

# Chapitre 2: Méthodologie



### II.1 Architecture et représentation de la méthode :

La détection des vertèbres dans les images radiographiques est une tâche difficile qui est due à leurs structures complexes, et aussi au manque de contraste entre les structures osseuses et les tissus musculaires. Dans ce chapitre nous allons décrire l'architecture de la méthode proposée pour résoudre ce problème. Notre processus de détection est divisé en trois principales parties :

La première étape implique une procédure d'amélioration de la qualité des images par ajustement d'intensité, afin de clarifier les régions à extraire.

La deuxième étape qui est primordiale dans notre système, représente l'extraction de caractéristiques des régions des vertèbres.

La dernière étape met en œuvre un modèle de classification basée sur une méthode à base des machines à vecteur support –SVM – pour effectuer une bonne classification et renforcer la détection. Les résultats de cette étape sont ensuite raffinés par une méthode d'estimation RANSAC.

Un aperçu de notre système est montré dans le schéma ci-dessous (figure 1).

Dans la phase d'entraînement, le système prend comme entrée :

- 1) Un ensemble d'images contenant les vertèbres qui ont été aligné pour avoir la même taille.
- 2) Pour chaque modèle de notre base, une représentation intermédiaire, encapsulant les informations importantes concernant la classe des vertèbres, est calculée. Et cela pour donner un ensemble de vecteurs de caractéristiques positives et négatives. Ces vecteurs sont utilisés pour entraîner le classificateur afin de différencier les deux modèles de classes.

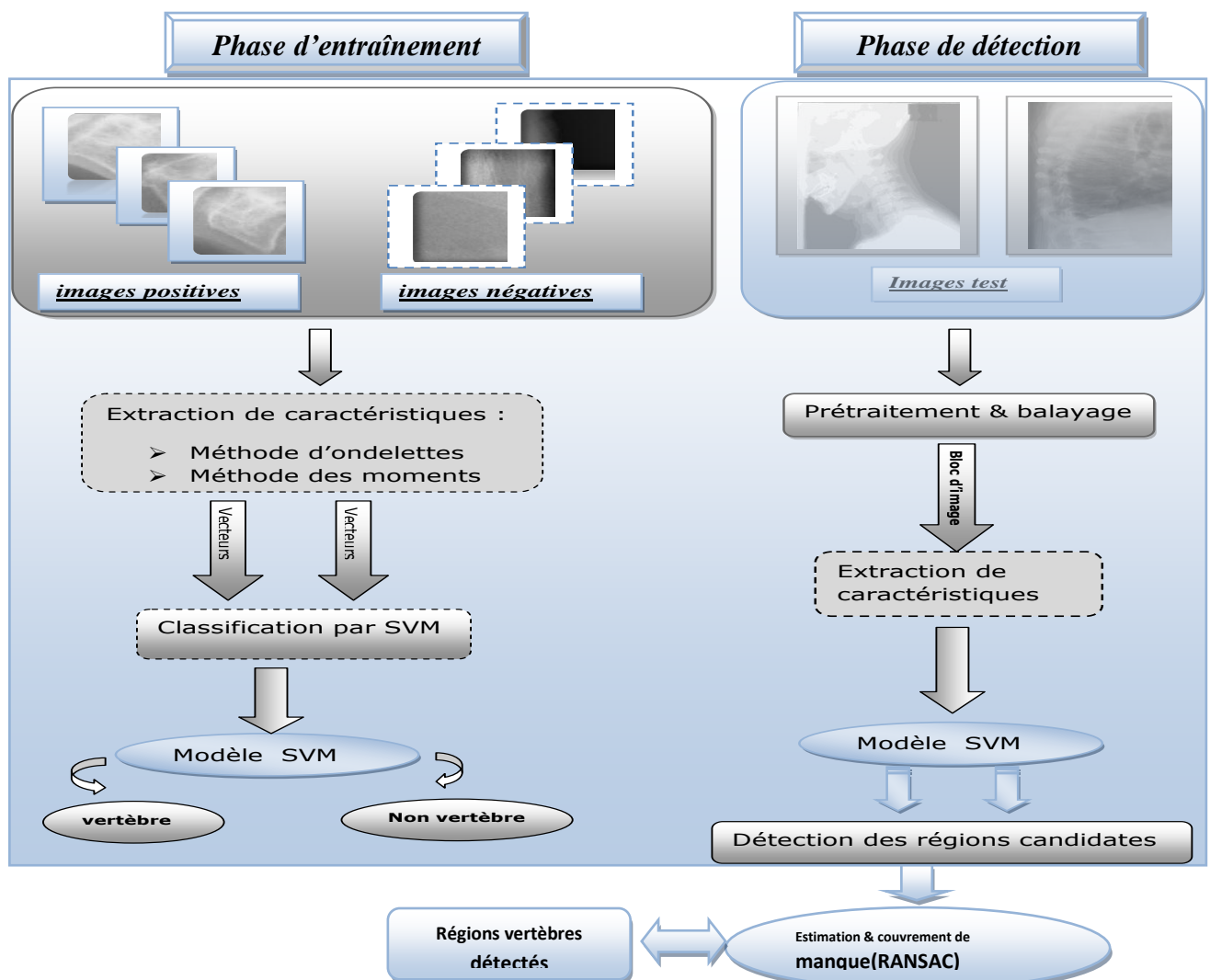
Cette représentation a été calculée par l'utilisation de deux descripteurs différents, qui sont les ondelettes, et les moments HU, afin de pouvoir distinguer l'information la plus pertinente pour une bonne caractérisation.

Afin d'améliorer de la performance de notre classificateur, la méthode de validation croisée a été utilisée dans le but d'aboutir à une bonne minimisation d'erreur de généralisation de nos données d'apprentissage.

Dans la phase de test, nous nous intéressons à la détection des vertèbres dans une nouvelle image. Le système fait glisser une fenêtre de taille fixée dans toute l'image pour décider si le modèle peut être assimilé à une région d'intérêt.

A chaque position de la fenêtre, on extrait les mêmes caractéristiques prédéfinies dans la phase d'entraînement, ensuite nous les injectons dans le classificateur. La sortie va déterminer si la région correspond ou pas à une région de vertèbre. Une méthode d'estimation est ensuite utilisée afin de raffiner les résultats de détections (fausse ou manque de détections).

Nous notons que pour une détection multi-échelle, il suffira de redimensionner l'image en entrée et effectuer le même traitement, en utilisant toujours la même fenêtre de balayage.



**Figure1 : Schéma général de la méthode de détection des vertèbres par SVM.**

### II.2 Le prétraitement :

Le but de cette étape est d'éliminer, le plus possible, les informations non pertinentes dues au bruit résultant de l'acquisition de l'image, et par conséquent, faciliter l'extraction des informations utiles à l'analyse. Un traitement de rehaussement de contraste est appliqué aux images rayon X de notre base. Ce traitement consiste à accentuer les caractéristiques d'une image afin de rendre son affichage plus convenable à l'analyse.

L'objectif est d'augmenter le contraste d'images afin d'accroître la séparabilité des régions.

La méthode que nous avons retenue est une égalisation adaptative d'histogramme :

#### II.2.1 Modification d'histogramme :

L'histogramme d'une image est une fonction  $H$  définie sur l'ensemble des entiers naturels :

$$H(x) = \text{card} \{p: I(p) = x\}$$

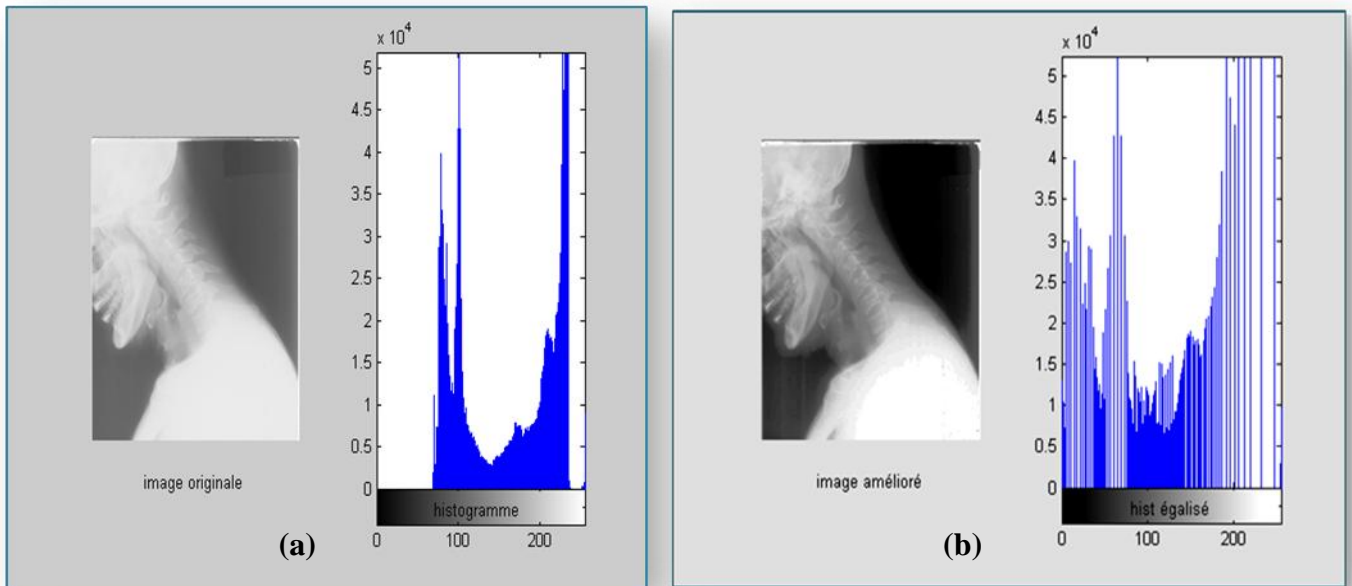
$H(x)$  correspond au nombre d'occurrences de niveaux de gris présent dans l'image. Autrement dit, l'histogramme est une représentation graphique de la distribution des valeurs des niveaux de gris.

Le principe de la modification d'histogramme est d'appliquer une linéarisation, afin de répartir uniformément les valeurs des pixels sur l'ensemble de l'histogramme.

#### II.2.2 L'égalisation d'histogramme :

Cette opération consiste à calculer à partir de l'histogramme  $H(x)$  de l'image  $I$  une fonction de rehaussement des niveaux de gris  $f$  telle que l'image rehaussée  $J$ , définie par :  $J(p) = f(I(p))$  puisse avoir son histogramme  $H_j$  se rapprochant le plus possible d'une fonction plate.

La figure 2 suivante présente un exemple d'une image de la base avant et après l'égalisation d'histogramme.



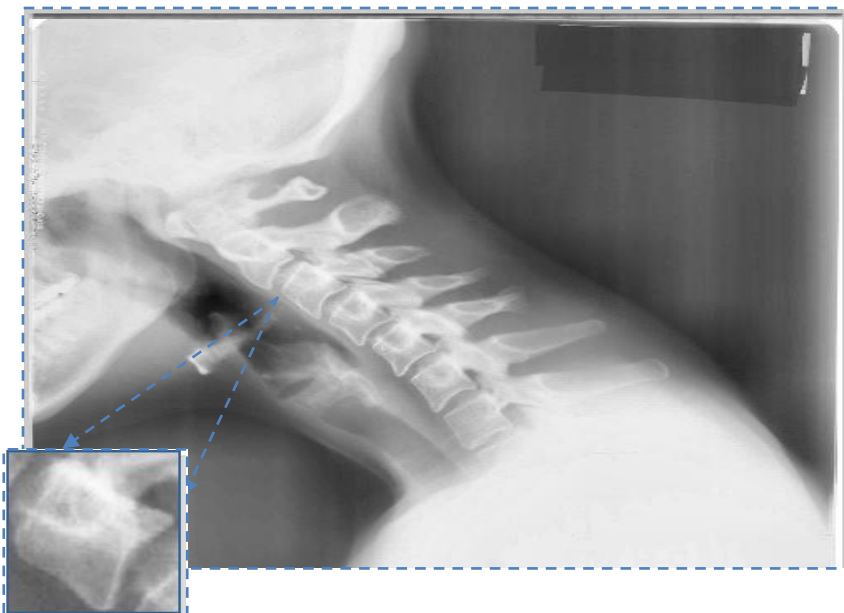
**Figure 2 : Exemple d'image et son histogramme (a) : avant égalisation (b) après égalisation**

### II.2.3 L'égalisation adaptative d'histogramme :

*Son principe consiste à appliqué sur chaque pixel ainsi que « sa région contextuel » une égalisation d'histogramme.*

*Cette région représente en effet les pixels voisions entourant le pixel traité.*

Toutes les images de notre base ont été améliorées pour mettre en valeur les caractéristiques visuelles des différentes régions des vertèbres (Figure 3) afin de faciliter la tâche d'extraction des caractéristiques.



**Figure 3 : Exemple d'image amélioré par égalisation d'histogramme adaptative**

### II.3. L'extraction des caractéristiques :

L'analyse des images se focalise généralement autour des attributs de bas niveaux tels que la texture, la forme, et la couleur.

Il y'a principalement deux approches de caractérisation :

La première est la construction de descripteurs **globaux** à toute l'image c.à.d. fournir des observations sur la totalité de l'image.

L'avantage de ces descripteurs est la simplicité de leur mise en œuvre, ainsi que le nombre réduit d'observations que l'on obtient. Cependant leur inconvénient majeur est la perte de l'information de localisation des éléments de l'image.

La seconde approche est **locale** et consiste à calculer des attributs sur des portions restreintes de l'image.

L'avantage des descripteurs locaux est de conserver une information localisée dans l'image, évitant ainsi que certains détails ne soit noyés par le reste de l'image. L'inconvénient majeur de ces techniques est que la quantité d'information produite est très grande.

Le choix des caractéristiques extraites est souvent guidé par la volonté d'invariance ou robustesse par rapport aux transformations de l'image.

L'approche utilisée dans notre solution est une approche à la fois locale car basée sur le choix de la texture et la forme par la méthode d'ondelettes et globale par l'utilisation des moments de HU.

#### II.3.1. La représentation par la transformée en ondelettes :

La texture peut être vue comme un ensemble de pixels (niveaux de gris) spatialement organisés selon un certain nombre de relations spatiales, ainsi créant une région homogène. De ce fait, la modélisation des textures est portée sur la caractérisation de ces relations spatiales.

Parmi les modèles les plus connus nous pouvons citer : les méthodes statistiques (Matrice de cooccurrence, différence de niveaux de gris, etc.), les méthodes

fréquentielles (Transformée de Fourier, Filtre de Gabor, et les ondelettes, qui sont utilisées dans notre travail).

### II.3.1.1. Introduction a la théorie d'ondelette :

L'imagerie médicale a révolutionné les pratiques médicales. Néanmoins, de nombreux problèmes liés au traitement d'image sont encore ouverts et leur résolution (même partielle) peut aboutir à une amélioration des diagnostics et des actes chirurgicaux.

Nous pouvons citer par exemple :

- Le problème de la réduction des radiations administrées lors d'un examen scanner (problème de la tomographie locale).
- La chirurgie assistée par ordinateur (incluant des problèmes de segmentation automatique, de recalage de données et de reconstruction temps-réel 3D).
- La détection et l'analyse de structures malignes dans des données d'échographie, mammographie, ou spectroscopie RMN (incluant par exemple l'analyse d'images texturées), etc.

L'utilisation des bases d'ondelettes en traitement d'images c'est généralisée durant les vingt dernières années [33]. Leur intérêt pour la compression et le dé-bruitage a été démontré puisqu'elles ont intégré le dernier standard de compression des images numériques JPEG 2000.

Leur application à l'imagerie médicale date de 1992 et c'est largement répandu depuis [14,32]. Dans ce contexte les ondelettes sont utilisées pour la compression et le dé-bruitage, mais aussi pour l'analyse fonctionnelle des données médicales (en vue d'établir un diagnostic), la tomographie locale, la segmentation et le rehaussement d'images, ou encore la description de textures.

Nous nous intéressons ici à la contribution des ondelettes à l'analyse et à la caractérisation des régions des vertèbres dans les images radiologiques.

### II.3.1.2. L'analyse en ondelettes :

Depuis les travaux de Grossman et Morlet [14], la transformation en Ondelettes est apparue comme un outil performant permettant de résoudre des problèmes relevant de

différents domaines d'application [45]. Très tôt, un intérêt soutenu c'est manifesté à l'égard de la Transformation en ondelettes en traitement d'images [20,32].

La notion d' "Ondelettes" ou "Wavelets" a été utilisée pour la première fois au début des années '80' par le géophysicien français J.Morlet [32] pour désigner des fonctions mathématiques utilisées dans la représentation des données sismiques. Les ondelettes sont des fonctions de base de variation multi-échelles, ou multi-résolution, utilisées dans le but de l'approximation et/ou de la compression des données.

La transformée en ondelettes décompose le signal d'entrée, équation (1), en une série de fonctions d'ondelettes  $\psi_{a,b}(t)$  qui dérivent d'une fonction mère  $\psi(t)$  donnée par des opérations de dilatation et de translation, équation (2).

$$C_{a,b} = \int_{-\infty}^{+\infty} x(t) \psi_{a,b}(t) dt \quad (1)$$

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \text{ avec } a \quad (2)$$

Où les paramètres :

- **a** est le facteur d'échelle.
- **b** est le paramètre de translation.

L'analyse par ondelettes est un outil mathématique capable de transformer un signal d'énergie finie dans le domaine spatial en un autre signal d'énergie finie dans le domaine spatio-fréquentiel.

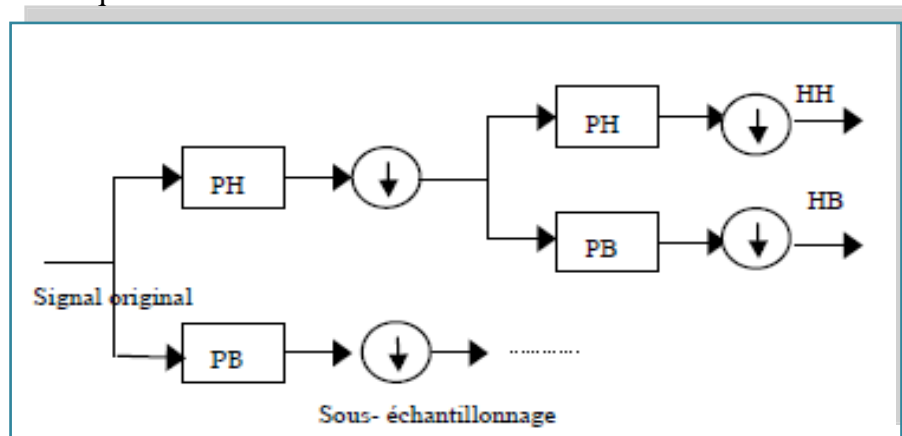
Les composantes de ce nouveau signal sont appelées *les coefficients d'ondelettes*. Ces coefficients renseignent sur la variation locale des niveaux de gris autour d'un pixel donné de l'image. Ils sont d'autant plus grands que cette variation est importante.



En 1989, Mallat a proposé un algorithme de décomposition multi-résolution basé sur la transformation en ondelettes. L'algorithme décompose une image en entrée en un ensemble d'images de détails et une image d'approximation. À chaque niveau de décomposition la taille des images transformées est réduite par un facteur de deux.

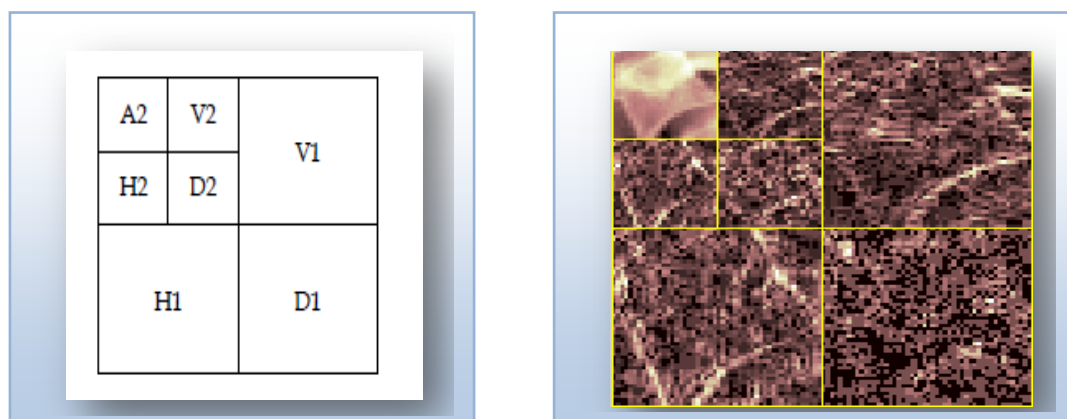
Mallat montra que la transformée en ondelette discrète (TOD) peut être implémentée grâce à un banc de filtres comprenant un filtre passe bas (PB) et un filtre passe-haut (PH).

Dans la figure 4 qui illustre ces bancs de filtre, le signal d'entrée subit un filtre passe-haut et un filtre passe-bas. Après une opération de sous-échantillonnage, le filtrage reprend sur chaque sous-bande.



**Figure 4 : Décomposition d'un signal en approximation et détails**

Pour le cas d'un signal 2D, la TOD est appliquée d'abord ligne par ligne, puis colonne par colonne. Quatre images sont alors générées à chaque niveau comme le montre la figure 5 :



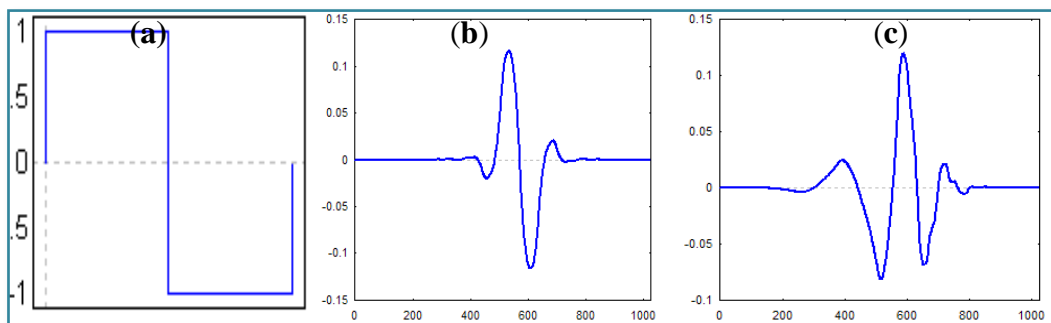
**Figure 5 : Un exemple de décomposition de l'image sur deux niveaux.**

### II.3.1.3 Choix de l'ondelette :

Dans la littérature, nous pouvons trouver une multitude de fonctions de base d'ondelettes. Le choix de l'ondelette dépend essentiellement du type de l'application (compression, segmentation, etc.). Prenons l'exemple de la classification de textures, les conclusions des auteurs ayant utilisé différents types d'ondelettes sont souvent contradictoires. En effet, de bons résultats ont été trouvés par exemple pour des ondelettes à support étendu [20], tandis que pour ce même type d'ondelettes, les résultats obtenus n'apportent aucune amélioration significative [47].

Dans notre travail, nous avons utilisé les ondelettes de Harr. Nous ne sommes pas fixés sur la recherche de l'ondelette optimale mais notre objectif est plutôt de mettre l'accent sur les phases d'extraction et d'analyse après une transformation en ondelettes sur des images gradients.

La figure 6 présente quelques exemples d'ondelettes.



**Figure 6 : Exemples classiques d'ondelettes : (a) ondelettes de Harr, (b) ondelette bi-orthogonale, (c) ondelettes de Daubechies**

### II.3.1.4 Ondelettes de Haar :

Le mérite revient à Alfred Haar d'avoir construit en 1909 des bases considérées aujourd'hui comme le fondement de la théorie des ondelettes (Meyer, 1994). Haar a défini une fonction  $h(x)$  telle que présentée dans l'équation 3 :

$$h(x) = \begin{cases} 1 & \text{pour } 0 \leq x \leq 1/2 \\ -1 & \text{pour } 1/2 \leq x \leq 1 \\ 0 & \text{ailleurs} \end{cases} \quad (3)$$

Pour  $n \geq 1$ , il a construit une base orthonormée de  $L^2 [0 ; 1]$  avec des fonctions définies par l'équation 4 :

$$\left\{ \begin{array}{l} h_n(x) = 2^{j/2} \cdot h(2^j x - k) \text{ Avec} \\ n = 2^j x + k, j \geq 0, 0 \leq k \leq 2^j \end{array} \right. \quad (4)$$

Dans l'équation (4),  $L^2 [0 ; 1]$  est l'espace des fonctions de carré intégrable sur l'intervalle  $[0 ; 1]$ . Le support de  $h_n(x)$  est l'intervalle dyadique défini par l'équation 5 :

$$I_n = [k2^{-j} ; (k + 1)2^{-j}] \subset [0; 1[ \quad (5)$$

$h_n(x)$  peut aussi s'écrire selon l'équation (6) :

$$h_n(x) = 2^{j/2} \cdot h(2^j x - k) = \frac{1}{\sqrt{2^{-j}}} h\left(\frac{x - k2^{-j}}{2^{-j}}\right) \quad (6)$$

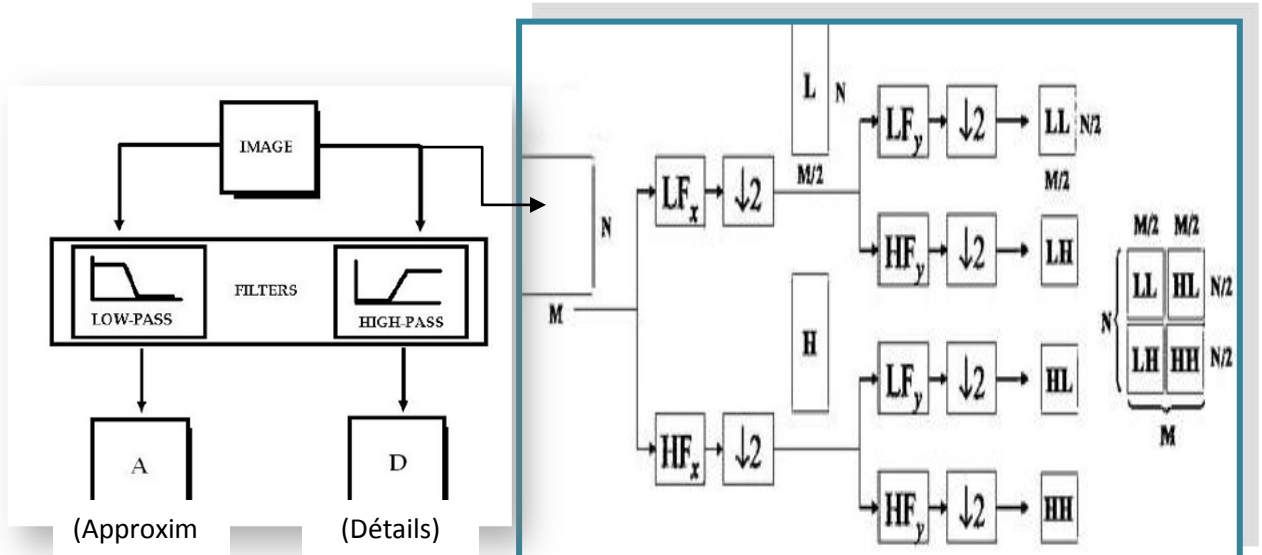
Par analogie avec la formule d'ondelettes définie précédemment, le facteur d'échelle de  $h_n$  est  $a = 2^{-j}$  et le paramètre de dilatation est  $b = k2^{-j}$ .

### II.3.1.5 Représentation par ondelettes proposée:

Notre représentation va s'intéresser aux petites régions locales contenant les vertèbres, en observant les différences d'intensités présentes.

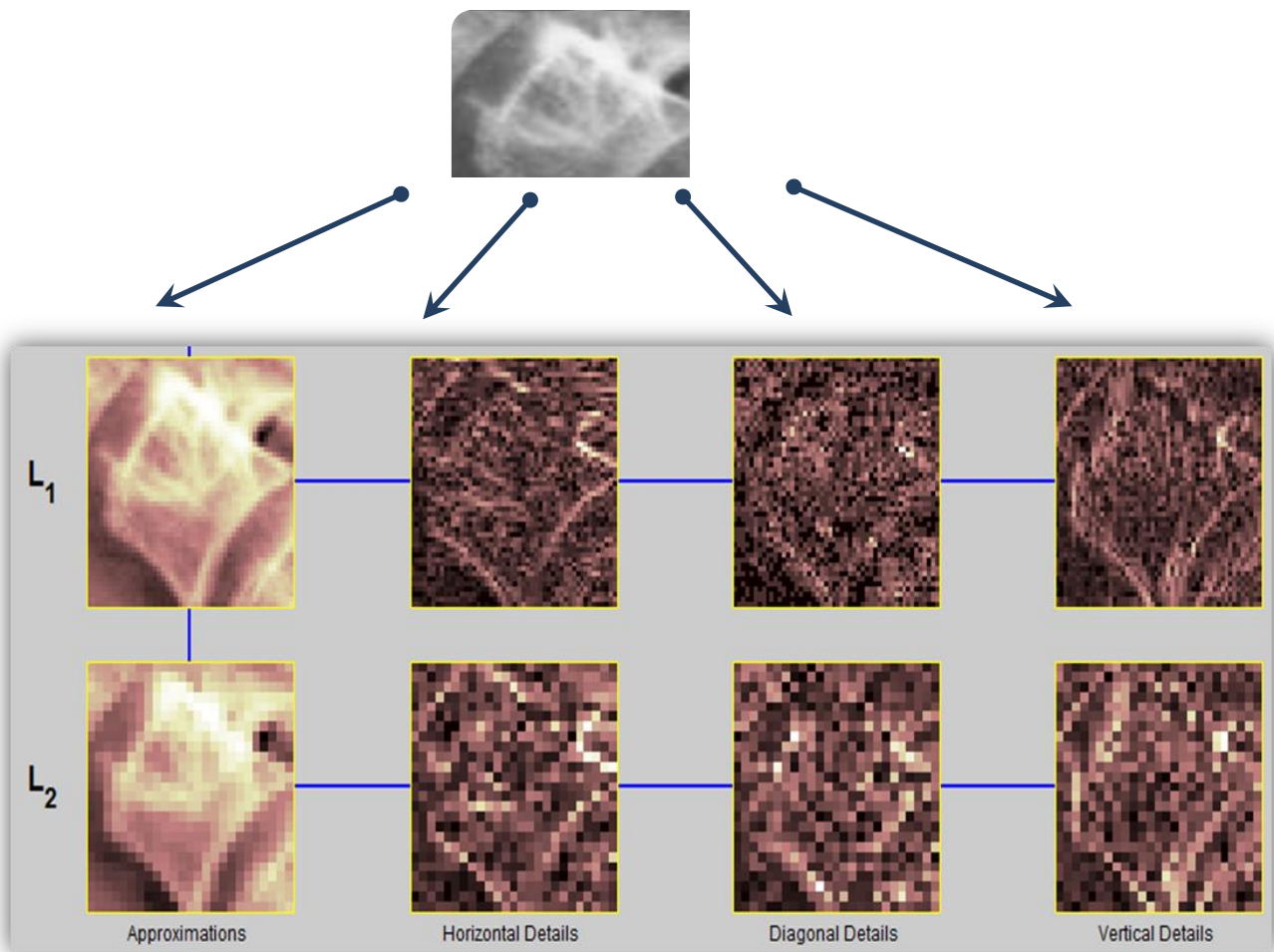
La transformation par ondelettes de Harr est appliquée à chaque image, et le résultat est un ensemble de coefficients à plusieurs échelles (Levels), qui représente les différentes orientations de différence d'intensité : verticale, horizontale, et diagonale. Par conséquent nous allons avoir trois ensembles de coefficients représentant les orientations d'ondelette. (Voir figure 7).

La figure 8 montre différentes décompositions et détails d'ondelettes pour un exemple de notre base d'apprentissage.



**Figure 7 : Décomposition de l'image par transformée en ondelette**

- LF : représente le filtre passe bas de haar [0.71, 0.71].
- HF : représente le filtre passe haut de haar [-0.71, 0.71].



**Figure 8 : Exemple d'analyse en ondelette à différents niveaux de détails aux 1<sup>ers</sup> et 2<sup>èmes</sup> niveaux**

Au même titre que la texture, la forme est généralement une description très riche d'un objet.

De nombreuses solutions ont été proposées dans la littérature pour représenter une forme. Nous distinguons deux catégories de descripteurs de formes : les descripteurs basés sur les frontières (ou contours), et les descripteurs basés sur les régions.

Les premiers font référence aux descripteurs de Fourier et portent sur une caractérisation des contours de la forme. La seconde approche fait référence, entre autres, aux moments invariants qui sont utilisés pour caractériser l'intégralité de la forme région. Ces attributs sont robustes aux transformations géométriques comme la rotation et le changement d'échelle. La méthode que nous utilisons dans cette étude concerne une description de forme à base de région par le moyen de moments géométriques.

### *II.3.2 La représentation par les moments invariants de HU :*

#### *II.3.2.1 les moments géométriques :*

Les moments géométriques [29] permettent de décrire une forme à l'aide de propriétés statistiques. Ils représentent les propriétés spatiales de la distribution des pixels dans l'image. Ils sont facilement calculés et implémentés, par contre cette approche est très sensible au bruit et aux déformations.

Les moments sont utilisés depuis longtemps pour calculer la position et le centre d'une distribution mais aussi sa variance. En vision par ordinateur, les moments permettent de calculer la position et l'orientation d'un objet.

A partir des moments géométriques, HU [28] a proposé un ensemble de sept moments invariants aux translations, rotations et changement d'échelle.

Une première utilisation des moments pour la reconnaissance de motifs géométriques a été proposée par HU [35].

Les moments offrent un cadre théorique puissant pour résoudre des problèmes rencontrés dans plusieurs applications d'imagerie y compris l'imagerie médicale, surtout lorsqu'il s'agit d'observer un objet rigide (comme les vertèbres) à différents

niveaux d'échelle, rotations et translations dans une image 2D (tomographie, projection, etc.).

Récemment, une application remarquable des moments a été présentée par [21] en combinaison avec la transformée en ondelettes pour la catégorisation des images médicales, en utilisant les cartes d'auto-organisation avec un taux de réussite de 81.8%.

### II.3.2.2. Théorie des moments :

La notion de moment en mathématiques (notamment en calcul des probabilités) a pour origine la notion de moment en physique. Le moment  $m_{pq}$  d'ordre  $n = p + q$  d'une fonction de distribution  $f(x, y)$  est définie par l'équation (7) suivante :

$$m_{pq} = \int_{-\infty}^{-\infty} \int_{-\infty}^{-\infty} x^p y^q f(x, y) dx dy \quad (7)$$

Pour une image digitale  $g(x, y)$  de taille  $M \times N$ , la formule ci-dessus devient :

$$m_{pq} = \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} x^p y^q g(x, y) \quad (8)$$

Le moment centrale est défini par :

$$\mu_{pq} = \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} (x - \bar{x})^p (y - \bar{y})^q g(x, y) \quad (9)$$

Les moments apportent différentes informations statistiques sur la forme :

- ✚ ordre 0, surface de la forme :  $m_{00}$ .
- ✚ ordre 1, centre de gravité de la forme, et est calculé par la formule (10) :

$$\bar{x} = \frac{m_{10}}{m_{00}}, \bar{y} = \frac{m_{01}}{m_{00}} \quad (10)$$

Pour caractériser la forme des images, nous utilisons la famille de moments HU [35] qui sont invariants aux différents changements d'image :

Les moments de Hu sont 7 moments issus de produits et quotients des moments centrés normés d'ordre 3, définis comme suit :

Moments d'ordres 2 :

$$\Phi_1 = m_{20} + m_{02}$$

$$\Phi_2 = (m_{20} + m_{02})^2 + 4m_{11}^2$$

Moments d'ordres 3 :

$$\Phi_3 = (m_{30} - 3m_{12})^2 + (3m_{21} - m_{03})^2$$

$$\Phi_4 = (m_{30} + m_{12})^2 + (m_{21} + m_{03})^2$$

$$\Phi_5 = (m_{30} - 3m_{12})(m_{30} + m_{12})[(m_{30} + m_{12})^2 + (m_{21} + m_{03})^2] - 6m_{11}(m_{30} + m_{12})(m_{21} + m_{03})$$

$$\Phi_6 = (m_{20} - m_{02})[(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2] - 4m_{11}(m_{30} + m_{12})(m_{21} + m_{03})$$

$$\Phi_7 = (3m_{21} - m_{03})(m_{30} + m_{12})[(m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2] - 6m_{11}(m_{30} + m_{12})(m_{21} + m_{03})$$

Les six premiers moments sont invariants aux translations, aux changements d'échelles, et aux rotations ainsi qu'aux réflexions. Or, l'invariance aux réflexions peut être problématique quand il s'agit de reconnaître des images « miroirs ». C'est pourquoi M. K. Hu exploite aussi dans son système le 7<sup>ème</sup> moment qui n'est pas invariant aux

réflexions. Ce dernier change de signe lorsqu'une telle transformation est appliquée à l'image et permet donc de détecter celle-ci.

### II.4.Classification SVM :

L'enjeu essentiel de l'apprentissage artificiel est l'aptitude à généraliser des résultats obtenue à partir d'un échantillon limitée. Dans ce travail, nous choisissons la méthode à base des machines à vecteurs de support comme un moyen opérationnel pour ce problème.

Nous allons présenter dans cette section le fondement théorique de cette méthode ainsi que sa mise en œuvre dans notre problème de détection des régions vertèbres dans les images à rayons-X.

#### II.4.1 Notions sur l'apprentissage statistique :

##### 2.4.1.1 Généralités :

D'après Mari & Napoli (1996) [1], effectuer une classification, c'est mettre en évidence, d'une part, des relations entre des objets et, d'autre part les relations entre ces objets et leurs paramètres. Il s'agit de construire une partition de l'ensemble des objets en un ensemble de classes qui soient les plus homogènes possible.

La classification, a donc deux objectifs à atteindre :

- Trouver un modèle capable de représenter la répartition des données (catégorisation).
- Définir de manière formelle l'appartenance à l'une ou l'autre des classes de toute nouvelle donnée (généralisation).

En effet la classification a pour but de réduire l'espace de recherche dans une base de données lors du processus d'identification.

En pratique, on peut rencontrer deux catégories de problèmes de classification, la classification supervisée et non-supervisée :

- **Apprentissage supervisé** : dans ce type d'apprentissage, on cherche à :
  - Estimer une fonction  $f(x)$  qui est la relation entre les objets et leurs classes.
  - Les objets utilisés comme données d'apprentissages sont accompagnés par la classe à laquelle ils appartiennent.



- Apprentissage non supervisé : on ne cherche pas cette fois à estimer une fonction mais on cherche à regrouper les objets ayant des caractéristique commune, les objets utilisés comme données d'apprentissage sont présentés sans leur classes.

### II.4.1.2 Formulation d'un problème de classification (supervisé) :

Le classificateur doit estimer une fonction  $f(x)$  qui est l'estimation de la fonction qui représente la relation entre l'objet et sa catégorie. Cette fonction est appelée fonction de décision :

$$f: X \rightarrow Y$$

X: L'ensemble des objets à classifier (appelé espace d'entrée).

Y: L'ensemble des catégories (appelé espace d'arrivée).

### 2.4.1.3 Minimisation du risque structurel :

Deux types de données sont utilisés pour un problème d'apprentissage : les données d'entraînement (données d'origine pour calculer le modèle) et les données de test (pour évaluer la performance de généralisation du modèle).

La qualité de ce modèle est alors jugée par rapport à sa capacité à réduire l'erreur de test ou de « généralisation ». Cependant, comme le modèle n'est pas construit en utilisant l'ensemble de test, l'erreur de généralisation ne peut pas être évaluée exactement car elle dépend de la distribution de probabilité des données.

Suivant la théorie de Vapnik [2], nous supposons que les données sont générées selon une distribution de probabilité inconnue  $P(x,y)$ . De plus, nous supposons que les données sont indépendantes et identiquement distribuées (iid).

L'erreur moyenne commise sur toute la distribution  $P(x,y)$  par la fonction  $f(x)$  est donnée par:

$$R[f] = \int \frac{1}{2} Q(x) dP(x, y) \quad (11)$$

Où:

- $Q$  : est la fonction d'erreur (erreur absolue dans le cas des SVM).
- $x$  : est le vecteur d'entrée.
- $y$  : est l'ensemble des classes.

Ainsi la fonction  $f$  devra être optimale : la fonction  $f^{opt}$  devra être calculée de sorte que l'erreur moyenne sur toute la distribution soit minimale.

$$f^{opt} = \arg \min_f (R[f]) \quad (12)$$

Le critère formulé dans (12) est malheureusement inutilisable en pratique. En effet, pour calculer le risque, nous devrions disposer d'une estimation de la distribution  $P(x,y)$ , ce qui n'est pas le cas. La seule information dont nous disposons comme évaluation de l'erreur est l'erreur d'entraînement appelée Risque empirique :

$$R_{emp} = \frac{1}{l} \sum_{i=1}^l \frac{1}{2} Q(x_i) \quad (13)$$

Où :

- $l$  est le nombre d'objets d'entraînements.

Ceci n'est pas suffisant. La raison en est que l'on peut facilement trouver un modèle minimisant l'erreur d'entraînement mais pour lequel l'erreur de généralisation sera très grande. Donc cette dernière est liée à la famille de fonction utilisée comme modèle. Cette dépendance est nommée « risque structurel ».

Dans la théorie de l'apprentissage statistique, Vapnik et Chervonenkis ont prouvé qu'il est possible de définir une majoration du risque structurel en fonction de la famille de fonctions utilisée pour le modèle [2]. L'une de ces majorations peut être calculée en utilisant la dimension de Vapnik–Chervonenkis (dimension VC) qui représente le plus grand nombre de points pouvant être séparés de toutes les façons possibles par un membre de l'ensemble de fonctions de  $F$ . Cela veut dire qu'il doit exister une configuration de  $h (=VC(F))$  points, telle que les fonctions  $f \in F$  peuvent leur assigner les  $2^h$  combinaisons des labels (classes) possibles.