criterion, usually the criterion of maximum value or mean value, the quantities-coefficients are selected that produce the final, fused image.

The flow chart of method is the following:



Fusion of images segmented and interpreted.

Summary: feature-level VS pixel-level image fusion

Advantages:

- Compared to pixel-level methods, feature-level techniques extract feature information from images. The amount of data is dramatically reduced while the most significant features are preserved.
- Although the data volume is reduced dramatically, feature information is distilled and retained in the edge/region maps or feature vectors so that analytical methodologies can be used for analysis and classification

Disadvantages-open problems:

- Loss of information: loss of information is inevitable in the abstraction process
- Feature selection and feature extraction algorithms.
- The ability of extraction algorithms and fusion rules to conserve image features.
- Feature refinement methods.
- Performance evaluation of fusion algorithms.

A.3. Decision-level techniques

Symbol level or decision level fusion consists of merging information at a higher level of abstraction. It represents high-level information where the symbol represents an input in the form of a decision, where the fusion describes both a logical and a statistical conclusion. The most significant improvement of using symbol-level fusion is the increase in the truth values.

An example of this type of fusion is the construction of probabilities associated with object recognition.

High-level data fusion or decision fusion occurs where sensor data, with or without pre-processing, is combined with other data or a priori knowledge. Each sensor makes an independent decision based on its own observations and passes these decisions to a central fusion unit where a global decision is made. Alternatively, in a decentralized multi-sensor system each node functions performs fusion based on local observations and the information communicated from neighbouring nodes.

In this level of description, sensor data are processed separately and a "determination of identity" is applied using

- **voting techniques:** they provide discrimination with a simple determination of majority, where the most possible object is detected.
- **scoring models:** they compose a set of weights and determine the maximum result with weights
- **other ad hoc methods.**

The advantage of methods of this category is that all knowledge about sensors can be applied separately. Each sensor expert knows the most about the capabilities and limitations of sensors which belong to his cognitive field and can use this information in order to optimize the detection performance.

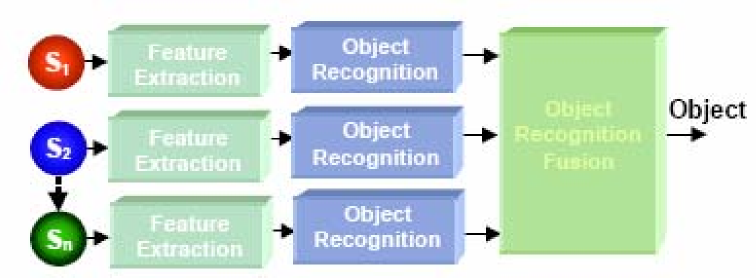
The techniques of this category follow two approaches:

1. **Based on knowledge**, using logical patterns, syntactic rules and relevant combination.
2. **Based on identity,** using assumptions and probabilities in order to classify objects.

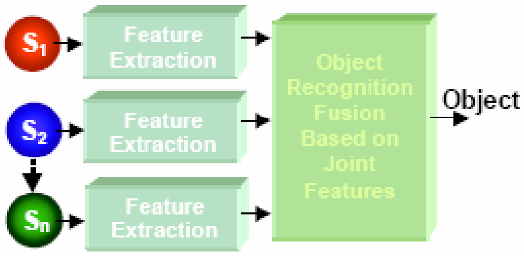
The most basic methods mentioned in international bibliography are the following and are based on mathematic models [30, 31]:

- **Bayes approaches:** this method makes use of complicated mathematic models, mainly probabilistic theory.
- **Dempster-Shafer theory:** this method is non-statistical; nevertheless it can be described with discriminant functions in sensor confidence space with corresponding error functions. Dempster-Shafer evidential theory is an extension of the Bayesian approach, where unknown values are defined as ignorance until "new" information arrives to change the opinion of the system. The uncertainty measures define how well the sensory information that has been captured from sensors match.
- **Fuzzy probabilities:** they are used in order to indicate that there is uncertainty in the estimation of the probability. The fuzzy membership function defines the range of possible probabilities and the 'likehood' of each possible probability.
- **Rule-based method:** it forms an intuitive and flexible approach as it is very easy to incorporate any available a priory knowledge into the system

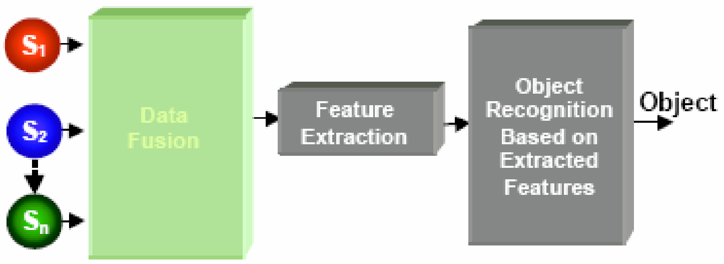
The two images below can be very helpful to understand the different description levels and how the different kinds of information look like within the various stages of processing



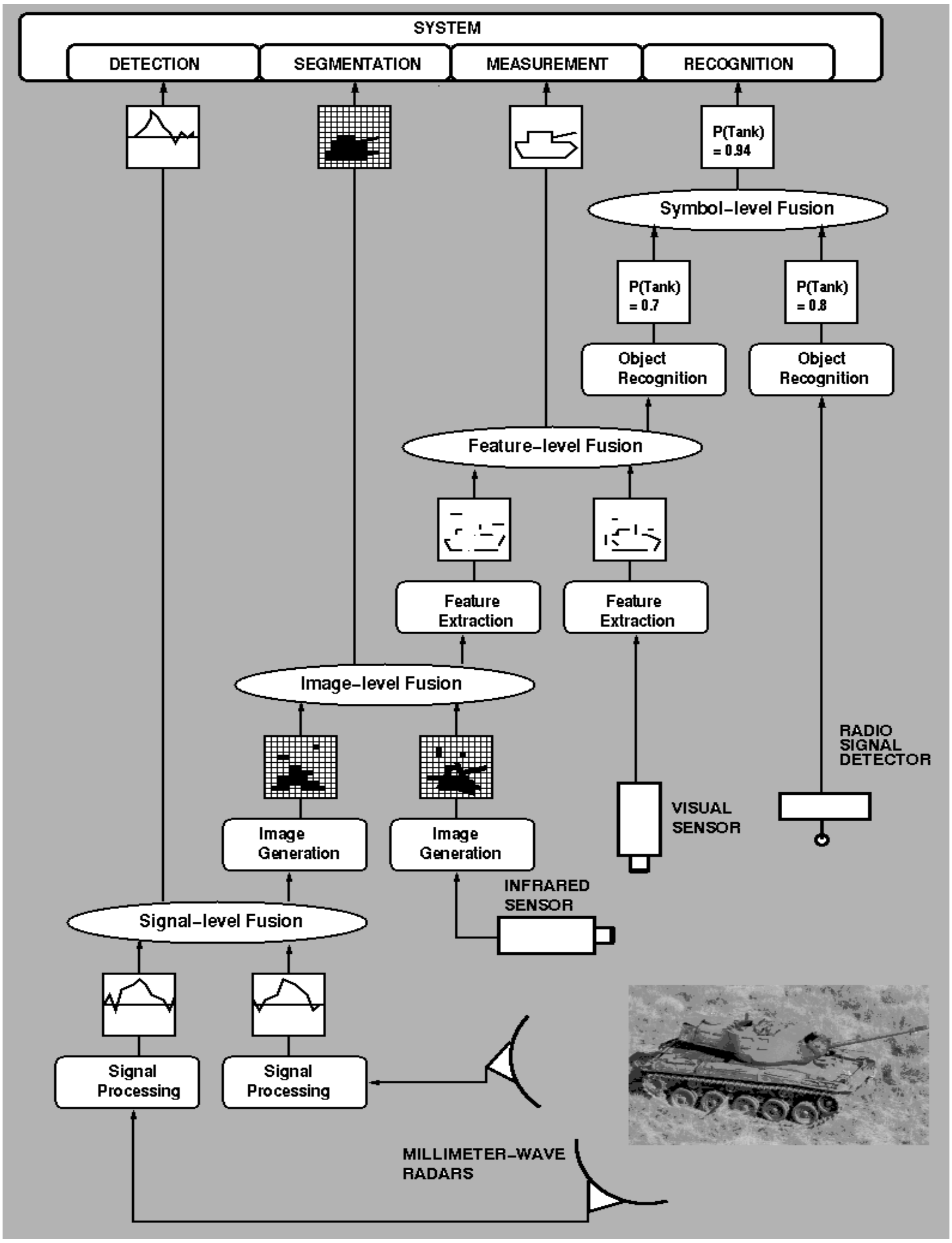
Decision level fusion



Feature level fusion



Pixel level fusion



An example on how a simple "tank image " looks like in the different description levels.

Discussion topics for pixel, feature and decision-level image fusion

Given the problem of image fusion that captured by one or more sensors and going through the implementation stage the following questions immediately arise:

- In which of the three levels should fusion be performed?
- In which mathematic background is the problem modeled?
- Which technique should be applied?
- How can we evaluate its performance?

As already mentioned the answer on these questions that concern image fusion is not easy and absolute. On the contrary: it is ambiguous! Despite the researches on the various fusion theories, their capabilities for feature preservation when combined with image processing techniques have not been studied in fusion models. In any case, the main query is still the problem of evaluation of the performance of image fusion methods, as, despite the existing bibliography and study, the evaluation is not determined by absolutely objective and commonly acceptable criteria for all the applications, methods and inputs.

An important notice is that as long as the fusion level gets higher (by increasing sequence pixel → feature → decision) as much information is lost during the abstractive process of information. Yet, the amount of data to be processed is dramatically decreased, the problem is simplified and the computation load becomes reasonable, specifically in real-time applications. At the same time, the preservation of this information is easier during processing. It appears that the more time the relative information is kept during the abstractive process the higher should the level of fusion implementation. Nevertheless, due to the complexity in understanding it is difficult to be assured that information is not lost, making hard to decide which level of description should be preferred in order to preserve the useful data.

The choice of the suitable level of description depends on the type of the available sensors. When sensors are similar pixel-level description is adopted, otherwise when they differ too much decision-level analysis is more suitable and more computationally efficient. Analysis in feature-level is most suitable when features captured by the sensors can be correlated. In order to achieve the greatest accuracy and information, it is desirable to fuse at the signal level before any information can be lost due to reconstruction or feature extraction.

Another issue that emerges is the handling of inaccuracy and uncertainty of source images. Fuzzy logic approaches, theory of evidence and the Bayesian model are the three basic tools used for this aim. However, their contribution to the interpretation of the initial images should be searched further.

.

Graphic categorization synopsis

A sequence of tables follows as an effort to summarize all the above information in a more readable, compact and presentable

Table1.Levels of description of information

<table>
<tr><th colspan="2">Level</th><th>Description</th></tr>
<tr><td colspan="2">Pixel level
(Data or image or signal level)</td><td>The lowest of the three levels, description of information in its raw form..The goal is the representation of the optical information provided by source images in a fused image without distortion and loss of information</td></tr>
<tr><td rowspan="2">**Feature**</td><td>**intermediate**</td><td rowspan="2">Requires algorithms capable to recognize objects from the various sources of information based on statistical features of dimension, shape and edges. Segmentation algorithms have been proved useful for this category.</td></tr>
<tr><td>**feature**</td></tr>
<tr><td colspan="2">**Decision level
(Symbol level)**</td><td>Processes the source images separately to extract information and applies decision rules to achieve common interpretation and throw differences</td></tr>
</table>

Schema1.Special categorization of image fusion methods

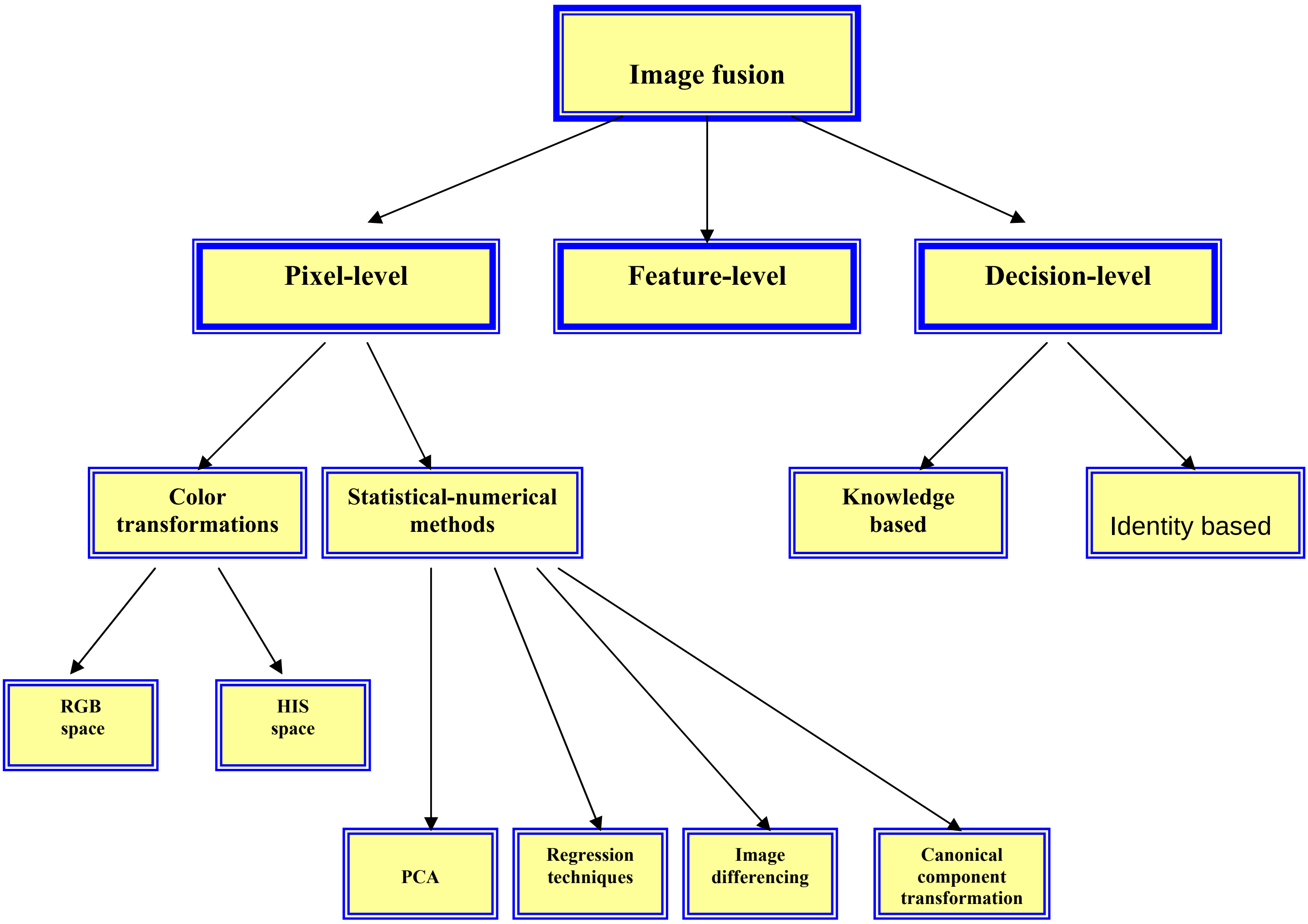


Table2.Categorization of pixel-level methods

Category	Methods	Description
Linear superposition	PCA analysis,IHS ,RGB method,averaging	Probably the most direct way to fuse images. Implements fusion as a linear combination weighted by all input images.
Non linear methods	Bayesian model	One more simple approach by using a simple non-linear operator e.g. min, max or a morphological operator or, to another aspect, image algebra
Optimization approaches	Hidden Markov trees methods	Image fusion is expressed as a Bayesian optimization problem, using a-priory model of the fusion result in order to find the image that optimizes the a-posteriori possibility
Artificial neural networks	Newman and Hartline approach	Inspired by fusion of signals taken by sensors in biological systems, researchers established neural networks to pixel-level image fusion
Image pyramids	Gradient,difference, morphological, averaging, contrast,ratio pyramids	Image pyramids are a sequence of images where each image is costructed by lowpass filtering and downsampling of its “ancestor” image
Wavelet transform	MWD 2DWT	Relevant to the category above with the difference that this transform leads to non-redundant representation of the image having as a disadvantage its shift dependency
Multiresolution fusion scheme	Includes the categories above, Toet’s method, Burt’s method, Wilson’s method, Yocky’s method, Region-based method	The basic idea lies to the fact that the human vision system especially sensitive to local contrast changes e.g.edges.The fused image arises as a combination of multiscale edge representation

IMAGE FUSION	**PIXEL LEVEL**	**Multiresolution Analysis**	**Image pyramids**	**-Gaussian -Laplacian -Contrast -Gradient -Morphological**
			Wavelets	**-DT-CWT -DDWT -UDWT -RWT**
			Other	**-Maximum Contrast Selection -Salient Patern Selection -Perceptually SPS -MWD-based -Additive MWD fusion -Region-based MWD fusion -Gradient-based fusion**
		Linear superposition		**-PCA -HIS,RGB models -Averaging**
		Non-linear methods		**-Bayesian model**
		Neural Networks		
		Optimization approaches		**-Hidden Markov Trees**
	FEATURE LEVEL	**Parametric templates**		**-Multi-spectral Cooperative Segmentation -Binary Edge Map Fusion by Majority Voting Rule -Influence Factor Modification Fusion(IFMF) -Fusion of segmented images**
		Hierarchical clusters		
		Neural Networks		
		Knowledge-based approaches		
	DECISION LEVEL	**Voting techniques**		**-Bayes approach -Dempster-Shafer theory -Fuzzy probabilities -Rule-based method**
		Scoring models		
		Other *ad hoc* methods		

B. Basis Functions

Basis functions, in general, are carefully selected and well defined functions, a linear combination of which can approximate a source signal. A simple example comes from the digital world: Every two dimensional vectors(x, y) is a combination of the vectors (1, 0) and (0, 1). These two vectors constitute the basis "functions" (vectors) because x multiplied by (1, 0) and y multiplied by (0, 1) result, by adding them, to the original vector (x, y). In addition, the basis vectors have an important property: they are orthogonal to each other.

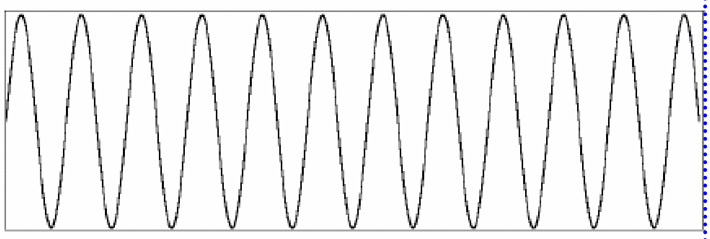
In a similar way, basis functions in the analog world (functions) are treated. A simple example is a voice signal: we can approximate it by adding sines and cosines using combinations of amplitudes and frequencies. Sine and cosine are the basis functions now. Expanding this concept we can create the **scale-varying basis functions** by diving data using different scale sizes. As an example let's imagine that we have a signal in the domain (0, 1). We can divide it by using two step functions in the intervals (0, ½) and (1/2, 1). Similarly we can divide it by using other two step functions in each interval, resulting in four step functions in the intervals (0, 1/4), (1/4, 1/2), (1/2, ¾), (3/4, 1) and so on. So we have managed to represent the original signal by basis functions within a particular scale [7].

C. Wavelet families

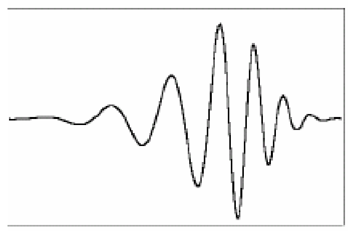
Wavelet theory is based on analyzing signals to their components by using a set of basis functions. One basic characteristic of wavelet basis functions is that they relate to each other by simple scaling and dilation. The original function, known as mother wavelet, is designed under some desired characteristics and rules and is used to generate all basis functions. In most applications it is required that the original signal be synthesized from the wavelet coefficients. To do so, wavelets should satisfy some conditions. For example, if we use the same wavelets for both decomposition and reconstruction they should satisfy the orthogonality condition, otherwise, if we use two different sets of wavelets, one for the analysis and one for the synthesis procedure, they should satisfy the biorthogonality condition.

It is not an easy part to design a uniform procedure in order to develop the best mother wavelet for a given class of signals. However, based on several general characteristics of the wavelet functions, it is possible to determine which wavelet is more suitable for a given application.

A wavelet is a small wave with finite energy, which has its energy concentrated in time or space. It still has the oscillating wave-like characteristics but has also the ability to allow simultaneous time and frequency analysis with a flexible mathematical approach [15].



Sine wave with infinite energy

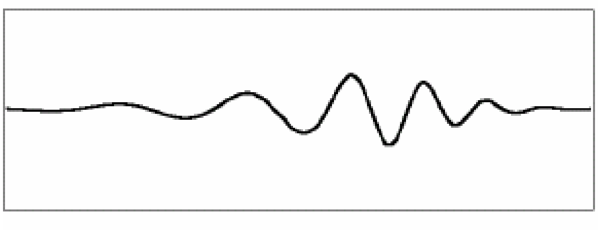


Wavelet of finite energy

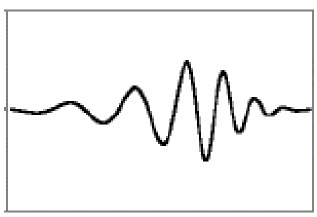
The wavelet transform is a two-parameter expansion of a signal in terms of a particular wavelet basis function. If $\psi(x)$ represents the mother wavelet, all other wavelets are computed by the following equation:

$$\psi_{\alpha,\tau}(t)=\left(\frac{1}{\sqrt{\alpha}}\right)\psi\left(\frac{t-\tau}{\alpha}\right)$$

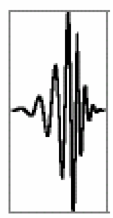
where the scaling factor α, usually $\alpha=2^{-j}$, defines the scale and the factor τ, usually $\tau=kT2^{-j}$, defines the dilation. The parameterization of time or space by integer k and the frequency or scale by integer j turns out to be very effective. Selecting different values of j, k the different wavelets are computed



mother wavelet



wavelet at scale 1

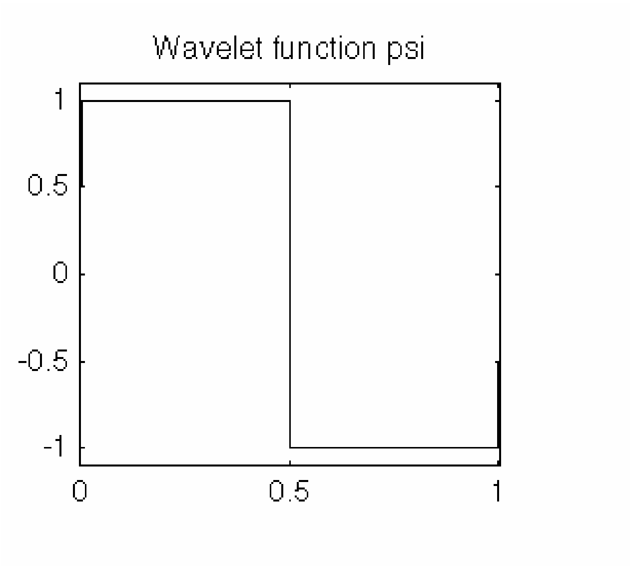


scale 2

A list of popular wavelets follows

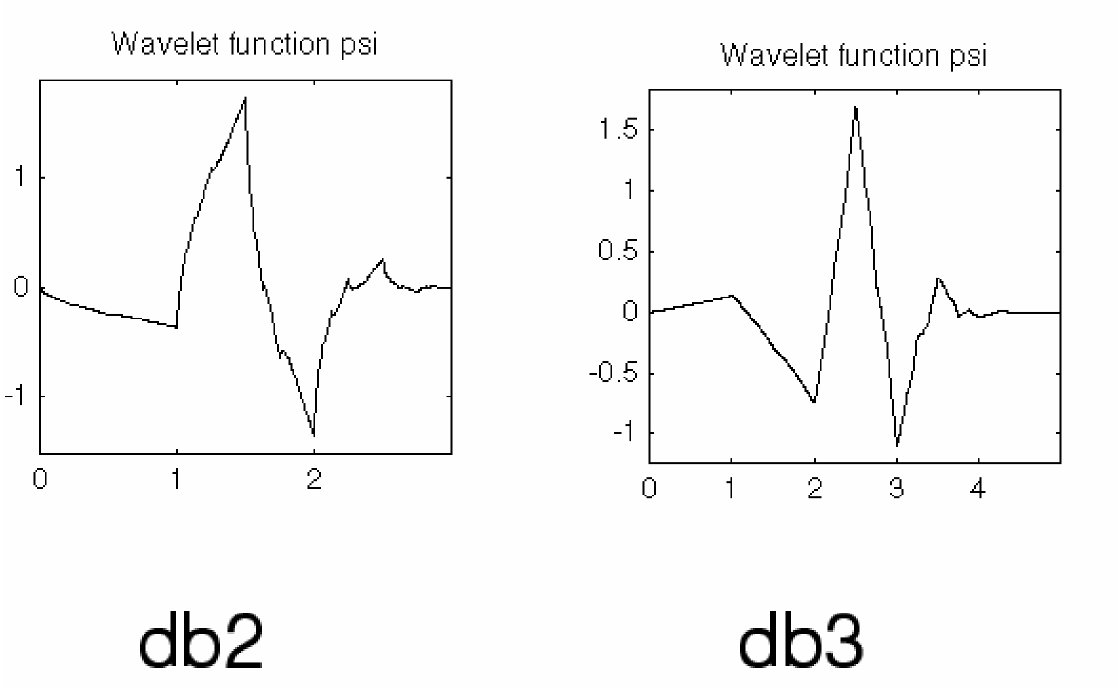
Haar wavelet

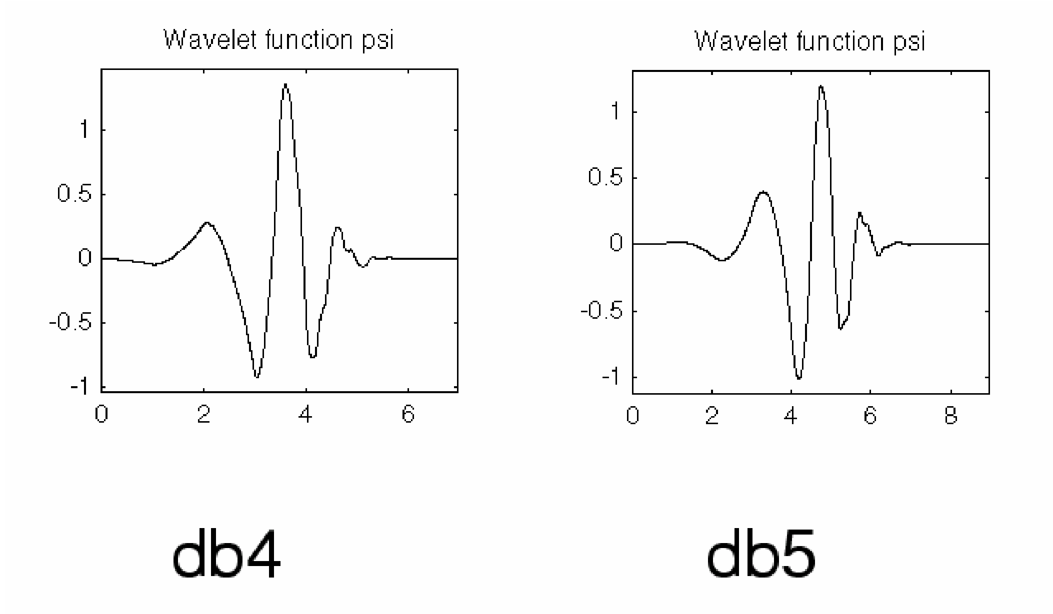
It is the simplest wavelet, the only orthogonal that has symmetric analysis and synthesis filters. Due to its simplicity and fast computational efficiency, it's a good choice for image processing. It is also known as Daubechies1.



Daubechies wavelet family

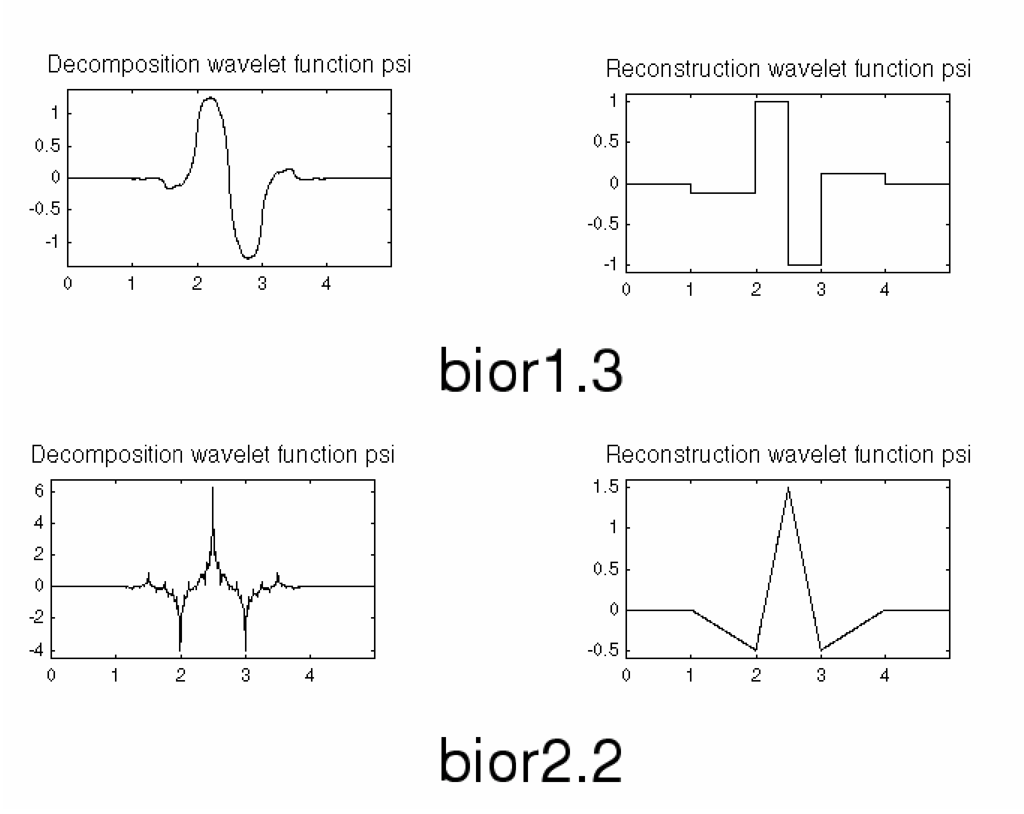
Ingrid Daubechies, one of the greatest researchers in wavelet theory, invented what are called compactly supported orthonormal wavelets, which make discrete wavelet analysis practicable. Daubechies wavelets have good compression property for wavelet coefficients but not for the approximation ones. Depending on the order N of the wavelet, the basis function can take the following form:



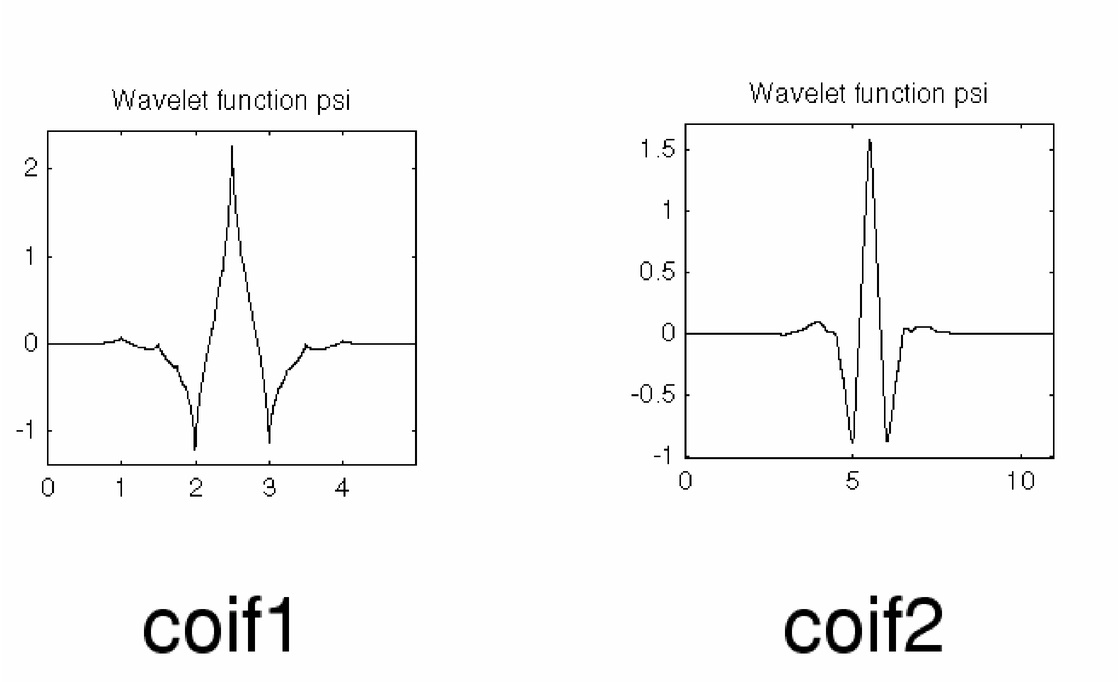


Biorthogonal wavelet family

This wavelet family satisfies the biorthogonality condition, which is essential for perfect image reconstruction. By using two different wavelets, one for the decomposition and one for the reconstruction procedure, interesting properties are derived.

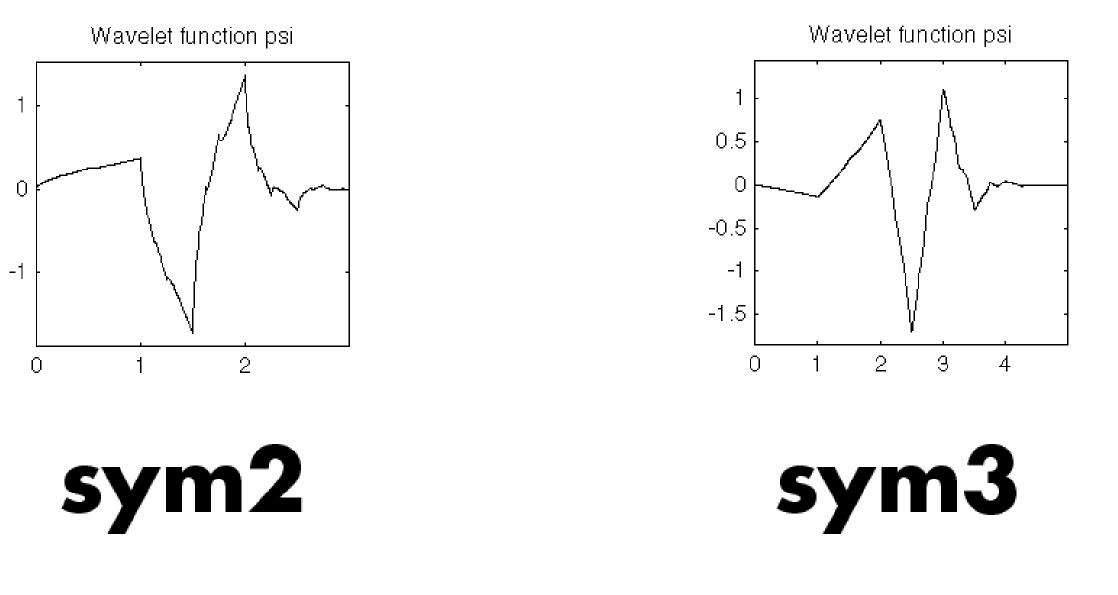


Coiflets wavelet family

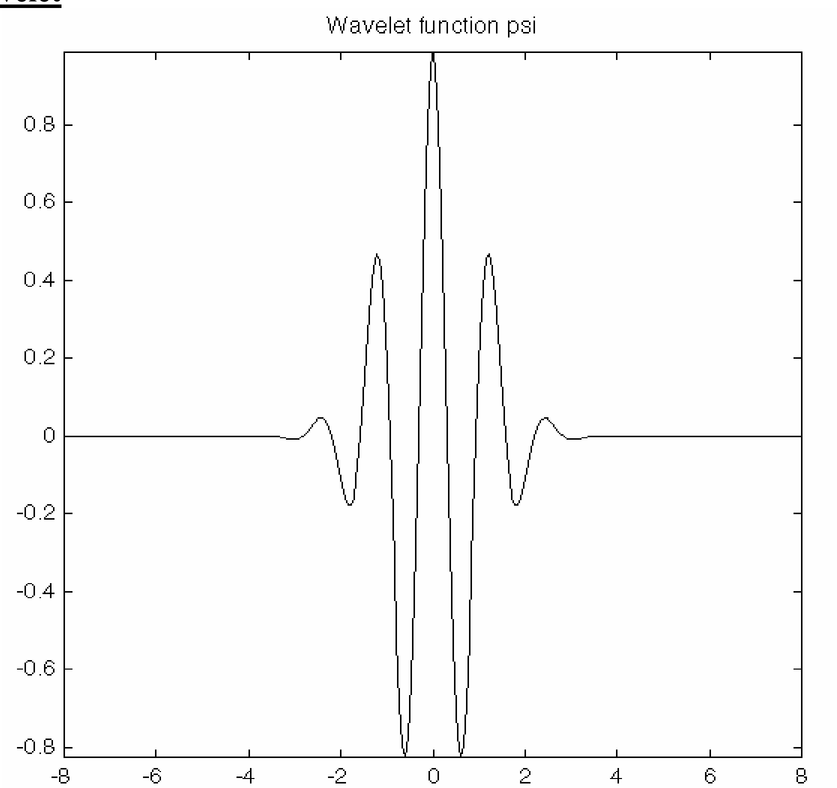


Symlets wavelet family

The symlets are nearly symmetrical wavelets and were proposed by Daubechies as a modification of the Daubechies wavelet family.

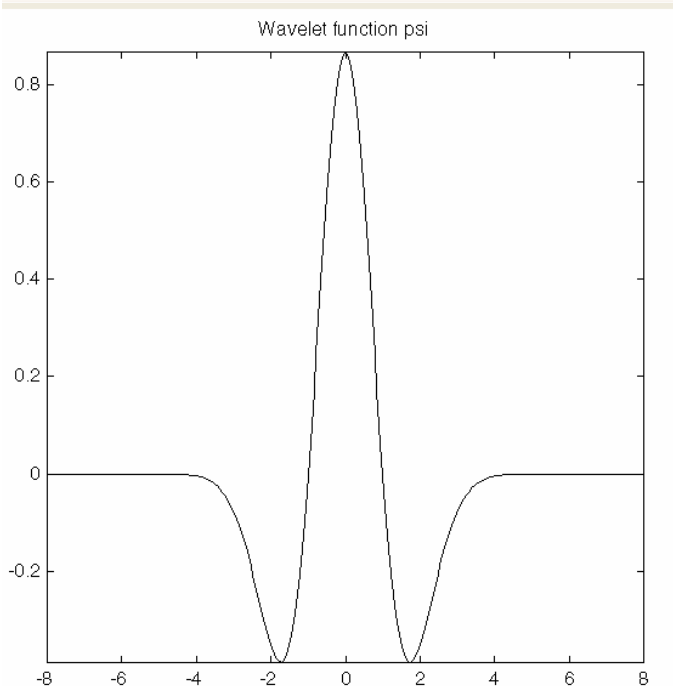


Morlet wavelet

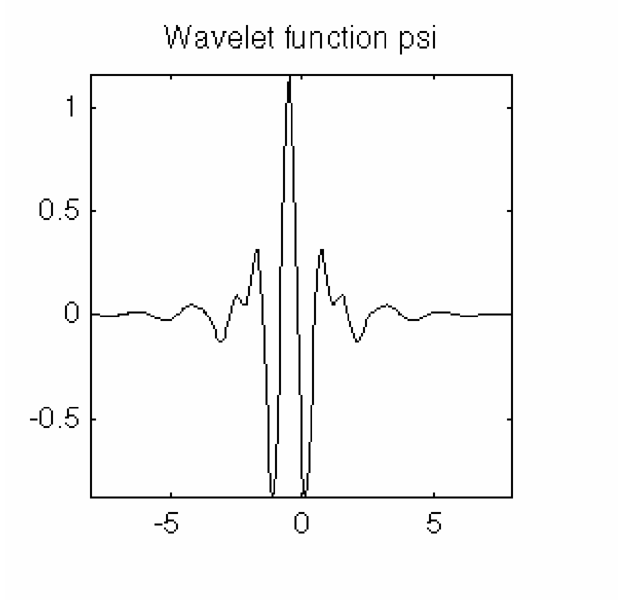


Mexican hat wavelet

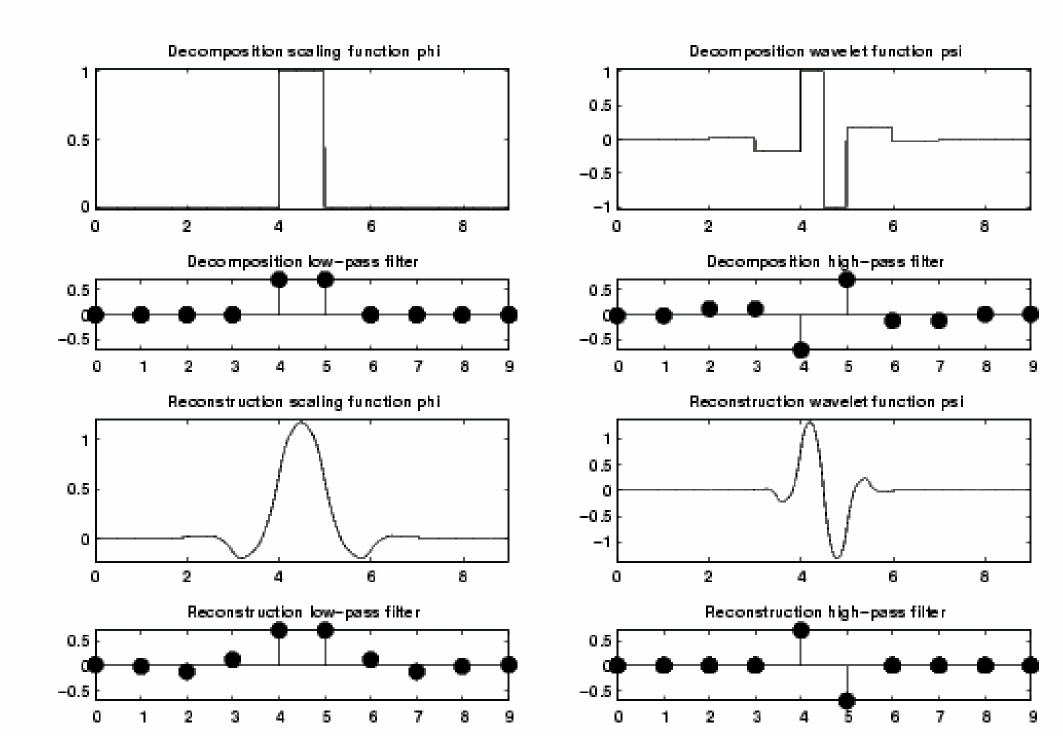
It is derived from a function that is proportional to the second derivative of the Gaussian probability function



Meyer wavelet



Reverse biorthogonal wavelet family



D. Complex and real wavelets

As already mentioned, complex wavelet transforms offer some basic advantages against real wavelet transforms. However their complexity, the demand for a good mathematical background and mainly their implantation difficulty makes their use in image applications unattractive. Some basic complex wavelets are listed below:

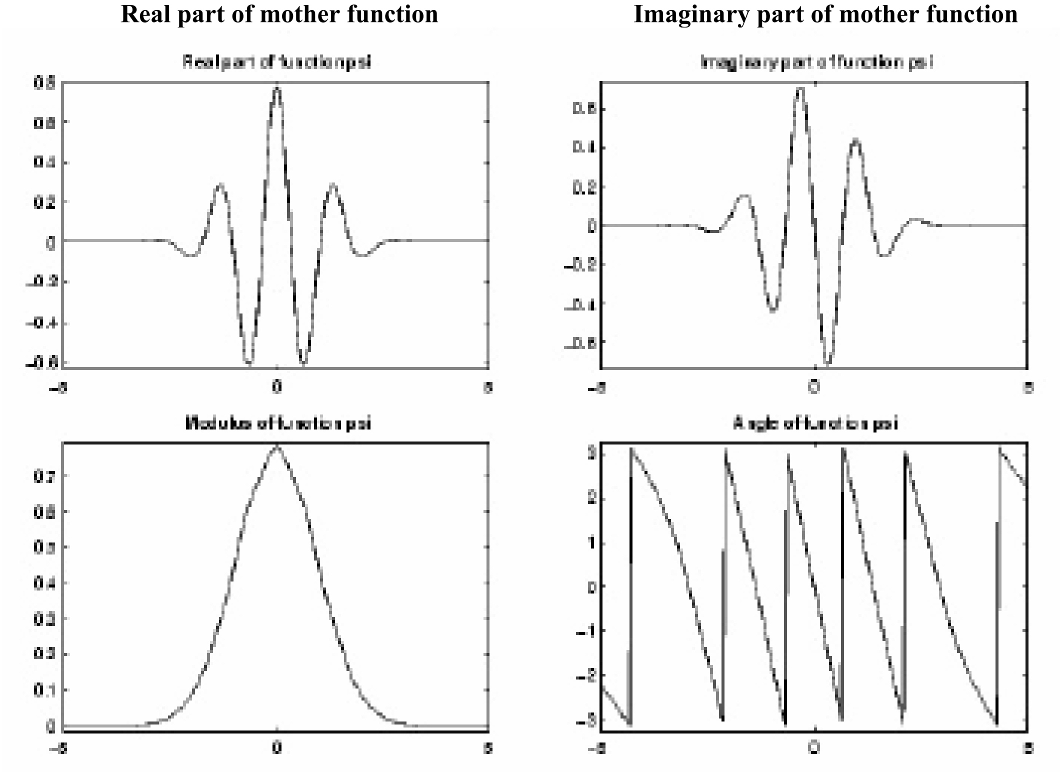
Complex Gaussian Wavelets

This family is built starting from the complex Gaussian function and taking the Nth derivative, where N defines the order of the wavelet family

Definition: derivatives of the complex Gaussian function

f(x)=$C_p e^{-ix} e^{-x^2}$ where C_p is such that $\left| f^{(p)} \right|^2 = 1$ where $f^{(p)}$ is the p-th derivative of f

Real part of mother function **Imaginary part of mother function**



Modulus of mother function **Angle of mother function**

Morlet wavelet

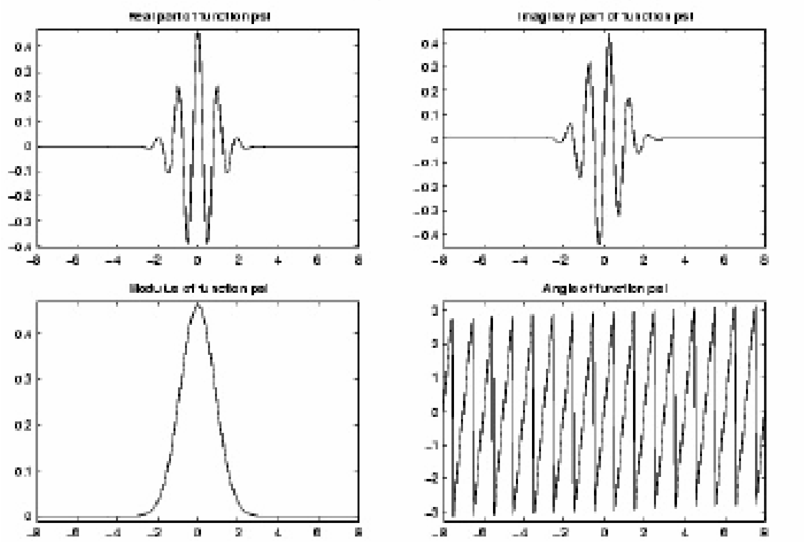
Definition: a complex Morlet wavelet is

$$\psi(x) = \sqrt{\pi f_b} \cdot e^{2i\pi f_c x} \cdot e^{-\frac{x^2}{f_b}}$$

depending on two parameters:

Fb is a bandwidth parameter

Fc is a wavelet center frequency



Complex frequency B-spline wavelets

Definition: a complex Frequency B-Spline wavelet is

$$\psi(x) = \sqrt{f_b} \cdot e^{2i\pi f_c x} \cdot \left(\sin c(\frac{f_b x}{m}) \right)^m$$

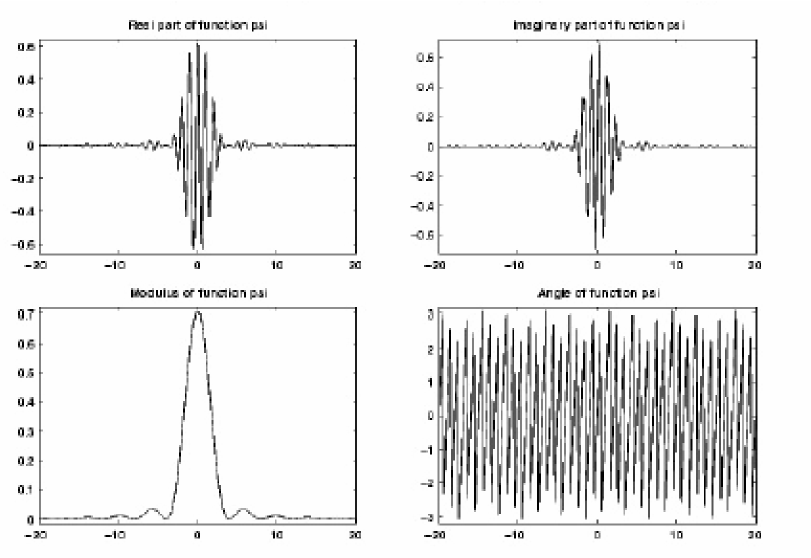
depending on three parameters:

M is an integer order parameter (>=1)

Fb is a bandwidth parameter

Fc is a wavelet center frequency

For M = 1, the condition Fc > Fb/2 is sufficient to ensure that zero is not in the frequency support interval.



Complex Shannon Wavelets

Definition: a complex Shannon wavelet is

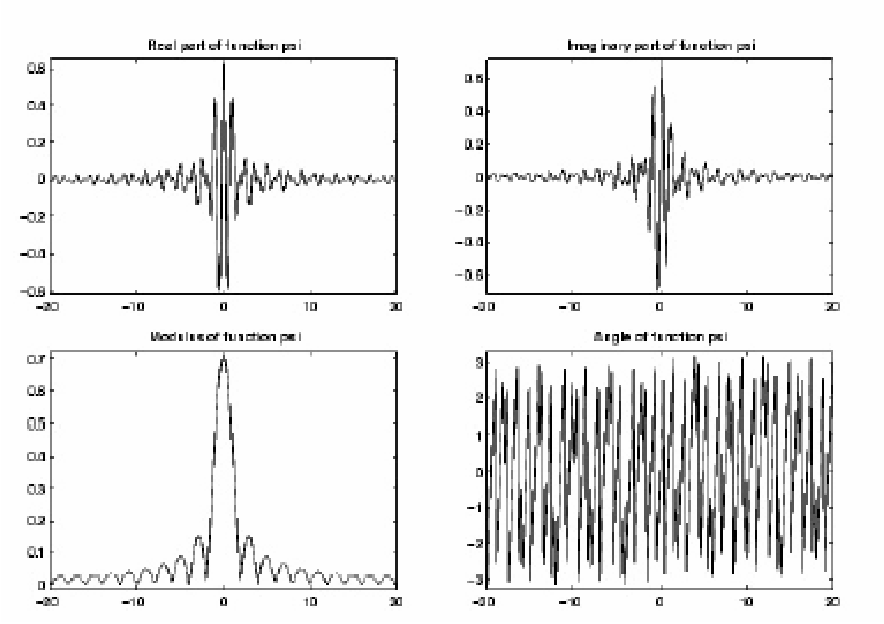
$$\psi(x) = \sqrt{f_b} \cdot e^{2i\pi f_c x} \cdot \sin c(f_b x)$$

depending on two parameters:

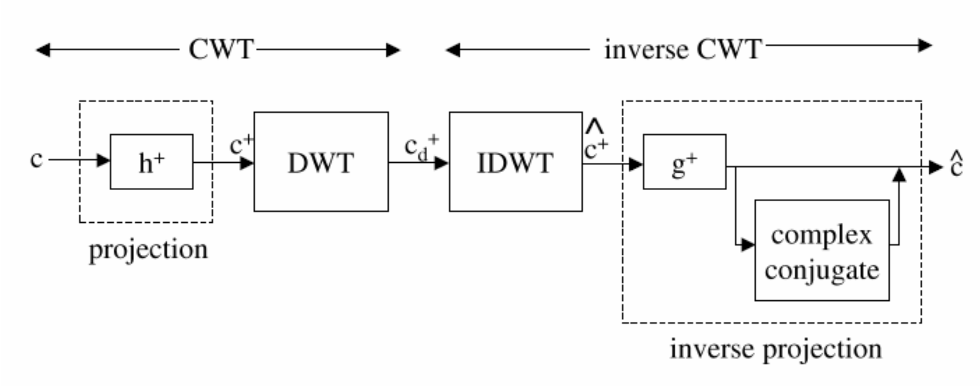
Fb is a bandwidth parameter

Fc is a wavelet center frequency

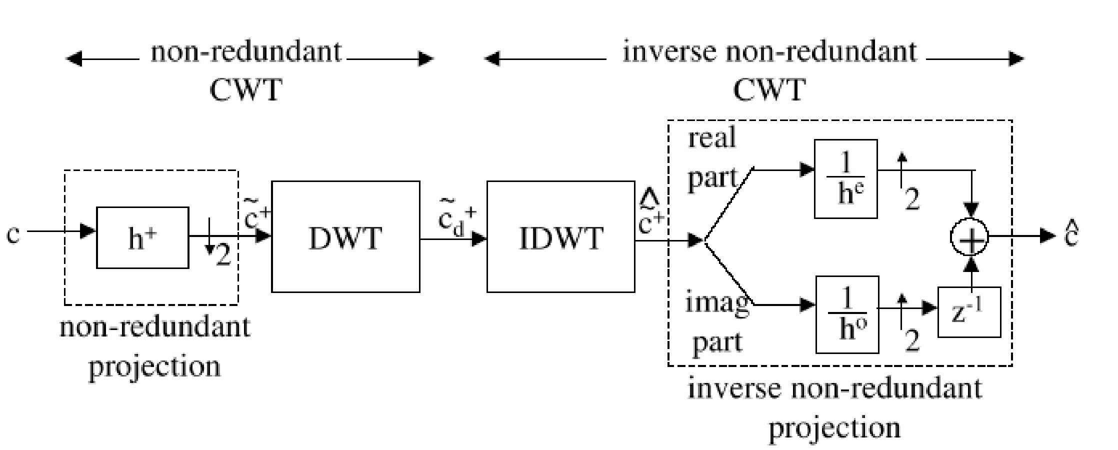
The condition Fc > Fb/2 is sufficient to ensure that zero is not in the frequency support interval.



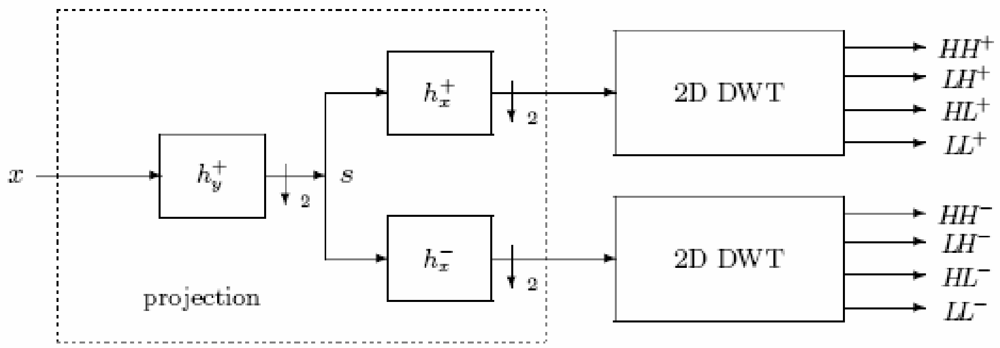
The Discrete Wavelet Transform using filters obtained by complex wavelets is difficult to be implemented. Instead, Fernandez proposed a projection-based Complex Wavelet Transform in order to exploit the applicability of real Discrete Wavelet Transform [21]. There is a lot of theoretical background beyond this idea but the main themes are briefly mentioned. The main idea is that complex wavelet coefficients can be obtained by projecting the input signal onto the Hardy space and then computing its wavelet transform using the DWT associated with a real wavelet, as illustrated in the figure below



As an improved version Fernandez proposed the non-redundant Complex wavelet transform by simply inserting the up/downsampling procedure in the projection phase



The projection filter h^+ has certain characteristics [21]
This transform can be expanded to two dimensions in a similar way



The projection filter h^+_x is performed along rows and the filter h^+_y along columns. The role of the projection phase is to eliminate the redundant negative-vertical frequencies and decouple the positive horizontal frequencies from the negative ones.

R e f e r e n c e s

1. **"Digital Image Fusion-Literature review"** by http://dfwm.ygent.be/forman/projecten/lake_chad001/DigitalImageFusion/...
2. **"Image Segmentation by Multisensor data fusion"** by G.Seetharaman and G.H.Chu, 22nd Southerneastern Symposioum on System Theory, 1990, pp583-587
3. **"Multi-Sensory Image Fusion Techniques at Image- and Feature-Levels",** A Qualifying Examination Report by Yuhua Ding Presented to the Qualifying Examination Comitee, School of Electrical and Computer Engineering Georgia Institute of Technology, February 9, 2001
4. **"An introduction to Multisensor Data Fusion"** by D.L Hall and J.Llinas, Proc.ieee, vol.85, no.1, Jan.1991, pp.6-23
5. **"Investigations of image fusion",** Electrical Engineering and Computer Science Department, Lehigh University, http://www.ece.lehigh.edu
6. **"Tools and methods for fusion of images of different spatial resolution",** C. Pohl, International Archives of Photogrammetry and Remote Sensing, Vol. 32, Part 7-4-3
7. **"An introduction to wavelets",** Amara Graps
8. **"Wavelets",** G.Strang, *American Scientist,* Vol.82,1992, pp. 250-255
9. **"Wavelets for image fusion",** Stavri Nikolov, Paul Hill, David Bull, Nishan Canagarajah
10. Z.Wang and A.C.Bovik, **"A universal image quality index",** IEEE Signal Processing Letters, vol. 9, no.3, pp.81-84, March 2002
11. **"A new quality metric for image fusion",** Gemma Piella and Henk Heijmans
12. **"Shift-invariance in the Discrete Wavelet Transform",** Proc.VIIth Digital Image Computing: Techniques and Applications, Sun C., Talbot H., OurselinS. And Adriaansen T. ,Dec.2003
13. N. G. Kingsbury, **"A dual-tree complex wavelet transform with improved orthogonality and symmetry properties"**, Proceedings of the IEEE Int. Conf. on Image Proc. (ICIP), 2000
14. **"An introduction to wavelets",** Bruce Knuteson, December 1996
15. **"Wavelet characteristics"**, Ali M. Reza, Spire Lab, UWM, October 19, 1999, White Paper
16. **"Image processing with complex wavelets",** Nick Kingsbury, Phil.Trans. Royal Society Lond. A (1999) 357, 2543-2560
17. **"Complex wavelet transforms with allpass filters",** Felix C. Fernandes and Ivan W. Selesnick, Rutger L.C. van Spaendonck, C. Sidney Burrus, December 18, 2002
18. **"Wavelets and multiscale edge detection",** Brody Dylan Johnson, SAINT LOUIS UNIVERSITY
19. **"Multiscale Image Edge Detection",** Ivan Christov, 1.130/18.327, Spring 2004
20. **"Characterization of signal from multiscale edges",** Stefan Mallat and Sifen Zhong, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE

INTELLIGENCE, VOL. 14, NO. 7, JULY 1992

21. **"Singularity Detection and Processing with wavelets",** IEEE TRANSACTIONS ON INFORMATION THEORY, VOL. 38, NO 2, MARCH 1992
22. **"Medical image fusion by wavelet transform modulus maxima**", Guihong Qu, Dali Zhang and Pingfan Yan, OPTICS EXPRESS 184, Vol. 9, No 4, 13 August 2001
23. **"Image fusion methods"** by http://www.metapix.de
24. **"Hierarchical Image Fusion",** Machine Vision and Applications, by A.Toet, vol.3, no.1, 1990, pp 699-705
25. **"Gradient-Based Multiresolution Image Fusion "** by Valdimir S.Petrovic,ieee,and Costas S. Xydeas
26. **"Fusion Tecnhiques in Remote Sensing"** by Farhad Samadzadegan, University of Tehran, Department of Surveying and Geomatics, Tehran, Iran
27. **"A syrvey of Fusion Techniques for Multi-spectral Images"** by Tiranee Achalakul
28. **"A new way of image data fusion: The Multi-spectral Cooperative Segmentation"** by P.Bonnin, B. Hoeltzener-Douarin, and E. Pissaloux, Proc. Intl Conf Image Processing, vol.3, 1995, pp. 572-575
29. **"Multisensor Image Fusion Using Influence Factor Modification and the Anova Methods"** by Y. Li, Y. V. Venkatesh, and K.C. Chung IEEE Tran. Geoscience and Remote Sensing ,vol.38, no.4, July 2000, pp.1976-1988
30. **"A comparison of decision-level sensor fusion methods for anti-personnel landmine detection"** by F.Cremer,K.Schutte,J.G.M.Schavemaker,E.den Breejen
31. **"Binary labeling and decision-level fusion"** by Terry Windeatt, RezaGhaderi, Univ.of SurreyGuildford, Surrey GU2 5XH