Modélisation Hiérarchique de Données Multidimensionnelles dans des Espaces Régulièrement Décomposés : Tome 4 : Synthèse et Perspectives (2016 -2018)

Olivier Guye

MODÉLISATION HIÉRARCHIQUE DE DONNÉES MULTIDIMENSIONNELLES DANS DES ESPACES RÉGULIÈREMENT DÉCOMPOSÉS

TOME 4 : SYNTHÈSE ET PERSPECTIVES

(2016 – 2018)

- 2019 -

Olivier Guye

Résumé du tome 4 :

Ce quatrième et dernier tome a pour objectif de détailler les travaux envisagés dans un projet présenté dans le tome précédent. Il porte sur une nouvelle approche dédiée au codage des images fixes et animées, établissant ainsi un pont entre les corps de normes MPEG-4 et MPEG-7.

Ce projet a pour objectif de définir les principes du codage vidéo auto-descriptif. Pour les établir le document est décomposé en cinq chapitres qui détaillent les diverses techniques envisagées pour mettre au point une telle approche en codage visuel:

- segmentation d'images,
- calcul de descripteurs visuels,
- calcul de regroupements perceptuels,
- construction de dictionnaires visuels,
- codage d'images et de vidéos.

Fondé sur les techniques de calcul multirésolution, il se propose de développer une segmentation d'image en composantes régulières par morceaux, de calculer des attributs portant sur le support et le rendu des formes ainsi produites, indépendamment des transformations géométriques que celles-ci peuvent subir dan le plan image, et de les assembler en groupements perceptuels de manière à pouvoir mettre en œuvre une reconnaissance des formes en parties cachées.

Grâce à la quantification vectorielle du support et du rendu des formes, il apparaîtra que les formes simples peuvent être assimilées à un alphabet visuel et que les formes complexes deviennent alors des mots rédigés sur cet alphabet qui pourront être enregistrés dans un dictionnaire. A l'aide d'un balayage au plus proche voisin appliqué sur les formes de l'image, l'encodage auto-descriptif produira alors une phrase formée de mots rédigés à partir de l'alphabet des formes simples.

Mots-clés : modèles de perception en vision artificielle, analyse d'images mono- et multispectrales, analyse multirésolution, variétés surfaciques, localisation, calcul d'attributs invariants aux transformations géométriques, indexation, mesure de similarité, groupements perceptuels, codage auto-descriptif, dictionnaires de primitives visuelles, reconnaissance statistique des formes, reconnaissance structurelle des formes, images généralisées

Domaines : Modélisation et simulation, Algorithme et structure de données, Traitement des images, Vision par ordinateur et reconnaissance des formes, Apprentissage

Table des matières

Introduction

Ce quatrième et dernier tome vient détailler les travaux envisagés dans un projet présenté dans le tome précédent. C'est un projet qui porte sur une nouvelle approche pour le codage des images fixes et animées, cherchant à établir un pont entre les corps de normes MPEG-4 et MPEG-7. Ce projet a pour objectif de définir les principes d'un codage vidéo auto-descriptif.

Le paradigme envisagé repose sur un modèle de représentation en couches qui structurera l'information présente dans les images fixes et les vidéos, c'est-à-dire l'information formée par les lignes de pixels, les structures géométriques, les relations entre régions et les objets visuels. Ces différentes sortes d'informations se complémentent et peuvent être retrouvées l'une par rapport à l'autre selon une démarche progressive menant naturellement vers une organisation pyramidale de l'information visuelle.

Le premier objectif est alors le développement d'un modèle de représentation de niveau intermédiaire qui permet de manipuler le contenu d'images et de vidéos à partir de leurs composantes homogènes. Une description géométriquement invariante pour ces différentes composantes est aussi définie.

Ces composantes sont des primitives visuelles constituées par apprentissage à partir du contenu acquis, puis enregistrées dans un dictionnaire. Elles peuvent être gérées indépendamment les unes des autres ou regroupées entre elles de manière à rendre plus facile l'édition et la recherche basées sur le contenu.

Des descripteurs de forme sont utilisés pour coder, enregistrer et accéder aux différentes primitives visuelles et éventuellement au contenu complet des images fixes et animées.

1. Description des travaux

Le programme de travail est divisé en différentes étapes permettant de mettre au point les technologies nécessaires à la mise en œuvre des objectifs du projet de recherche.

Il comporte les étapes suivantes :

- segmentation d'images,
- calcul de descripteurs visuels,
- calcul de regroupements perceptuels,
- construction de dictionnaires visuels,
- codage d'images et de vidéos.

La première étape va s'intéresser à la structure des images interprétées comme des surfaces régulières par morceaux et tenter d'identifier l'ordre de ces morceaux surfaciques en distinguant les singularités dans les irrégularités rencontrées. Elle va permettre de développer une segmentation d'image, non pas guidée par la recherche de points adjacents selon une topologie métrique, mais plutôt privilégier les relations ultra-métriques des régions issues de la segmentation par morceaux : la notion d'objet visuel est alors remplacée par celle de visème qui est une unité homogène et compacte de plus petite taille.

La seconde étape cherche à mesurer ces nouvelles formes visuelles en utilisant les moments généralisés, qui permettent de localiser et de fournir des mesures invariantes aux transformations géométriques que le support de ses formes simples peuvent subir dans le plan de l'image, et la réduction des expressions analytiques du rendu de ces formes dans l'espace tridimensionnel associé au morceau surfacique correspondant, sachant qu'un visème comportera autant de formes tridimensionnelles qu'il y a de bandes spectrales dans l'image.

L'étape suivante s'intéresse aux formes composées obtenues par agrégation hiérarchique des formes simples dans l'image sous la forme de regroupements perceptuels ([15],[16]) qui seront plus proches de la notion classique d'objets visuels et que permettront de mettre en œuvre une technique de reconnaissance des formes en parties cachées à mi-chemin des reconnaissances des formes statistique ([9]) et structurelle ([10]). Le calcul des descripteurs associés aux formes simples sera étendu aux formes complexes pour produire un seul mode de calcul de descripteurs quelle que soit la nature de la forme.

Puis notre attention se portera sur la quantification de ses descripteurs qui serviront de clés pour retrouver l'information visuelle dans une base de données et les techniques de synthèse employées pour régénérer une image à l'aide de l'enregistrement des formes en base de données. Il apparaîtra qu'une fois quantifiées les formes simples constituent un alphabet visuel et que les bases de données dans lesquelles les formes complexes sont enregistrées sont alors des dictionnaires de mots rédigés sur cet alphabet visuel.

Enfin pour encoder les formes présentes dans une image, nous nous intéresserons aux courbes qui remplissent le plan et plus particulièrement à la courbe de Peano-Hilbert ([8]) qui a pour propriété de produire un parcours au plus proche voisin dans l'espace qu'elle qu'en soit sa dimension. Il apparaîtra alors que le parcours forme une phrase sur le dictionnaire des formes précédemment décrit et que leur syntaxe devrait pouvoir permettre de désigner plus globalement la nature d'une scène et le type de plan employé pour réaliser sa prise de vue.

2. Segmentation d'images

2.1. Introduction

Le projet se base sur les travaux menés dans le passé à ADERSA pour mettre au point une nouvelle technique de modélisation fondée sur une régression multiple par morceaux en contrôle des processus continus ([28]).

Il s'agit d'une technique de régression basée sur le découpage récursif d'un ensemble de données appliqué orthogonalement à son grand axe d'inertie par l'hyperplan passant au centre de gravité du nuage de points associé, jusqu'à obtenir une erreur d'approximation minimale.

Le résultat de ce découpage conduit à organiser les données sous la forme d'un arbre binaire où les données sont regroupées en sous-ensembles de données voisines vérifiant un même modèle linéaire pour une erreur d'approximation donnée.

Cette approche relève du domaine des arbres de classification et de régression ou CARTs (« Classification And Regression Trees »).

2.2. Fondements expérimentaux d'une décomposition régulière par morceaux

2.2.1. Prise en compte des irrégularités d'un modèle

Ce procédé de modélisation a été utilisé pour mettre au point des lois de commande paramétrique en matière de contrôle de processus continus.

Pour obtenir une linéarisation d'un système, les données sont rééchantillonnées en introduisant autant de retards sur la sortie que nécessaires afin de prendre en compte l'ordre du système physique lorsque celui-ci est connu. Cela permet d'introduire un schéma aux différences finies parmi les données initiales et d'estimer le comportement dynamique du système, notamment en vitesse et en accélération.

A l'usage cette méthode d'approximation générait des problèmes de continuité lorsqu'on s'éloignait des centres de gravité des nuages pour passer d'un sous-domaine à une autre voisin dans le découpage récursif des données de modélisation.

Pour réduire la génération d'irrégularités dans l'application d'un modèle linéaire par morceaux, il a été mis au point une technique de raccord continu reposant sur l'interpolation barycentrique des données estimées sur deux morceaux connexes dans l'arbre de modélisation au voisinage de leur hyperplan séparateur.

2.2.2. Raccordement de modèles linéaires adjacents

Le raccord continu est calculé comme la composition de formes d'ordre directement supérieur de la manière suivante :

$$f = \frac{|\overrightarrow{C_g M} . \overrightarrow{C_g C_d}|}{\|\overrightarrow{C_g C_d}\|^2} \cdot f_g + \frac{|\overrightarrow{MC_d} \cdot \overrightarrow{C_g C_d}|}{\|\overrightarrow{C_g C_d}\|^2} \cdot f_d$$

où M est le point de l'espace dont on veut connaître la valeur estimée,

C_g et C_d sont les centres de gravité des nuages de points associés aux fils gauche et droit d'un nœud dans l'arbre de modélisation,

f_g et f_d sont les formes de régression calculées sur ces deux nuages.

Cela s'apparente à un schéma d'ajustement multidimensionnel d'une fonction B-Spline dont les points d'appui sont les centres de gravité des nuages de points connexes selon leur hyperplan séparateur. Il génère un nouveau modèle d'ordre directement supérieur aux deux modèles, qu'il agrège en régularisant le passage de l'un à l'autre au voisinage de leur frontière commune, et cela en toutes les variables des deux modèles initiaux, c'est-à-dire de manière multiple.

Si les données de modélisation suivent des formes infiniment continues et différentiables alors le schéma agrégatif peut être appliqué de manière récursive jusqu'à la racine de l'arbre de modélisation : le nombre de niveaux de l'arbre définira l'ordre du modèle continu ainsi généré.

En pratique, il a été mis en œuvre de manière contrôlée pour qu'il ne s'applique qu'au voisinage des hyperplans séparateurs du modèle linéaire par morceaux et il a donné satisfaction aux utilisateurs pour réduire les rebonds dans l'application de lois de commande.

2.2.3. Génération de modèles continus d'ordre quelconque

La production d'un modèle linéaire par morceaux régularisé jusqu'à l'ordre permis par la décomposition hiérarchique d'un jeu de données numériques, est contrôlée par un seuil d'erreur fourni à la génération du modèle linéaire avant sa régularisation. Il s'agit d'une méthode d'approximation qui cherche à trouver un modèle paramétrique qui satisfasse à ce seuil.

L'ajustement local d'un modèle sur les données numérisées a permis de déterminer à quelle précision les données peuvent suivre un comportement linéaire et continu en lissant le bruit présent dans les données de modélisation.

Oublions momentanément l'existence de ce bruit et intéressons-nous au seul schéma d'interpolation offert par l'ajustement par B-Spline.

En se basant sur un jeu de données multidimensionnelles modélisées hiérarchiquement dans un espace régulièrement décomposé, appliquons maintenant directement aux données présentes aux nœuds du niveau le plus profond dans l'arbre de représentation du jeu de données et fusionnons les nœuds deux à deux en remontant progressivement dans l'arbre jusqu'à atteindre sa racine.

Au niveau le plus profond de l'arbre, ces données se présentent sur la forme d'une fonction étagée dont on pourra progressivement calculer les interpolants d'ordre supérieur en appliquant successivement en chaque variable de l'espace le schéma suivant :

$$f = \frac{x_i - x_{g,i}}{x_{d,i} - x_{g,i}} \cdot f_g + \frac{x_{d,i} - x_i}{x_{d,i} - x_{g,i}} \cdot f_d$$

où i est l'indice de cette direction et x_i la coordonnée correspondante du point M ,

$x_{g,i}$ et $x_{d,i}$ sont les coordonnées de même indice des centres de gravité des nuages de points associés aux fils gauche et droit d'un nœud paternel,

f_g et f_d sont les formes calculées récursivement à partir des fonctionnelles enregistrées dans les nœuds terminaux à l'aide de ce schéma d'interpolation.

En agrégeant ainsi les formes rencontrées en remontant dans l'arbre de cette pyramide d'interpolants, si les formes associées à deux nœuds filiaux sont identiques alors le schéma récursif d'interpolation générera la même forme au nœud paternel. Il sera alors localement atteint l'ordre maximal d'interpolation du jeu de données. Si la fonction étagée correspond à la numérisation d'une fonction continûment différentiable, on obtiendra son développement polynomial d'ordre maximal en chaque variable du repère de numérisation des données.

2.3. Décomposition régulière par morceaux d'une image

2.3.1. Représentation numérique d'une image

Une image numérique peut être représentée sous la forme d'une ou plusieurs applications de $[0, x_M] \times [0, y_N] \subset \mathbb{R}^2$, support des N lignes et M colonnes de l'image, vers :

- $[0,z_G]\subset\mathbb{R}$, pour une image monochromatique ou la luminescence d'une image couleur ;
- $[0,z_{C_R}]\times[0,z_{C_B}]\subset\mathbb{R}^2$, pour les chrominances rouge et bleue d'une image couleur ;
- $[0,z_1]\times[0,z_{C_2}]\times\ldots\times[0,z_{C_{N_S}}]\subset\mathbb{R}^{N_S}$, où N^S est le nombre de bandes spectrales d'une image multi-spectrale.

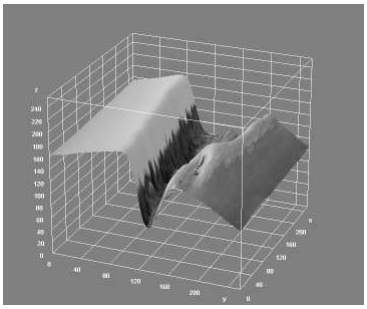
Soit $z=\begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{N_S} \end{pmatrix}$, alors la numérisation de l'image produit un ensemble de valeurs discrètes sur le support maillé de l'image:

- celui-ci est plan et s'exprime par $[0,N_C-1]\times[0,N_L-1]\subset\mathbb{N}^2$ où N_L et N_C sont les nombres de lignes et de colonnes de l'image ;
- et sur lequel reposent les valeurs de l'image numérique, appartenant à l'un des espaces possibles $[0,N_G-1]$, $[0,N_{C_R}-1]\times[0,N_{N_B}-1]$, $\prod_{k=1}^{N_S}[0,N_k-1]\subset\mathbb{N}^{N_S}$ où généralement $N_G=N_{C_R}=N_{C_B}=N_1=\cdots=N_{N_S}=256$.

A titre d'exemple, voici présentée ci-après une image couleur de laquelle ont été extraites d'une part l'image des luminescences, d'autre part les trois composantes couleur rouge, verte et bleue. L'image des luminescences constitue une image mono-chrome alors que l'ensemble des trois composantes colorées forme une image multi-chrome. Celles-ci sont présentées sous la forme d'images à niveaux de gris, puis sous la forme d'un tracé surfacique qui décrit à chaque fois la nappe des intensités lumineuses dans chaque bande de fréquences de manière à bien ressentir le caractère globalement continu des données à échelle grossière ou à mi-échelle.

Image multi-chrome

Image des luminescences

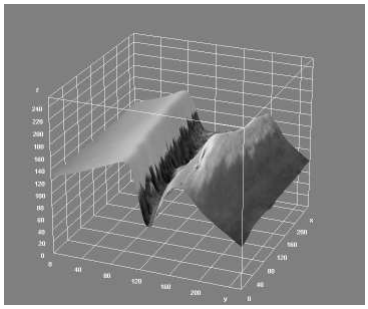


Tracé surfacique des luminescences

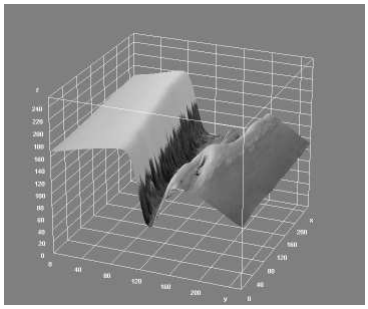
Composante rouge

Composante verte

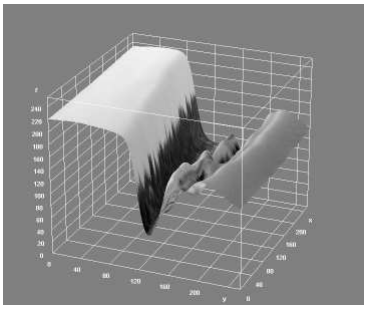
Composante bleue



Tracé surfacique de la composante rouge



Tracé surfacique de la composante verte



Tracé surfacique de la composante bleue

Figure 1 : Représentation d'une image numérique

2.5.2. Estimation linéaire de l'image au sens des moindres carrés

Une image numérique est alors la collection de points :

$$\{((x_1,y_1),z_1);\cdots;((x_{N_C\times N_L},y_{N_C\times N_L}),z_{N_C\times N_L})\}=\{((x_k,y_k),z_k)\}$$

où $k=i+(j-1)\times N_C$ avec $i\in[1,N_C], j\in[1,N_L]$ et $z_k=\begin{pmatrix} z_{1,k} \\ z_{2,k} \\ \vdots \\ z_{N_S,k}\end{pmatrix}\in[0,N_G-1]^{N_S}$

Alors l'estimée linéaire de l'image au sens des moindres carrés s'écrit :

$\hat{z}=\begin{pmatrix} f_1 \\ \vdots \\ f_l \\ \vdots \\ f_{N_S}\end{pmatrix}\begin{pmatrix}1\\x\\y\end{pmatrix}$ où les f_l sont des vecteurs lignes tels que :

$$\hat{z}=R_{z\begin{pmatrix}x\\y\end{pmatrix}}\cdot R^{-1}_{\begin{pmatrix}x\\y\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}}\cdot\left(\begin{pmatrix}x\\y\end{pmatrix}-M\begin{pmatrix}x\\y\end{pmatrix}\right)+M_z$$

où $M\begin{pmatrix}x\\y\end{pmatrix}=\frac{1}{Card(V)}\begin{pmatrix}\sum_{p\in V}x_p\\ \sum_{p\in V}y_p\end{pmatrix}=\begin{pmatrix}\bar{x}\\\bar{y}\end{pmatrix}$ et $M_z=\frac{1}{Card(V)}\begin{pmatrix}\sum_{p\in V}z_{1,p}\\ \sum_{p\in V}z_{2,p}\\ \cdots \\ \sum_{p\in V}z_{N_S,p}\end{pmatrix}=\begin{pmatrix}\bar{z}_1\\\bar{z}_2\\\cdots\\\bar{z}_{N_S}\end{pmatrix}$

et $p=\begin{pmatrix}x\\y\end{pmatrix}\in V$ point d'un sous-ensemble appartenant au support de l'image,

où $R_{z\begin{pmatrix}x\\y\end{pmatrix}}=\frac{1}{Card(V)}\begin{pmatrix}\sum_{p\in V}(z_{1,p}-\bar{z}_1)\cdot(x_p-\bar{x}) & \sum_{p\in V}(z_{1,p}-\bar{z}_1)\cdot(y_p-\bar{y}) \\ \cdots & \cdots \\ \sum_{p\in V}(z_{N_S,p}-\bar{z}_{N_S,p})\cdot(x_p-\bar{x}) & \sum_{p\in V}(z_{N_S,p}-\bar{z}_{N_S,p})\cdot(y_p-\bar{y})\end{pmatrix}$

et $R_{\binom{x}{y}\binom{x}{y}} = \frac{1}{Card(V)} \begin{pmatrix} \sum_{p \in V} (x_p - \bar{x})^2 & \sum_{p \in V} (x_p - \bar{x}) \cdot (y_p - \bar{y}) \\ \sum_{p \in V} (y_p - \bar{y}) \cdot (x_p - \bar{x}) & \sum_{p \in V} (y_p - \bar{y})^2 \end{pmatrix}$

Comme $R_{\binom{x}{y}\binom{x}{y}}$ est une matrice symétrique définie positive, elle est alors inversible et peut être décomposée en valeurs singulières de la manière suivante :

$R_{\binom{x}{y}\binom{x}{y}} = U \Lambda U^{-1}$, où $U^{-1} = U^T$,

c'est-à-dire où Λ est la matrice des valeurs propres rangées de manière décroissante et U est la matrice des vecteurs propres de $R_{\binom{x}{y}\binom{x}{y}}$.

Étant donné que l'image repose sur un support plan, on peut encore écrire :

$$R_{\binom{x}{y}\binom{x}{y}} = \begin{pmatrix} r_{xx} & r_{xy} \\ r_{yx} & r_{yy} \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$$

avec $r_{yx} = r_{xy}$ et où

$$\lambda_1 = \frac{1}{2}\left(r_{xx} + r_{yy} + \sqrt{(r_{xx} - r_{yy})^2 + 4 r_{xy}^2}\right),$$

$$\lambda_2 = \frac{1}{2}\left(r_{xx} + r_{yy} - \sqrt{(r_{xx} - r_{yy})^2 + 4 r_{xy}^2}\right),$$

$$\theta = \arctan\left(\frac{\lambda_1 - r_{xx}}{r_{xy}}\right) \text{ à } \pi \text{ près,}$$

car λ_1 et λ_2 sont les racines du polynôme caractéristique de $R_{\binom{x}{y}\binom{x}{y}}$, c'est-à-dire les valeurs qui annulent l'équation $\lambda^2 - Tr\left(R_{\binom{x}{y}\binom{x}{y}}\right)\cdot\lambda + Det\left(R_{\binom{x}{y}\binom{x}{y}}\right) = 0$ et U une matrice de rotation d'angle θ.

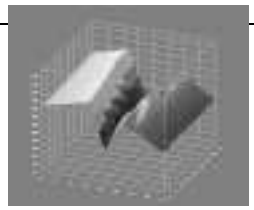
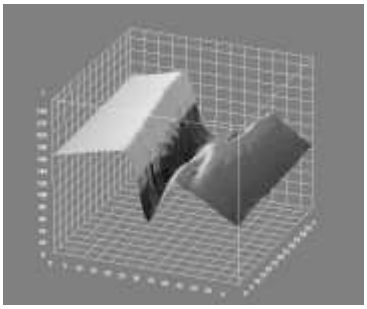
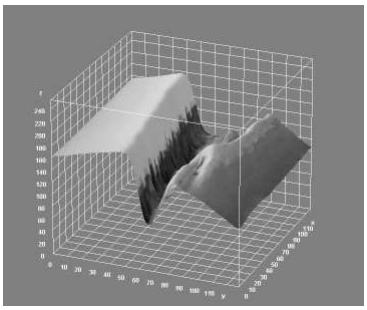
 Multi-chrome 8x8	 Mono-chrome 8x8	
 Multi-chrome 16x16	 Mono-chrome 16x16	
 Multi-chrome 32x32	 Mono-chrome 32x32	
 Multi-chrome 64x64	 Mono-chrome 64x64	
 Multi-chrome 128x128	 Mono-chrome 128x128	 Tracés surfaciques correspondant

Figure 2 : Approximation multi-résolution d'une image

Pour se rapprocher de l'effet de la mise en œuvre d'une approximation linéaire au sens des moindres carrés d'une image à différents niveaux de résolution, nous avons entrepris d'en

montrer le résultat sur un moyennage multi-résolution de cette même image, mais en se restreignant à la seule image des luminescences. Pour chaque niveau de résolution est tracé la nappe correspondante en complément de l'image moyennée.

2.3.3. Détection des irrégularités et division de l'ensemble des données

L'objectif est de localiser les irrégularités à l'ordre courant et d'en extraire une séparatrice dans le support plan de l'image permettant de diviser l'ensemble courant de données en deux sous-ensembles plus réguliers.

Partant de l'hypothèse que les singularités sont les données qui sont à la source des erreurs d'approximation les plus élevées, celles-ci sont calculées en chaque point de l'ensemble des données : $\varepsilon_p^{\infty}=\lfloor Max\{|z_{1,p}-f_1(x_p,y_p)|,\cdots,|z_{N_S,p}-f_{N_S}(x_p,y_p)|\}\rfloor$.

Ensuite l'histogramme des erreurs d'approximation est construit sur l'intervalle de valeurs entières: $[0,2^{N_G}-1]$

Puis un seuil est déterminé pour isoler les singularités des autres erreurs d'approximation de la manière suivante:

- soit l'histogramme est mono-modal et les singularités sont masquées par les erreurs d'approximation, alors le seuil est positionné devant les erreurs $\sqrt{Card(V)}$ les plus élevées en vue de pouvoir y ajuster une droite de régression;
- soit l'histogramme est bi- ou multimodal, alors le seuil sera défini par la vallée d'indice le plus élevé pour distinguer les singularités des autres erreurs que l'on cherchera progressivement à réduire.

Soit V_S le sous ensemble de V où les erreurs d'approximation sont les plus importantes, notons par :

- $\bar{x}_S=\frac{1}{Card(V_S)}\sum_{p\in V}x_p$, $\bar{y}_S=\frac{1}{Card(V_S)}\sum_{p\in V}y_p$ les coordonnées du centre de V_S ,
- $\sigma_{x^2}=\frac{1}{Card(V_S)}\sum_{p\in V}(x_p-\bar{x}_S)^2$, $\sigma_{xy}=\frac{1}{Card(V_S)}\sum_{p\in V}(x_p-\bar{x})\cdot(y_p-\bar{y}_S)$ les corrélations associées.

Alors la droite de régression ajustée sur le lieu des singularités a pour équation dans le plan du support de l'image $y=\frac{\sigma_{xy}}{\sigma_{x^2}}\cdot(x-\bar{x}_S)+\bar{y}_S$, soit encore $(x-\bar{x}_S)\cdot\sigma_{xy}-(y-\bar{y}_S)\cdot\sigma_{x^2}=0$.

A l'aide de cette équation, on pourra diviser le support de V en deux sous-ensembles de données plus régulières que celle de l'ensemble initial en évaluant le signe de cette expression pour chacun de ses points :

- les points donnant une valeur négative à cette expression seront versés dans $V_-=\{p\in V/(x-\bar{x}_S)\cdot\sigma_{xy}-(y-\bar{y}_S)\cdot\sigma_{x^2}<0\}$;

- ceux offrant une valeur positive ou nulle à cette expression le seront dans $V_+=\{p\in V/(x-\bar{x}_S)\cdot\sigma_{xy}-(y-\bar{y}_S)\cdot\sigma_{x^2}\geqslant 0\}$;

- au final $V_-\cup V_+=V$ et $V_-\cap V_+=\emptyset$.

2.3.4. Décomposition linéaire par morceaux de l'image

A l'aide des résultats précédents, un processus récursif de décomposition dichotomique de l'image peut être réalisé de la manière suivante :

- initialement, l'ensemble des données à analyser est formé de toutes les données de l'image, c'est-à-dire de $\{((x_k,y_k),z_k)\}$ où $k=i+(j-1)\times N_C$ avec $i\in[1,N_C], j\in[1,N_L]$ et $z_k=\begin{pmatrix}z_{1,k}\\z_{2,k}\\\vdots\\z_{N_S,k}\end{pmatrix}\in[0,N_G-1]^{N_S}$;

- l'estimée linéaire de l'ensemble de ces données au sens des moindres carrés s'écrit $\hat{z}=\begin{pmatrix}f_1\\\vdots\\f_l\\\vdots\\f_{N_S}\end{pmatrix}\begin{pmatrix}1\\x\\y\end{pmatrix}$ où les f_l sont des vecteurs lignes $f_l=(a_{1,l},a_{x,l},a_{y,l})$ et où $\hat{z}_l=a_{1,l}+a_{x,l}\cdot x+a_{y,l}\cdot y$ est l'estimée de la $l^{ième}$ bande spectrale ;

- les éventuelles irrégularités présentes dans l'ensemble des données de construction se détectent de la manière suivante :

- en construisant l'histogramme des erreurs d'approximation $H_\infty(V)=\{Card(p\in V/\epsilon_p^\infty=l), l\in[0, N_G-1]\}$ où $\varepsilon_p^\infty=Max_{l\in[1,N_G]}\{[|z_{k,p}-f_k(x_p,y_p)|]\}$,

- en sélectionnant dans V , le sous-ensemble V_S des points p dont l'erreur ε_p^∞ est supérieure à un seuil évalué sur l'histogramme $H_\infty(V)$ comme indiqué précédemment, c'est-à-dire $V_S=\{p\in V/\epsilon_p^\infty\geqslant seuil(H_\infty(V))\}$,

- puis en calculant dans le support plan de l'image la droite de régression des points appartenant à V_S , c'est-à-dire la droite d'équation : $(x-\bar{x}_S)\cdot\sigma_{xy}-(y-\bar{y}_S)\cdot\sigma_{x^2}=0$ où $\bar{x}_S$ et $\bar{y}_S$ sont les moyennes des coordonnées des points $p\in V_S$ et où σ_{x^2} et σ_{xy} sont leurs variance et covariance,

- enfin l'expression analytique de cette droite est employée pour diviser l'ensemble initial V en deux sous-ensembles V_- et V_+ tels que : $V_-=\{p\in V/a_S\cdot x+b_S\cdot y+c_S<0\}$ et $V_+=\{p\in V/a_S\cdot x+b_S\cdot y+c_S\geqslant 0\}$ où $a_S=\sigma_{xy}$, $b_S=-\sigma_{x^2}$ et $c_S=\bar{y}_S\cdot\sigma_{x^2}-\bar{x}_S\cdot\sigma_{xy}$.

Grâce à cette procédure, un ensemble V a été décomposé en deux sous-ensembles V_- et V_+ de manière à concentrer à leur frontière les irrégularités les plus fortes pour en réduire le nombre dans chaque nouveau sous-ensemble.

Le processus de division peut être réitéré sur chaque sous-ensemble issu de la décomposition jusqu'à parvenir à satisfaire globalement à la même erreur d'approximation sur l'ensemble initial de l'image : elle représentera la précision d'ajustement du modèle linéaire par morceaux de l'image ainsi généré.

Pour pouvoir appréhender les résultats que produiraient une telle démarche, la présentation multi-résolution des différents niveaux de moyennage d'une même image est complétée maintenant avec les images des différences entre la moyenne d'un niveau donné et sa valeur avant moyennage. Puis ces images des différences sont seuillées selon un mécanisme similaire à celui qui est décrit dans ce paragraphe de manière à ne conserver que les irrégularités observées dans le niveau courant et tracées sous forme surfacique. A un facteur d'échelle près, les zones irrégulières conservent une certaine stabilité de niveau en niveau, ce qui doit permettre de pouvoir caler progressivement les séparatrices sur les crêtes de celles-ci en analysant successivement chacun de ses niveaux jusqu'à atteindre la pleine résolution de l'image initiale. Cela reviendrait à passer d'un schéma d'ondelettes conventionnelles à un schéma d'ondelettes géométriques en cherchant à localiser les aires

sur lesquelles l'image à un comportement régulier afin d'être approchées par des développements limités de fonctionnelles.

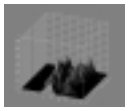
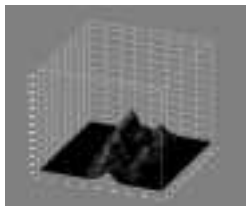
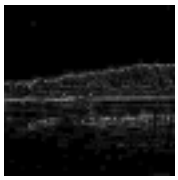
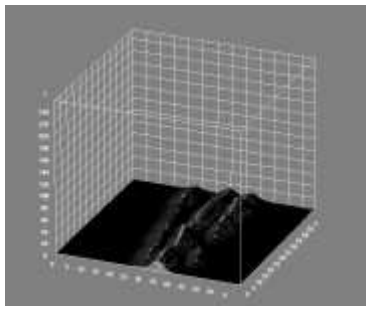
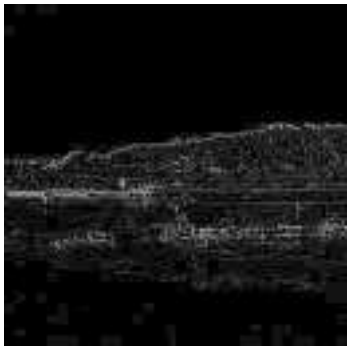
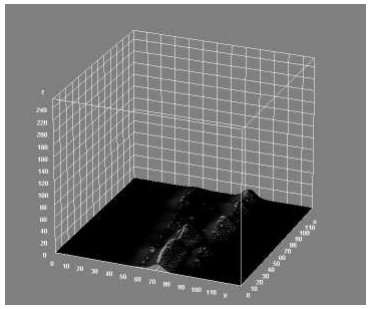
 Différences 8x8	 Seuillées 8x8	
 Différences 16x16	 Seuillées 16x16	
 Différences 32x32	 Seuillées 32x32	
 Différences 64x64	 Seuillées 64x64	
 Différences 128x128	 Seuillées 128x128	 Tracés surfaciques correspondant

Figure 3 : Détection des irrégularités dans une image

2.3.5. Agrégation à l'ordre supérieur de deux morceaux d'image

En commençant ce processus itératif sur l'image entière, on obtient progressivement une décomposition hiérarchique linéaire par morceaux sous la forme d'un arbre binaire de sous-ensembles continus et différentiables à l'ordre 1.

Pour distinguer les continuités d'ordre supérieur des singularités présentes dans l'image, il est possible de lever l'incertitude concernant l'origine de ces irrégularités en agrégeant successivement les modèles de la structure arborescente en modèles d'ordre directement supérieur. En appliquant le principe aux deux modèles issus d'un même nœud de l'arbre, les nœuds non terminaux sont alors associés avec de nouveau modèles agrégeant ceux trouvés au niveau des fils de ce nœud. Il est nécessaire d'évaluer l'erreur d'approximation du nouveau modèle sur la réunion des données présentes dans les deux morceaux associés aux deux nœuds à agréger. Si l'erreur du modèle associé respecte la précision d'ajustement recherchée, alors il s'agit bien d'une irrégularité d'ordre supérieure et non d'une singularité divisant le jeu de données localement en deux : les deux nœuds filiaux peuvent être remplacés par le nœud paternel auquel on attribuera comme modèle le nouveau modèle issu de l'agrégation des deux anciens modèles. Si cela n'est pas le cas, alors l'agrégation doit s'interrompre et les deux nœuds filiaux sont conservés en l'état avec leurs propres modèles.

Il existe une troisième possibilité après avoir commencé à agréger des modèles : il est aussi possible que chacun des modèles associés aux deux nœuds filiaux s'appliquent aussi à la réunion des données associées à ces deux nœuds : cela signifie que l'ordre du modèle ne croît temporairement plus et que les deux modèles filiaux sont équivalents à la précision d'ajustement recherchée. Les deux nœuds filiaux peuvent être agrégés en un seul en lui associant le modèle du nœud offrant la meilleure erreur d'ajustement sur l'ensemble des deux nœuds ou encore la moyenne des deux modèles filiaux. Le phénomène peut s'observer plusieurs fois de suite en remontant vers la racine dans l'arbre de modélisation.

L'idée d'agréger deux morceaux d'un certain ordre pour en produire un nouveau d'ordre directement supérieur sur la réunion de leur deux supports repose sur l'exemple des polynômes de Bernstein qui permettent de mettre progressivement en place un schéma d'approximation polynomiale sur un réseau de point représentant une courbe dont on veut connaître l'expression analytique ([7]).

Le schéma est le suivant pour une courbe décrite par $n+1$ points $P_0,\ldots,P_n\in\mathbb{R}^m$, $m\geqslant 2$:

$\forall i\in[0,n]$, $\forall t\in[0,1]$, $B(t)=\sum_{k=0}^{n} P_k^0 B_k^n(t)=\sum_{k=0}^{n-i} P_k^i B_k^{n-i}(t)$, où les points P_i^j sont définis par récurrence comme les moyennes barycentriques $P_i^j=(1-t)P_i^{j-1}+t\,P_{i+1}^{j-1}$ et en appliquant la propriété des polynômes de Bernstein $B_i^n(t)=(1-t)B_i^{n-1}+t\,B_{i+1}^{n-1}$.

Ce schéma est à la base de la construction des courbes de Bézier et de leur version composite les B-splines.

Grâce à la création de points intermédiaires par moyenne barycentrique, les polynômes de Bézier sont bâtis sur une succession de combinaisons convexes qui entraîne que la courbe correspondante reste à l'intérieur de l'enveloppe convexe des points de la porteuse de cette courbe. Cela permet de contrôler les variations de la courbe polynomiale et de lui assurer une bonne régularité.

Pour ce qui concerne les surfaces, il faut introduire la notion de carreau bi-paramétrique $0 \leqslant u \leqslant 1$ et $0 \leqslant v \leqslant 1$ où une surface évolue dans un espace à trois dimensions et appliquer les règles des produits tensoriels : $S(u,v) = \sum_{i=0}^{n} \sum_{j=0}^{n} P_{i,j} B_i^n(u) B_j^n(v)$ pour une approximation polynomiale d'ordre n en chaque variable du carreau.

La version proposée pour procéder à l'agrégation de modèles en vue d'obtenir un modèle d'ordre supérieur reprend l'utilisation de moyennes barycentriques en s'appuyant sur les centres de gravité de chacun des sous-ensembles de données issus de la décomposition hiérarchique de l'image. Elle mène alors vers une approximation polynomiale d'ordre n non plus en chaque variable du support de la variété surfacique, mais en toutes les variables de celui-ci.

En reprenant les notations employées pour décrire la procédure de division d'un ensemble V en deux sous-ensembles V_- et V_+, appelons :

- M_- et M_+ les centres de gravité tels qu'ils ont été calculés lors du calcul de l'estimation linéaire des morceaux d'image au sens des moindres carrés ,
- $\widehat{z_-}$ et $\widehat{z_+}$ les estimées en question sur ces deux sous-ensembles,

alors le schéma d'interpolation barycentrique sur ces deux sous-ensembles V_- et V_+ génère l'estimée d'ordre directement supérieur en toutes les variables comme suivant :

$$\hat{z} = \frac{\langle \overrightarrow{M_- M} . \overrightarrow{M_- M_+} \rangle}{\| \overrightarrow{M_- M_+} \|^2} \cdot \widehat{z_-} + \frac{\langle \overrightarrow{M M_+} \cdot \overrightarrow{M_- M_+} \rangle}{\| \overrightarrow{M_- M_+} \|^2} \cdot \widehat{z_+} \quad .$$

C'est-à-dire en développant de manière matricielle :

$$\begin{pmatrix} f_1 \\ \vdots \\ f_l \\ \vdots \\ f_{N_s} \end{pmatrix} = \frac{(x-\overline{x_-})\cdot(\overline{x_+}-\overline{x_-})+(y-\overline{y_-})\cdot(\overline{y_+}-\overline{y_-})}{(\overline{x_+}-\overline{x_-})^2+(\overline{y_+}-\overline{y_-})^2}\cdot\begin{pmatrix} f_1^- \\ \vdots \\ f_l^- \\ \vdots \\ f_{N_s}^- \end{pmatrix} + \frac{(\overline{x_+}-x)\cdot(\overline{x_+}-\overline{x_-})+(\overline{y_+}-y)\cdot(\overline{y_+}-\overline{y_-})}{(\overline{x_+}-\overline{x_-})^2+(\overline{y_+}-\overline{y_-})^2}\cdot\begin{pmatrix} f_1^+ \\ \vdots \\ f_l^+ \\ \vdots \\ f_{N_s}^+ \end{pmatrix}$$

où M est le point de coordonnées $\begin{pmatrix} x \\ y \end{pmatrix} \in V = V_- \cup V_+$.

Comme dans le schéma de construction des polynômes de Bézier, le processus d'agrégation des modèles peut être appliquer de manière récursive en remontant dans l'arbre des modèles jusqu'à s'arrêter lorsque l'erreur d'approximation n'est plus maîtrisée. Si le processus se poursuit jusqu'à la racine de l'arbre alors l'image n'est formée que d'une seule variété surfacique dont l'ordre de continuité et de différentiabilité est égal à la profondeur de l'arbre de modélisation avant une quelconque opération d'agrégation.

La première passe d'agrégation va produire l'expression d'une quadrique, c'est-à-dire un polynôme d'ordre 2 en toutes les variables x et y du support de chaque fonctionnelle f_l. La seconde passe d'agrégation, l'expression d'une cubique, c'est-à-dire un polynôme d'ordre 3 pour chacune de ces mêmes fonctionnelles. Pour calculer des descripteurs de rendu en utilisant ces informations, il n'est pas nécessaire d'aller au-delà de l'ordre 3, comme cela sera présenté par la suite.

Travailler sur des estimées linéaires localement continues permet de s'affranchir dans une certaine mesure de l'influence du bruit de capture et de numérisation des données sans avoir à prétraiter celles-ci avant leur utilisation. Cette mesure est contrôlée par la précision d'ajustement recherchée sur l'erreur quadratique moyenne de l'estimée sur son support.

L'agrégation hiérarchique par des polynômes d'ordre croissant ne permet pas de fournir une segmentation d'image selon des systèmes de voisinages classiquement associés à des distances métriques, mais plutôt une partition de l'image selon l'ultra-métrique induite par le découpage arborescent de l'image. Il s'agit d'une topologie plus grossière qu'une topologie métrique qui a pour effet de produire une segmentation plus morcelée qu'une segmentation classique. Cette caractéristique ne devrait pas entraver les processus d'encodage d'images et de reconnaissance des formes.

Par contre le prédicat employé pour segmenter des images n'est pas un prédicat d'iso-coloration, mais un prédicat d'iso-modélisation.

En place d'une combinaison barycentrique de modèles, une autre méthode peut être aussi envisagée pour réaliser l'agrégation de deux modèles fraternels de l'arborescence en un nouveau modèle d'ordre supérieur, il s'agit alors de la linéarisation du problème à l'ordre supérieur des modèles initiaux et de sa résolution au sens des moindres carrés, c'est-à-dire trouver :

- l'estimation d'ordre 2 en toutes les variables qui minimise l'erreur quadratique moyenne sur le nouvel ensemble agrégé de données si les modèles initiaux sont linéaires, soit les quadriques $\hat{z}_l = a_{1,l} + a_{x,l} \cdot x + a_{y,l} \cdot y + a_{x^2,l} \cdot x^2 + a_{xy,l} \cdot xy + a_{y^2,l} \cdot y^2$,

- puis au niveau supérieur, les cubiques,

$$\hat{z}_l = a_{1,l} + a_{x,l} \cdot x + a_{y,l} \cdot y + a_{x^2,l} \cdot x^2 + a_{xy,l} \cdot xy + a_{y^2,l} \cdot y^2 + a_{x^3,l} \cdot x^3 + a_{x^2y,l} \cdot x^2 y + a_{xy^2,l} \cdot xy^2 + a_{y^3,l} \cdot y^3 ,$$

- etc.

Si les expressions linéarisées sont similaires à celles produites par combinaison barycentrique, alors il pourra être considéré que l'agrégation par combinaison barycentrique est la transformée rapide des modèles linéarisés d'ordre croissant. Comme les deux méthodes produisent des développements limités à un ordre donné en toutes les variables, il est possible de s'appuyer sur les propriétés des développements de Taylor qui assurent l'unicité de ces développements quel que soit l'ordre, lorsqu'ils existent ([6]).

Quelle que soit l'approche employée, il faudra vérifier préalablement que les modèles d'ordre inférieur ne s'appliquent pas directement à la réunion des deux supports en respectant la précision de modélisation recherchée. Si c'est le cas le meilleur modèle d'ordre inférieur est propagé au niveau supérieur dans l'arbre. Si il n'est plus possible d'agréger les modèles filiaux en remontant dans l'arbre, c'est que l'ordre maximal de modélisation a été atteint.

Après agrégation des modèles jusqu'à l'ordre p , ainsi que celle des nœuds qui leur sont associés, l'ensemble des parties résultantes forme une partition de l'image en sous-ensembles qui peuvent être approchés par des modèles d'ordre p par morceaux. Notons alors $\{V_i\}_{i \in 1, n(p)}$ où $n(p)$ est le nombre de parties de classe C^p représentant l'image initiale à la précision recherchée.

2.4. Algorithme général de décomposition régulière d'ordre quelconque d'une image

Sur la base des développements précédents, nous pouvons établir l'algorithme qui permettra de réaliser la décomposition régulière d'ordre quelconque par morceaux d'une image :

DÉBUT

A partir de tous les points de l'image constitué l'ensemble de données :

$V = \{((x_k, y_k), z_k)\}$ où $k = i + (j-1) \times N_C$

avec $i \in [1, N_C], j \in [1, N_L]$ et $z_k = \begin{pmatrix} z_{1,k} \\ z_{2,k} \\ \vdots \\ z_{N_S,k} \end{pmatrix} \in [0, N_G - 1]^{N_S}$

ordre ← 1, $\hat{z}$ ← Décomposition régulière par morceaux (V , ordre, précision)

FIN

FONCTION Décomposition régulière par morceaux (V , ordre, précision)

DÉBUT

Calcul de l'estimée linéaire de V au sens des moindres carrés :

$\hat{z} = \begin{pmatrix} f_1 \\ \vdots \\ f_l \\ \vdots \\ f_{N_S} \end{pmatrix} \begin{pmatrix} 1 \\ x \\ y \end{pmatrix}$ où les f_l sont des vecteurs lignes tels que :

$$\hat{z} = R_{z\binom{x}{y}} \cdot R^{-1}_{\binom{x}{y}\binom{x}{y}} \cdot \left(\begin{pmatrix} x \\ y \end{pmatrix} - M \begin{pmatrix} x \\ y \end{pmatrix} \right) + M_z ,$$

où $R_{\binom{x}{y}\binom{x}{y}} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$, $\theta = \arctan\left(\frac{\lambda_1 - r_{xx}}{r_{xy}}\right)$

et $\lambda_1 = \frac{1}{2}\left(r_{xx} + r_{yy} + \sqrt{(r_{xx} - r_{yy})^2 + 4r_{xy}^2}\right)$, $\lambda_2 = \frac{1}{2}\left(r_{xx} + r_{yy} - \sqrt{(r_{xx} - r_{yy})^2 + 4r_{xy}^2}\right)$

Calcul de l'erreur maximale d'approximation :

$$\varepsilon_V^{\infty} = \lfloor Max_{p \in V} \{ |z_{1,p} - f_1(x_p, y_p)|, \cdots, |z_{N_S,p} - f_{N_S}(x_p, y_p)| \} \rfloor$$

SI (ε_V^{∞} > précision) ALORS FAIRE

Calcul de l'histogramme des erreurs :

$$H_{\infty}(V) = \{ Card(p \in V / \epsilon_p^{\infty} = l), l \in [0, N_G - 1] \}$$

SI (histogramme mono-modal)

ALORS $V_S \leftarrow$ { $\sqrt{Card(V)}$ points de plus forte erreur }

SINON FAIRE

Définir comme seuil la vallée la plus élevée de l'histogramme

$$V_S \leftarrow \quad \{p \in V / \epsilon_p^{\infty} > seuil\}$$

FIN

Calculer les coordonnées $\bar{x}$ et $\bar{y}$ du centre

et les variance σ_{x^2} et covariance σ_{xy} de V_S

Diviser l'ensemble V en deux sous-ensembles :

$$V_- = \{p \in V / (x - \bar{x}) \cdot \sigma_{xy} - (y - \bar{y}) \cdot \sigma_{x^2} < 0\}$$

$$V_+ = \{p \in V / (x - \bar{x}) \cdot \sigma_{xy} - (y - \bar{y}) \cdot \sigma_{x^2} \geq 0\}$$

$\widehat{z_-}$ ← Décomposition régulière par morceaux (V_- , ordre, précision)

$\widehat{z_+}$ ← Décomposition régulière par morceaux (V_+ , ordre, précision)

SI (ordre < 3) ALORS FAIRE

Calcul de la combinaison barycentrique :

$$\hat{z} = \frac{(x - \overline{x_-}) \cdot (\overline{x_+} - \overline{x_-}) + (y - \overline{y_-}) \cdot (\overline{y_+} - \overline{y_-})}{(\overline{x_+} - \overline{x_-})^2 + (\overline{y_+} - \overline{y_-})^2} \cdot \widehat{z_-} + \frac{(\overline{x_+} - x) \cdot (\overline{x_+} - \overline{x_-}) + (\overline{y_+} - y) \cdot (\overline{y_+} - \overline{y_-})}{(\overline{x_+} - \overline{x_-})^2 + (\overline{y_+} - \overline{y_-})^2} \cdot \widehat{z_+}$$

Calcul de l'erreur d'approximation :

$$\varepsilon_V^{\infty} = \lfloor Max_{p \in V} \{|z_{1,p} - f_1(x_p, y_p)|, \cdots, |z_{N_S,p} - f_{N_S}(x_p, y_p)|\} \rfloor$$

SI (ε_V^{∞} <= précision) ALORS FAIRE

Remplacer le modèle actuellement associé à V

par ce modèle d'ordre supérieur

ordre ← ordre + 1

FIN

FIN

FIN

RETOUR ($\hat{z}$)

FIN

Dans cette version de l'algorithme, l'ordre maximal d'agrégation de modèles est limité, mais il pourrait être envisagé de ne pas le restreindre ainsi et ne conserver que les coefficients analytiques jusqu'à l'ordre 3, sachant que les coefficients d'ordre supérieur procurent une contribution moins importante à la modélisation si l'on se réfère à la formulation des accroissements de Taylor.

Avant d'agréger les modèles, il n'est pas non plus vérifié si l'un des modèles filiaux s'applique aussi aux données du nœud voisin, car il est probable que si c'est le cas l'autre modèle doit aussi s'appliquer sur les données voisines et à ce moment-là l'agrégation va produit l'équivalent de la moyenne simple des deux modèles si l'on exclut les coefficients d'ordre supérieur qui ne doivent présenter qu'une faible contribution à la qualité de la modélisation.

Enfin, parmi les tests à réaliser, il serait souhaitable de vérifier par ailleurs si la moyenne barycentrique donne un résultat similaire à l'estimation linéaire au sens des moindres carrés pour des données linéarisées à l'ordre correspondant à celui produit par l'agrégation employée dans l'algorithme précédent pour valider l'approche : il est possible que cela permette de mettre au point une transformée encore plus rapide.

Si l'on se reporte à la représentation sous forme d'ondelettes utilisée pour illustrer quels résultats pourraient être attendus en appliquant cette approche en modélisation d'image, il devrait être possible de simplifier l'algorithme qui vient d'être présenté, non en cherchant à faire une décomposition linéaire par morceaux d'une image mais en commençant à agréger sur la

représentation constante par morceaux fournie par le moyennage multi-résolution, c'est-à-dire par une fonction étagée à une erreur quadratique près en moyenne.

De manière plus générale, on remarque que la décomposition progressive du modèle linéaire global en modèles linéaires par morceaux s'arrête lorsque l'influence du bruit a été maîtrisée grâce à la méthode des moindres carrés sur l'ensemble des morceaux et que la distinction entre irrégularités et singularités s'effectue en agrégeant les modèles par interpolation à l'ordre supérieur jusqu'à ce que l'erreur d'approximation ne satisfasse plus la précision demandée : à ce moment-là, il existe localement une singularité que la méthode n'est pas parvenue à contenir.

3. Calcul de descripteurs

3.1. Description des formes visuelles

A l'issue de la segmentation, une image se présente sous la forme d'une décomposition en éléments visuels ayant les propriétés suivantes :

- chaque élément peut être représenté par une fonctionnelle ou une collection de fonctionnelles prenant appui sur un sous-ensemble connexe du support de l'image ;
- chacune de ces fonctionnelles est approchée par un polynôme d'ordre fini à une erreur près ;
- l'ensemble des supports de ces fonctionnelles forme une partition du support bidimensionnel de l'image.

On distinguera alors deux types de descripteurs de forme :

- ceux associés à la partition de l'image en morceaux ;
- ceux plus propres aux fonctionnelles qui s'appuient sur ces morceaux et que l'on peut approcher par un polynôme d'ordre fini.

La fonctionnelle et son support forment un élément visuel simple que l'on appellera encore visème.

3.2. Descripteurs associés aux supports des formes

3.2.1. Moments généralisés du support d'une forme

Les moments d'un support V à l'ordre 0, 1, 2 et 3 sont donnés par les formules suivantes sous forme continue :

- $M(0)=\iint_V dydx$,

surface du support ;

- $M(x)=\iint_V x\,dydx$, $M(y)=\iint_V y\,dydx$,

dont on extrait le centre de gravité du support ;

- $M(x^2)=\iint_V x^2\,dydx$, $M(xy)=\iint_V xy\,dydx$, $M(y^2)=\iint_V y^2\,dydx$,

décrivent l'ellipsoïde d'inertie du support dans le repère d'observation ;

- $M(x^3)=\iint_V x^3\,dydx$, $M(x^2y)=\iint_V x^2y\,dydx$,
- $M(xy^2)=\iint_V xy^2\,dydx$, $M(y^3)=\iint_V y^3\,dydx$,

exprimant les diverses asymétries du support selon les axes du repère.

Des moments d'ordre 2, sont déduits l'axe principal d'inertie et l'angle que celui-ci fait par rapport au repère d'observation. Celui-ci est calculé à 180° près et l'incertitude à 360° est levée en examinant le signe de l'asymétrie selon le grand axe après s'être replacé dans le repère propre du support.

3.2.2. Moments calculés sur une représentation cellulaire

L'expression discrétisée sur un réseau maillé régulièrement échantillonné de ces mesures est la suivante :

- $M(0)=\sum_{x,y\in V} dm$;
- $M(x)=\sum_{x,y\in V} xdm$ et $M(y)=\sum_{x,y\in V} ydm$;
- $M(x^2)=\sum_{x,y\in V} x^2dm$, $M(xy)=\sum_{x,y\in V} xydm$, $M(y^2)=\sum_{x,y\in V} y^2dm$;
- $M(x^3)=\sum_{x,y\in V} x^3dm$, $M(x^2y)=\sum_{x,y\in V} x^2ydm$, $M(xy^2)=\sum_{x,y\in V} xy^2dm$, $M(y^3)=\sum_{x,y\in V} y^3dm$,
- où $dm=1$.

3.2.3. Translation des moments au centre de gravité du support

$M(0)$ représente la surface S du support.

Les coordonnées du centre de gravité sont obtenues à l'aide des moments d'ordre 1 :

- l'abscisse du centre de gravité $x_G = M(x)/S$;
- l'ordonnée du centre de gravité $y_G = M(y)/S$.

L'invariance des moments en translation se déduit par correction de ceux-ci en fonction des coordonnées du centre de gravité du support.

Les valeurs des moments d'ordre supérieur dans le nouveau repère $(\overrightarrow{x_G X}, \overrightarrow{y_G Y})$ (cf. figure ci-dessous) deviennent :

- $M(X^2) = M(x^2) - x_G^2 S$
- $M(XY) = M(xy) - x_G y_G S$
- $M(Y^2) = M(y^2) - y_G^2 S$
- $M(X^3) = M(x^3) - 3x_G^2 M(X^2) - y_G^3 S$
- $M(X^2 Y) = M(x^2 Y) - y_G M(X^2) - 2x_G M(XY) - x_G^2 y_G S$
- $M(XY^2) = M(xy^2) - x_G M(Y^2) - 2y_G M(XY) - x_G y_G^2 S$
- $M(Y^3) = M(y^3) - 3x_G^2 M(Y^2) - y_G^3 S$

3.2.4. Rotation des moments dans les axes du support

Des moments d'ordre 2 se déduisent les axes d'inertie u_1, u_2 du support :

$$M(u_1^2)=\frac{1}{2}\left(M(X^2)+M(Y^2)+\sqrt{\left(M(X^2)-M(Y^2)\right)^2+4\,M(XY)^2}\right)$$
$$M(u_2^2)=\frac{1}{2}\left(M(X^2)+M(Y^2)-\sqrt{\left(M(X^2)-M(Y^2)\right)^2+4\,M(XY)^2}\right)$$

Cette opération revient à assimiler le support à son ellipsoïde d'inertie de grand axe $\vec{u_1}$ et son petit axe $\vec{u_2}$.

L'angle de rotation qui permet de passer dans le repère propre du support $(\overrightarrow{X_G u_1}, \overrightarrow{Y_G u_2})$, vaut :

$$\theta(\vec{X}, \vec{u_1})=arctan\left(\frac{M(u_1^2)-M(X^2)}{M(XY)}\right) \text{ à } \pi \text{ près.}$$

En effet, on en peut déduire de l'ellipsoïde d'inertie du support qu'une direction pour le grand axe et non un sens.

Le moment croisé $M(u_1 u_2)$ s'annule et les moments d'ordre 3 dans les directions u_1 et u_2 deviennent:

$$M(u_1^3)=\cos^3\theta\cdot M(X^3)+3\cdot\sin\theta\cdot\cos^2\theta\cdot M(X^2Y)$$
$$+3\cdot\sin^2\theta\cdot\cos\theta\cdot M(XY^2)+\sin^3\theta\cdot M(Y^3)$$
$$M(u_1^2u_2)=-\sin\theta\cdot\cos^2\theta\cdot M(X^3)+\cos^3\theta\cdot M(X^2Y)$$
$$-2\cdot\sin^2\theta\cdot\cos\theta\cdot M(X^2Y)+2\cdot\sin\theta\cdot\cos^2\theta\cdot M(XY^2)$$
$$-\sin^3\theta\cdot M(XY^2)+\sin^2\theta\cdot\cos\theta\cdot M(Y^3)$$
$$M(u_1u_2^2)=\sin^2\theta\cdot\cos\theta\cdot M(X^3)-2\cdot\sin\theta\cdot\cos^2\theta\cdot M(X^2Y)$$
$$+\sin^3\theta\cdot M(X^2Y)+\cos^3\theta\cdot M(XY^2)$$
$$-2\cdot\sin\theta\cdot\cos^2\theta\cdot M(XY^2)+\sin\theta\cdot\cos^2\theta\cdot M(Y^3)$$
$$M(u_2^3)=-\sin^3\theta\cdot M(X^3)+3\cdot\sin^2\theta\cdot\cos\theta\cdot M(X^2Y)$$
$$-3\cdot\sin\theta\cdot\cos^2\theta\cdot M(XY^2)+\cos^3\theta\cdot M(Y^3)$$

Le sens des axes $\vec{u_1}$ et $\vec{u_2}$ est fixé en forçant le moment $M(u_1^3)$ à une valeur positive.

Cela équivaut à donner pour sens à $\vec{u_1}$ celui qui offre la plus forte asymétrie selon cet axe.

L'asymétrie selon le grand axe d'un support est une caractéristique propre à la forme de celui-ci et est stable une fois déterminées les directions des axes d'inertie.

L'incertitude à 360° près est alors levée ainsi :

$$M(u_1^3) < 0 \Rightarrow \theta = \theta + \pi$$

et les moments d'ordre 3 sont corrigés par un changement de signe :

$$M(u_1^3) = -M(u_1^3)$$
$$M(u_1^2 u_2) = -M(u_1^2 u_2)$$
$$M(u_1 u_2^2) = -M(u_1 u_2^2)$$
$$M(u_2^3) = -M(u_2^3)$$

Ces valeurs représentent les différentes asymétries du support dans son repère propre.

3.2.5. Caractéristiques invariantes des supports des morceaux surfaciques

Nous venons de décrire comment obtenir :

- les coordonnées du centre de gravité X_G, Y_G du support d'une forme;
- l'angle θ que fait le support d'une forme dans le référentiel de l'image.

Les autres moments calculés dans le référentiel propre de la forme (X_G, Y_G, θ) permettent de disposer des caractéristiques spatiales du support invariantes en translation et rotation dans le plan de la prise de vue.

Ces caractéristiques sont :

- $M(0)$ la surface du support ;
- $M(u_1^2), M(u_2^2)$ les inerties du support ;
- $M(u_1^3), M(u_1^2 u_2), M(u_1 u_2^2), M(u_2^3)$ les asymétries du support.

Ces caractéristiques deviennent invariantes aux homothéties en abandonnant la surface du support et en normalisant les moments restant par la valeur du grand axe d'inertie. Il reste alors :

- $\epsilon = \frac{M(u_2^2)}{M(u_1^2)}$ l'excentricité du support de la forme;
- $\frac{M(u_1^3)}{M(u_1^2)}, \frac{M(u_1^2 u_2)}{M(u_1^2)}, \frac{M(u_1 u_2^2)}{M(u_1^2)}, \frac{M(u_2^3)}{M(u_1^2)}$ les asymétries mises à l'échelle .

Il faudra rajouter aux informations de localisation du support de la forme le facteur d'échelle constitué par le grand axe d'inertie $M(u_1^2)$ pour pouvoir replacer correctement celui-ci dans le plan de l'image.

3.3. Descripteurs du rendu des formes

3.3.1. Développement analytique des morceaux

Dans l'objectif de produire des mesures permettant de décrire les morceaux surfaciques issus de la décomposition régulière par morceaux d'une image numérique, nous nous focaliserons sur les modèles d'ordre au plus 3 pour calculer ces mesures ou à leur développement jusqu'à l'ordre 3.

Pour chaque sous-ensemble V_j , $j \in 1, n(3)$ de la partition de l'image par décomposition régulière jusqu'à l'ordre 3, chacune des bandes spectrales est approchée par la cubique :

$$\hat{z}_l^j = a_{1,l}{}^j + a_{x,l}^j \cdot x + a_{y,l}^j \cdot y + a_{x^2,l}^j \cdot x^2 + a_{xy,l}^j \cdot xy + a_{y^2,l}^j \cdot y^2 + a_{x^3,l}^j \cdot x^3 + a_{x^2y,l}^j \cdot x^2 y + a_{xy^2,l}^j \cdot xy^2 + a_{y^3,l}^j \cdot y^3$$

En oubliant les indices j et l , l'expression se simplifie en :

$$\hat{z} = a_1 + a_x \cdot x + a_y \cdot y + a_{x^2} \cdot x^2 + a_{xy} \cdot xy + a_{y^2} \cdot y^2 + a_{x^3} \cdot x^3 + a_{x^2y} \cdot x^2 y + a_{xy^2} \cdot xy^2 + a_{y^3} \cdot y^3$$

3.3.2. Translation des expressions analytiques aux centres de gravité des morceaux

Si $\begin{pmatrix} \bar{x}_j \\ \bar{y}_j \end{pmatrix}$ est le centre de gravité du morceau V_j, son estimée dans la $l^{ième}$ bande spectrale vaudra $\hat{z}_l^j$ = estimée d'ordre 3 en $\begin{pmatrix} \bar{x}_j \\ \bar{y}_j \end{pmatrix}$ qui s'écrira de manière simplifiée $\begin{pmatrix} \bar{z} \\ \bar{x} \\ \bar{y} \end{pmatrix}$. C'est-à-dire que l'on ne s'intéresse plus directement à l'image mais à son modèle au voisinage du centre de gravité du support de chaque morceau d'image.

Pour cela, plaçons-nous dans le repère du plan tangent de l'estimée d'une bande spectrale au centre de gravité M_j du morceau V_j.

3.3.3. Description de la forme dans le repère de son plan tangent

Pour y parvenir, il faut enchaîner une translation en $\begin{pmatrix} \bar{z} \\ \bar{x} \\ \bar{y} \end{pmatrix}$, puis une rotation définie à partir de l'équation du plan directeur déduit des coefficients d'ordre 1 de l'estimée, c'est-à-dire tel que le plan directeur ait pour équation en $\begin{pmatrix} \bar{z} \\ \bar{x} \\ \bar{y} \end{pmatrix}$ $Z = a_x \cdot X + a_y \cdot Y$ où $Z = z - \hat{z}$, $X = x - \bar{x}$, $Y = y - \bar{y}$ et donc pour coefficients directeurs :

- $\tan(\theta_{XZ}) = a_x$ dans le plan $\left(\begin{pmatrix} \bar{z} \\ \bar{x} \\ \bar{y} \end{pmatrix}, \vec{X}, \vec{Z}\right)$;
- $\tan(\theta_{YZ}) = a_y$ dans le plan $\left(\begin{pmatrix} \bar{z} \\ \bar{x} \\ \bar{y} \end{pmatrix}, \vec{Y}, \vec{Z}\right)$.

Ainsi se déplacer dans le nouveau repère associé au plan tangent de la surface centrée au centre de gravité du support du morceau V_j revient à appliquer la transformation linéaire suivante à chaque point de la surface concernée :

$$\begin{pmatrix} Z \\ X \\ Y \end{pmatrix} = \begin{pmatrix} \cos(\theta_{XZ}) & \sin(\theta_{XZ}) & 0 \\ -\sin(\theta_{XZ}) & \cos(\theta_{XZ}) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos(\theta_{YZ}) & 0 & \sin(\theta_{YZ}) \\ 0 & 1 & 0 \\ -\sin(\theta_{YZ}) & 0 & \cos(\theta_{YZ}) \end{pmatrix} \begin{pmatrix} z-\bar{z} \\ x-\bar{x} \\ y-\bar{y} \end{pmatrix},$$

soit encore :

$$\begin{pmatrix} Z \\ X \\ Y \end{pmatrix} = \begin{pmatrix} \cos(\theta_{XZ})\cos(\theta_{YZ}) & \sin(\theta_{XZ}) & \cos(\theta_{XZ})\sin(\theta_{YZ}) \\ -\sin(\theta_{XZ})\cos(\theta_{YZ}) & \cos(\theta_{XZ}) & -\sin(\theta_{XZ})\sin(\theta_{YZ}) \\ -\sin(\theta_{YZ}) & 0 & \cos(\theta_{YZ}) \end{pmatrix} \begin{pmatrix} z-\bar{z} \\ x-\bar{x} \\ y-\bar{y} \end{pmatrix}.$$

Comme $\tan(\theta) = \frac{\sin(\theta)}{\cos(\theta)}$ et $\cos(\theta) = \sqrt{\frac{1}{1+\tan(\theta)^2}}$, on emploiera $\sin(\theta) = \tan(\theta)\cdot\cos(\theta)$ pour conserver le signe de la tangente lors de la construction de la matrice de rotation.

Dans le repère associé au plan tangent centré au centre de gravité du morceau, l'équation de la surface associée à l'estimée locale de l'image prend l'expression réduite :

$$\hat{Z} = a_{X^2}\cdot X^2 + a_{XY}\cdot XY + a_{Y^2}\cdot Y^2 + a_{X^3}\cdot X^3 + a_{X^2Y}\cdot X^2 Y + a_{XY^2}\cdot XY^2 + a_{Y^3}\cdot Y^3$$

Car elle correspond encore au développement limité de l'équation de la surface au point M qui s'écrit en appliquant la formule de Taylor jusqu'à l'ordre 3 :

$$\hat{z} = \bar{z} + \frac{\delta z}{\delta x}(\bar{x},\bar{y})\cdot x + \frac{\delta z}{\delta y}(\bar{x},\bar{y})\cdot y + \frac{\delta^2 z}{\delta^2 x}(\bar{x},\bar{y})\cdot x^2 + \frac{\delta^2 z}{\delta x \delta y}(\bar{x},\bar{y})\cdot xy + \frac{\delta^2 z}{\delta^2 y}(\bar{x},\bar{y})\cdot y^2$$
$$+ \frac{\delta^3 z}{\delta^3 x}(\bar{x},\bar{y})\cdot x^3 + \frac{\delta^3 z}{\delta^2 x \delta y}(\bar{x},\bar{y})\cdot x^2 y + \frac{\delta^3 z}{\delta x \delta^2 y}(\bar{x},\bar{y})\cdot xy^2 + \frac{\delta^3 z}{\delta^3 y}(\bar{x},\bar{y})\cdot y^3 + o^3(x,y)$$

Exprimer l'estimée locale de l'image dans le repère associé au plan tangent centré au centre de gravité du morceau revient alors à annuler le terme constant et le gradient de la fonction dans son expression polynomiale.

3.3.4. Repère propre d'une forme et réduction de son expression analytique

Poursuivons la réduction de cette équation en calculant son expression dans le repère propre $(M,\vec{u},\vec{v})$ de la quadrique $a_{X^2}\cdot X^2 + a_{XY}\cdot XY + a_{Y^2}\cdot Y^2$ dans $(M,\vec{X},\vec{Y})$.

L'expression de cette quadrique peut encore s'écrire :

$$a_{X^2}\cdot X^2+a_{XY}\cdot XY+a_{Y^2}\cdot Y^2=\begin{pmatrix}X & Y\end{pmatrix}\begin{pmatrix}\frac{\delta^2 Z}{\delta^2 X}(0,0) & \frac{\delta^2 Z}{\delta X\,\delta Y}(0,0)\\ \frac{\delta^2 Z}{\delta X\,\delta Y}(0,0) & \frac{\delta^2 Z}{\delta^2 Y}(0,0)\end{pmatrix}\begin{pmatrix}X\\Y\end{pmatrix}=\begin{pmatrix}X\\Y\end{pmatrix}^T H_Z\begin{pmatrix}X\\Y\end{pmatrix}$$, où

H_Z est la matrice hessienne de Z .

Comme il s'agit d'une matrice symétrique réelle, on peut encore écrire $H_Z=P\begin{pmatrix}\lambda_u & 0\\0 & \lambda_v\end{pmatrix}P^{-1}$, où

λ_u et λ_v sont les valeurs propres du hessien de Z et P leur matrice des vecteurs propres.

Ces valeurs propres ont pour expressions :

$$\lambda_u=\frac{1}{2}\left(a_{X^2}+a_{Y^2}+\sqrt{(a_{X^2}-a_{Y^2})^2+4a_{XY}^2}\right) \text{ et } \lambda_v=\frac{1}{2}\left(a_{X^2}+a_{Y^2}-\sqrt{(a_{X^2}-a_{Y^2})^2+4a_{XY}^2}\right)$$

La matrice P est la matrice de rotation d'angle : $\theta_{Xu}=\arctan\left(\frac{\lambda_u-a_{X^2}}{a_{XY}}\right)$ à π près dans le plan

tangent, qui s'écrit alors : $P=\begin{pmatrix}\cos(\theta_{Xu}) & \sin(\theta_{Xu})\\-\sin(\theta_{Xu}) & \cos(\theta_{Xu})\end{pmatrix}$.

Après application de cette nouvelle rotation dans le plan tangent passant par M , on obtient une nouvelle équation pour représenter l'estimée :

$\hat{Z}_\theta=\lambda_u\cdot u^2+\lambda_v\cdot v^2+a_{u^3}\cdot u^3+a_{u^2v}\cdot u^2v+a_{uv^2}\cdot uv^2+a_{v^3}\cdot v^3$, que peut être normalisée en effectuant

la mise à l'échelle : $\hat{w}=u^2+\frac{\lambda_v}{\lambda_u}\cdot v^2+\frac{a_{u^3}}{\lambda_u}\cdot u^3+\frac{a_{u^2v}}{\lambda_u}\cdot u^2v+\frac{a_{uv^2}}{\lambda_u}\cdot uv^2+\frac{a_{v^3}}{\lambda_u}\cdot v^3$.

Si $\lambda_v>0$ alors l'équation réduite de la quadrique est celle d'un paraboloïde elliptique, sinon il s'agit celle d'un paraboloïde hyperbolique. Les coefficients a_{u^3} , a_{u^2v} , a_{uv^2} et a_{v^3} mesurent les déformations asymétriques de ces surfaces autour de ces quadriques.

3.3.5. Caractéristiques invariantes des formes associées aux morceaux surfaciques

Les caractéristiques de rendu des formes issues de la décomposition régulière par morceaux s'appuient sur le centre de gravité du support des formes, c'est-à-dire sur les vecteurs plan $\begin{pmatrix} X_G \\ Y_G \end{pmatrix}$, que l'on réécrira $\begin{pmatrix} \overline{z} \\ \overline{x} \\ \overline{y} \end{pmatrix}$ en y intégrant les valeurs moyennes des bandes spectrales sur le support V de chaque morceau, c'est-à-dire où $\overline{z} = \begin{pmatrix} \overline{z}_1 \\ \overline{z}_2 \\ \vdots \\ \overline{z}_{N_S} \end{pmatrix}$.

On se positionnera ensuit dans la collection des N_S plans tangents au point $\begin{pmatrix} \overline{z} \\ \overline{x} \\ \overline{y} \end{pmatrix}$ en réalisant les rotations d'angle suivantes :

- $\theta^l_{XZ} = \arctan(a^l_x)$ dans le plan $(\begin{pmatrix} \overline{z}^l \\ \overline{x} \\ \overline{y} \end{pmatrix}, \vec{X}, \vec{Z})$, pour $l \in \{1, \cdots, N_S\}$;
- $\theta^l_{YZ} = \arctan(a^l_y)$ dans le plan $(\begin{pmatrix} \overline{z}^l \\ \overline{x} \\ \overline{y} \end{pmatrix}, \vec{Y}, \vec{Z})$, pour $l \in \{1, \cdots, N_S\}$.

Puis on effectuera dans chaque plan tangent la rotation d'angle $\theta_{Xu} = \arctan(\frac{\lambda_u - a_{X^2}}{a_{XY}})$ pour parvenir dans le repère propre de la quadrique $a_{X^2} \cdot X^2 + a_{XY} \cdot XY + a_{Y^2} \cdot Y^2$.

Comme pour le support de la forme, on disposera après normalisation des caractéristiques invariantes aux similitudes pour le rendu de la forme dans chaque bande spectrale :

- $\frac{\lambda_v}{\lambda_u}$ dans l'espace $\left(\begin{pmatrix}\bar{z}\\ \bar{x}\\ \bar{y}\end{pmatrix}, \vec{X}, \vec{Y}, \vec{Z}\right)$;
- $\frac{a_{u^3}}{\lambda_u}, \frac{a_{u^2v}}{\lambda_u}, \frac{a_{uv^2}}{\lambda_u}, \frac{a_{v^3}}{\lambda_u}$ dans l'espace $\left(\begin{pmatrix}\bar{z}\\ \bar{x}\\ \bar{y}\end{pmatrix}, \vec{X}, \vec{Y}, \vec{Z}\right)$.

4. Calcul de regroupements perceptuels

4.1. Introduction

Au début du XX^{ième} siècle, des psychologues ont développé une théorie holistique de la perception nommée Gestalttheorie. Il s'est agi de rechercher quelles sont les lois auxquelles la perception visuelle humaine pouvait être assujettie. Des expériences psycho-sensorielles ont pu mettre en exergue six grandes lois :

- la loi de la bonne forme : un ensemble de parties informe tend à être perçu d'abord comme une forme simple, symétrique et stable;
- la loi de continuité : des points rapprochés sont perçus par prolongement les uns des autres comme une seule forme ;
- la loi de proximité : les points les plus proches des uns des autres sont regroupés en priorité ;
- la loi de similitude : si ils ne sont pas suffisamment proches, les plus similaires d'entre eux sont regroupés en une seule forme;
- la loi du destin commun : des parties en mouvement ayant la même trajectoire sont perçues comme faisant partie de la même forme ;
- la loi de familiarité : les formes les plus familières sont perçues comme étant les plus significatives

Cette approche de la perception repose sur la structuration spatiale des formes visuelles et permet d'envisager la mise au point de techniques de reconnaissance des formes en parties cachées en identifiant tout ou partie d'un objet dans le champ de vision.

Le calcul de descripteurs numériques sur des formes visuelles (loi de la bonne forme) issues d'une segmentation (loi de continuité) ne permet d'envisager de reconnaître que des objets entièrement visibles dans le champ de vision. Ce qui est rarement possible étant donné que la vision humaine n'offre qu'une vue plane sur un univers physique tridimensionnel au travers d'une transformation projective.

Pour résoudre ce problème, il a été déjà proposé dans le passé d'adapter le calcul d'attributs manière à intégrer certaines lois de la perception des formes par agrégation des attributs des formes les plus proches (loi de proximité) le long d'un parcours arborescent des formes présentes dans une image.

Cette approche comporte à la fois les avantages de la reconnaissance statistique des formes (loi de similitude) comme ceux de la reconnaissance structurelle des formes (loi de familiarité) et a pu déjà être

testée dans le cadre d'une application passée de reconnaissance faciale fondée sur l'analyse d'images planes en couleur.

Le suivi des regroupements image par image dans une vidéo permettra d'illustrer la loi du destin commun.

Le calcul d'attributs suppose que l'objet d'intérêt ne comporte pas de partie cachée et puisse apparaître sous la forme d'une seule composante connexe. Ces conditions ne sont en général pas réunies : seule une partie de l'objet tridimensionnel peut être observée par un observateur unique et la segmentation d'images produira une décomposition de l'objet observé en de multiples composantes connexes. Pour résoudre un tel problème, il est proposé d'introduire une étape de calcul intermédiaire fondé sur le schéma de construction holistique suivant: les régions voisines sont agrégées de manière successive sous la forme de régions composées jusqu'à pouvoir recomposer aussi complètement que possible l'objet d'intérêt. Le calcul de descripteurs est ainsi généralisé aux régions composées et un objet donné apparaît alors comme une suite de composantes imbriquées les unes dans les autres, c'est-à-dire une série de formes complexes. Chacune d'entre elle dispose de son propre vecteur d'attributs et peut être identifiée individuellement : ainsi il devient possible de reconnaître un objet sans le voir entièrement en identifiant une partie des parties qui le compose.

4.2. Quantification vectorielle des descripteurs de forme

Pour décrire numériquement des formes simples issues de la segmentation régulière par morceaux d'image, on dispose pour chaque forme:

- d'un vecteur d'attributs associé au support de cette forme, basé sur le calcul des moments généralisés, et constitué de valeurs invariantes aux transformations géométriques qu'un objet peut subir dans le plan de l'image ;
- et pour chaque composante spectrale de l'image, d'un vecteur complémentaire décrivant le rendu de la forme aussi de manière invariante mais dans l'espace fonctionnel de l'image.

Pour disposer d'une description jusqu'à une précision donnée de ces informations numériques, il est proposé d'employer une structure arborescente de ces vecteurs de données numériques ([13],[28]), pour représenter des images mais aussi des attributs d'objets visuels permettant de mettre en œuvre des procédures de reconnaissance de formes par apprentissage statistique.

Pour rappel, il s'agit d'un schéma d'indexation basé sur des structures d'arbre binaire permettant de représenter des espaces de dimension quelconque régulièrement décomposés et utilisé pour enregistrer les données numériques d'ensemble d'apprentissage étiquetées ou non avec les identifiants des objets à reconnaître ([11],[12]). Pour représenter un ensemble de données numériques appartenant à un espace à k dimensions, les données peuvent être enregistrées dans un arbre binaire employant les vecteurs d'attributs comme des clés de coordonnées géographiques pour localiser

l'information dans l'arbre. Pour retrouver la position d'un vecteur de données, il faut produire un parcours similaire à celui effectué pour une recherche dans un arbre quaternaire ou octernaire:

- selon le paradigme diviser pour conquérir, l'espace multidimensionnel est divisé moitié par moitié successivement le long de chaque direction de l'espace jusqu'à atteindre la précision de modélisation donné en paramètre de recherche ;
- à chaque étape de division de l'espace, les coordonnées du vecteur sont examinées et la décision d'aller vers tel ou tel demi-espace gauche ou droit est prise jusqu'à atteindre à la précision d'analyse demandée dans l'arbre.

Prenons l'exemple de l'addition d'un vecteur normalisé, comme c'est la cas des descripteurs de forme, à un arbre binaire modélisant le cube unitaire de dimension k à la précision r :

DÉBUT

Initialisation de l'addition d'un vecteur normalisé à un arbre binaire représentant un 2^k-arbre :

vecteur ← vecteur à ajouter dans les données déjà présentes dans un arbre donné

racine ← racine de l'arbre dans lequel va s'effectuer l'addition

minrac(1 : k) ← $\{0.,0.,\cdots,0.\}$

maxrac(1 : k) ← $\{1.,1.,\cdots,1.\}$

niveau ← niveau atteint dans l'arbre

dimens ← k

profondeur ← $k \cdot r$

APPEL ← Addition d'un vecteur normalisé à un arbre binaire (vecteur, racine, minrac, maxrac, 0, dimension, profondeur)

FIN

PROCÉDURE Addition d'un vecteur normalisé à un arbre binaire (vecteur, racine, minrac, maxrac, niveau, dimension, profondeur)

DÉBUT

SI ((niveau <>profondeur) ET (NON NOIR(racine)) ALORS FAIRE

Descente dans l'arbre guidée par les coordonnées du vecteur :

nudim ← (niveau % dimens) + 1

centre ← (minrac(nudim) + maxrac(nudim) / 2.

SI (vecteur(nudim) < centre) ALORS xmax(nudim) ← centre , côté ← GAUCHE

SINON xmin(nudim) ← centre , côté ← DROITE

SI (TERMINAL(racine) ALORS FISSION(racine)

APPEL Addition d'un vecteur normalisé à un arbre binaire (vecteur, FILS(racine,côté), minrac, maxrac, niveau + 1, dimension, profondeur)

FIN

SINON NOIRCIR(racine)

SI (NON TERMINAL(racine)) ALORS FUSION(racine)

FIN

Cette procédure récursive permet d'inscrire la présence d'un vecteur de dimension quelconque dans l'espace unitaire de même dimension en gérant cet espace sous la forme d'un arbre binaire, ce qui revient à paginer $\left\{ 0, \frac{1}{2^r}, \ldots, \frac{2^r-1}{2^r} \right\}^k$ dans $\left\{ 0, \frac{1}{2^{kr}}, \ldots, \frac{2^{kr}-1}{2^{kr}} \right\}$.

Ce dernier est créé à vide grâce à l'instruction racine ← BLANC, qui crée un nœud blanc représentant l'espace vide à une dimension et une précision quelconque. Les branches de l'arbre se développent progressivement à chaque addition grâce à l'opérateur FISSION et deux nœuds filiaux isocolores représentant deux vecteurs voisins à la précision r sont fusionnés en un seul nœud au niveau paternel grâce à l'opérateur FUSION. Les nœuds NOIR et BLANC sont des nœuds terminaux de l'arbre qui indiquent si une donnée est présente ou non dans la branche qui a été parcourue.

Ainsi paginé dans un arbre binaire, un espace multidimensionnel peut être vu comme une suite de quadrants, d'octants ou d'hexadécants emboîtés successivement les uns dans les autres en fonction de la

dimension de l'espace. Leur union redonne l'espace entier initial dans lequel ils sont paginés et une simple recherche binaire permet de retrouver n'importe quel point ou sous-ensemble appartenant à cet espace modélisé de cette façon. La précision est un paramètre employé lors de la construction ou la recherche d'informations dans cet espace et correspond à la profondeur du parcours de l'arbre de représentation.

Alors un ensemble de données d'apprentissage peut être représenté par une collection d'arbres étiquetés construits à une certaine précision où les étiquettes sont les identificateurs des objets à reconnaître, indexés par leurs vecteurs de description calculés lors de l'analyse d'image, à raison d'un arbre spécifique aux attributs du support et autant d'autres arbres qu'il y a de composantes spectrales. Plus la précision est faible, plus la quantification sera grossière et plus elle est élevée, plus elle sera fine et détaillée.

4.3. Similarité en modélisation hiérarchique multidimensionnelle

La modélisation d'ensembles de données multidimensionnelles par le moyen de structures hiérarchiques arborescentes implique que les ensembles sont gérés grâce à une distance ultramétrique. Pour des espaces régulièrement divisés, la distance de structuration est la distance de Hausdorff : la distance entre deux points de l'espace est la taille du plus petit sous-ensemble de l'espace qui contient à la fois ces deux points. Ce n'est pas une distance conventionnelle qui permet de mesurer la distance séparant deux points de l'espace mais une pseudo-distance employée pour comparer des sous-ensembles entre eux.

Comme les caractéristiques décrivant les objets sont normalisées, il aurait pu être possible d'utiliser une distance euclidienne non pondérée pour définir une mesure de similarité. Mais dans le cas présent, les coordonnées des caractéristiques sont directement gérées au niveau de la structure d'arbre qui représente l'ensemble d'apprentissage et cela permet de comparer directement tout nouveau vecteur avec l'un quelconque des vecteurs qui ont été enregistrés lors de l'apprentissage. Ainsi les conditions sont plus favorables pour évaluer la similarité entre objets connus et inconnus. Cette faculté pourra utilement être employée notamment pour détecter et identifier les objets en mouvement dans une séquence d'images et procéder éventuellement à un enrichissement de la base d'apprentissage avec les caractéristiques des nouveaux objets. De plus, cette mesure de similarité induite par la structure arborescente permet de trier tous les objets appartenant à une base de données selon leur similarité avec un vecteur échantillon (interrogation par l'exemple) ou selon leur auto-similarité (tri d'une base de données).

4.4. Agrégation de formes

Le calcul d'attributs fondé sur les moments suppose que les objets à reconnaître soient entièrement visibles. C'est rarement le cas avec les applications de reconnaissance des formes. L'apprentissage statistique construit à partir de l'enregistrement de plusieurs vues d'un même objet, observé selon différentes positions et attitudes, permet de créer un ensemble suffisamment dense de données

caractéristiques pour espérer le reconnaître toute nouvelle vue dans une situation proche de celles déjà enregistrées.

Mais les moments calculés aux premiers ordres ne permettent que d'identifier des formes compactes. Par contre, la segmentation régulière par morceaux d'image devrait produire une décomposition plus fine des objets numérisés en générant un plus grand nombre de régions. Pour profiter de cette propriété, il est proposé de reprendre la méthode d'agrégation des formes employée par le passé en reconnaissance faciale ([30]). Usuellement, le regroupement de régions se propose d'agréger des régions adjacentes. Pour éviter de construire un graphe d'adjacence régionale, une approche différente d'agrégation reposant sur la proximité des régions a été mise au point. Durant l'agrégation de régions, le calcul d'attributs est effectué sans centrage, normalisation et mise à l'échelle. Comme les moments sont des formules de calcul intégral, les attributs des régions à agréger peuvent être directement sommés avant d'être post-traités. En utilisant les coordonnées des centres de gravité, un arbre quaternaire des centres de gravité des régions était alors construit et employé pour réaliser une agrégation des régions les plus proches grâce au parcours ascendant de l'arbre. Comme le processus d'agrégation est répété récursivement jusqu'au sommet de l'arbre, les divers objets d'intérêt présents dans l'image sont décomposés successivement sous la forme de suites de régions emboîtées.

Après le cumul des attributs de l'ensemble des régions, les caractéristiques sont centrées, normalisées et mises à l'échelle à chaque niveau de décomposition du modèle de l'image. Ainsi un objet dans l'image peut être décrit par une suite de longueur variable de vecteurs de caractéristiques. Durant l'apprentissage les étiquettes sont attribuées à ces suites de vecteurs permettant de mettre en œuvre un étage de reconnaissance s'appliquant à tout ou partie d'un objet partiellement visible.

4.5. Exemple de la reconnaissance faciale

Le calcul d'attributs fondé sur les moments suppose que les objets à reconnaître soient entièrement visibles. C'est rarement le cas des applications de reconnaissance faciale. L'apprentissage statistique construit à partir de l'enregistrement de plusieurs vues d'une même personne, observée selon différentes positions et attitudes, permet de créer un ensemble dense de données caractéristiques dans l'objectif de reconnaître une personne vue dans une situation proche de celles déjà enregistrées.

L'exemple présenté ici est fondé sur la segmentation d'image couleur. Elle permet de réaliser une décomposition fine des objets numérisés et produit un grand nombre de régions. Pour profiter de cette propriété, il a été mis en œuvre une méthode d'agrégation des formes qui prend en compte la plupart des groupements de régions qui peuvent apparaître dans la décomposition d'un visage. Un regroupement de régions permet d'agréger des régions adjacentes. Durant l'agrégation de régions, le calcul d'attributs est effectué sans centrage, normalisation et mise à l'échelle. Celle-ci repose sur une segmentation isocolore classique. En utilisant les coordonnées des centres de gravité, un arbre quaternaire des centres de gravité des régions est alors construit et est employé pour réaliser une agrégation des régions les plus proches grâce au parcours ascendant de l'arbre. Comme le processus d'agrégation est répété récursivement jusqu'au sommet de l'arbre, le visage d'une personne peut être recomposé progressivement en un ensemble de régions emboîtées.

Après le cumul des attributs de l'ensemble des régions, les caractéristiques sont centrées, normalisées et mises à l'échelle à chaque niveau de décomposition du modèle d'un visage. Ainsi un visage peut être décrit par une suite de longueur variable de vecteurs de caractéristiques. Durant l'apprentissage les étiquettes sont attribuées à ces suites de vecteurs permettant de mettre en œuvre un étage de reconnaissance faciale s'appliquant à tout ou partie d'un objet partiellement visible. Cette approche comporte à la fois les avantages de la reconnaissance statistique des formes comme ceux de la reconnaissance structurelle des formes.

Un exemple d'agrégation progressive des formes d'un visage est montré ci-dessous. La liste des attributs des différents objets produits lors de l'agrégation correspondant à ces images est présentée à la suite. Ils comportent des attributs de localisation et des attributs de forme. Les coordonnées du centre de gravité, l'angle et le facteur d'échelle sont des attributs de localisation de l'objet dans le référentiel de prise de vue. La surface, l'excentricité et les asymétries composent les attributs de forme de l'objet. Ce sont des valeurs centrées, normalisées et mises à l'échelle, c'est-à-dire indépendantes de toute translation, rotation et homothétie dans le repère d'acquisition.

Dans cette application précise, la décision d'authentification ou d'identification d'une personne parmi plusieurs est prise en fonction de l'étiquette recueillant le nombre maximum de formes simples ou agrégées reconnues.

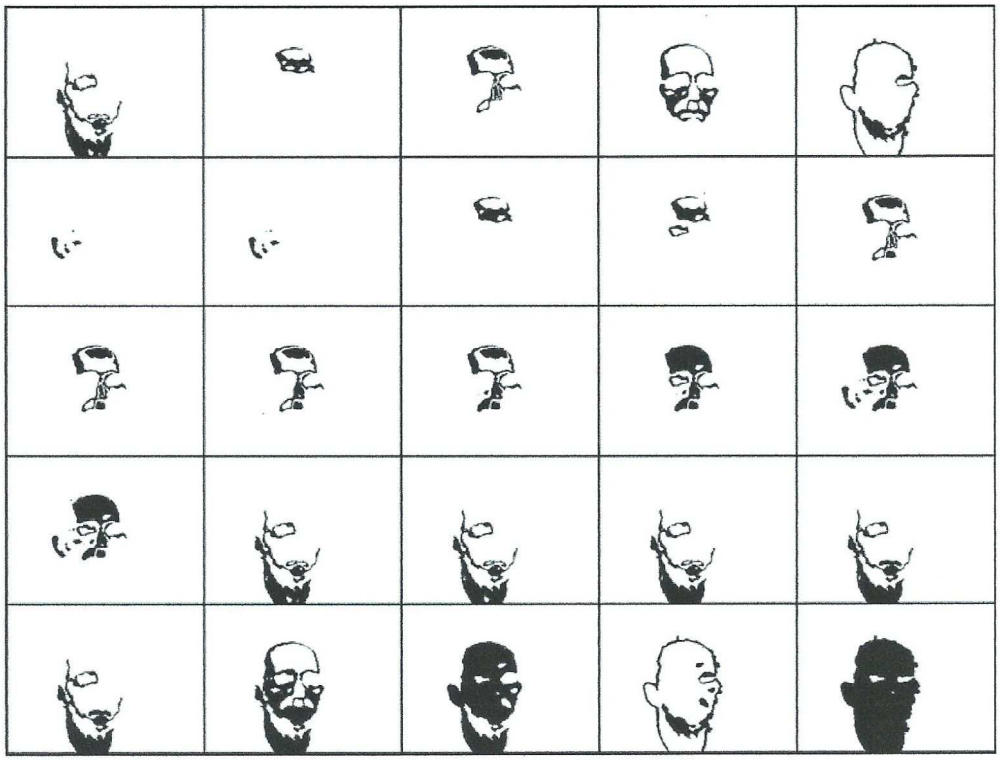


Figure 4 : Agrégation hiérarchique de régions

Object n°	Center Abs.	Center Ord	Angle	Scale	Surface	Eccentric.	1st Asym.	2nd Asym.
0	169.279	190.920	273.545	59.559	0.4747	0.7057	0.8579	0.3849
1	168.971	160.698	238.319	39.068	0.7220	0.7628	0.7980	0.5896
2	176.130	121.729	77.629	34.416	0.6487	0.7082	0.7952	0.1673
3	178.040	105.748	15.728	20.979	0.7661	0.5030	0.1476	0.3890
4	153.887	226.375	249.587	47.281	0.6608	0.5542	0.8516	0.4414
5	109.085	173.296	350.625	16.242	0.6306	0.7493	1.0041	0.2672
6	109.190	173.068	348.652	16.265	0.6320	0.7594	1.0039	0.3471
7	177.521	106.380	14.977	20.360	0.8164	0.5211	0.3752	0.4272
8	173.352	112.506	339.856	20.794	0.8888	0.8004	0.4257	0.5166
9	176.674	126.672	79.848	37.619	0.6162	0.6263	0.7472	0.2434
10	176.769	127.016	79.740	37.723	0.6167	0.6225	0.7359	0.2379
11	177.078	128.370	79.654	37.591	0.6319	0.6125	0.6774	0.2259
12	176.152	132.086	84.307	38.740	0.6347	0.5916	0.5893	0.2757
13	175.141	125.018	85.145	33.952	0.9060	0.6508	0.8530	0.3482
14	168.515	129.846	115.188	37.001	0.8765	0.7405	0.8251	0.2190
15	168.470	129.799	115.066	36.988	0.8773	0.7421	0.8255	0.2340
16	153.939	226.471	249.548	47.319	0.6609	0.5533	0.8515	0.4404
17	153.982	226.550	249.521	47.341	0.6610	0.5527	0.8518	0.4396
18	154.134	227.217	249.691	47.404	0.6645	0.5487	0.8589	0.4397
19	154.838	226.784	249.857	46.836	0.6810	0.5600	0.8577	0.4307
20	154.742	226.803	249.846	46.771	0.6829	0.5613	0.8577	0.4324
21	160.977	197.839	264.764	53.025	0.8036	0.5928	0.5511	0.4162
22	163.727	172.865	92.444	57.546	0.9308	0.5386	0.6141	0.2124
23	172.523	188.593	275.887	58.201	0.5090	0.7112	0.8155	0.5023
24	165.788	176.550	92.379	58.052	1.0544	0.5853	0.3648	0.2288

Tableau 1 : Liste des attributs des régions agrégées

4.6. Algorithmes d'agrégation de formes

4.6.1. Agrégation des descripteurs des rendus de formes

Dans le cas qui nous préoccupe, il est inutile de construire l'arbre de centres de gravité des différentes régions qui constitue l'image, car il est identique à celui produit par la décomposition régulière par morceaux de l'image.

Il suffit juste d'adapter l'algorithme de décomposition régulière en morceaux pour agréger les descripteurs de formes en supprimant le test sur l'ordre à ne pas dépasser pour celles-ci et se borner à ne conserver que les coefficients d'ordre 3 générés par la combinaison barycentrique . Voici l'algorithme de décomposition régulière par morceaux revu dans cette intention en introduisant la gestion explicite de la structure arborescente des données:

DÉBUT

ordre ← 1

arbre ← BLANC

APPEL Décomposition régulière par morceaux (V , arbre, ordre, précision)

FIN

PROCÉDURE Décomposition régulière par morceaux (V , arbre, ordre, précision)

DÉBUT

Calcul de l'estimée linéaire de V au sens des moindres carrés :

$\hat{z} = \begin{pmatrix} f_1 \\ \vdots \\ f_l \\ \vdots \\ f_{N_s} \end{pmatrix} \begin{pmatrix} 1 \\ x \\ y \end{pmatrix}$ où les f_l sont des vecteurs lignes résultant de l'estimation linéaire

Calcul de l'erreur maximale d'approximation :

$$\varepsilon_V^\infty = \lfloor Max_{p \in V} \{ |z_{1,p} - f_1(x_p, y_p)|, \cdots, |z_{N_s,p} - f_{N_s}(x_p, y_p)| \} \rfloor$$

SI (ε_V^∞ > précision) ALORS FAIRE

Calcul de l'histogramme des erreurs

SI (histogramme mono-modal)

ALORS $V_S \leftarrow$ { $\sqrt{Card(V)}$ points de plus forte erreur }

SINON FAIRE

Définir comme seuil la vallée la plus élevée de l'histogramme

$$V_S \leftarrow \{ p \in V / \epsilon_p^\infty > seuil \}$$

FIN

Calculer les coordonnées $\bar{x}$ et $\bar{y}$ du centre

et les variance σ_{x^2} et covariance σ_{xy} de V_S

Diviser l'ensemble V en deux sous-ensembles :

$$V_- = \{ p \in V / (x - \bar{x}) \cdot \sigma_{xy} - (y - \bar{y}) \cdot \sigma_{x^2} < 0 \}$$

$$V_+ = \{ p \in V / (x - \bar{x}) \cdot \sigma_{xy} - (y - \bar{y}) \cdot \sigma_{x^2} \geq 0 \}$$

SI (TERMINAL(racine) ALORS FISSION(racine)

$\widehat{z_-}$ ← Décomposition régulière par morceaux (V_- , FILS(racine,GAUCHE),ordre, précision)

$\widehat{z_+}$ ← Décomposition régulière par morceaux (V_+ , FILS(racine, DROIT), ordre, précision)

$\widehat{z_-}$ ← VALEUR(FILS(racine, GAUCHE))

$\widehat{z_+}$ ← VALEUR(FILS(racine, DROIT))

Calcul de la combinaison barycentrique :

$$\hat{z} = \frac{(x - \overline{x_-}) \cdot (\overline{x_+} - \overline{x_-}) + (y - \overline{y_-}) \cdot (\overline{y_+} - \overline{y_-})}{(\overline{x_+} - \overline{x_-})^2 + (\overline{y_+} - \overline{y_-})^2} \cdot \widehat{z_-} + \frac{(\overline{x_+} - x) \cdot (\overline{x_+} - \overline{x_-}) + (\overline{y_+} - y) \cdot (\overline{y_+} - \overline{y_-})}{(\overline{x_+} - \overline{x_-})^2 + (\overline{y_+} - \overline{y_-})^2} \cdot \widehat{z_+}$$

Calcul de l'erreur d'approximation :

$$\varepsilon_V^{\infty} = \lfloor Max_{p \in V} \{ |z_{1,p} - f_1(x_p, y_p)|, \cdots, |z_{N_s,p} - f_{N_s}(x_p, y_p)| \} \rfloor$$

VALEUR(racine) ← $\hat{z}$ tronquée à l'ordre 3

ordre ← ordre + 1

FIN

RETOUR

FIN

4.6.2. Agrégation des descripteurs des supports de formes

Par contre lors du calcul des descripteurs des supports de forme, il ne faut plus arrêter cette étape au niveau des nœuds terminaux de la décomposition, mais poursuivre celui-ci jusqu'à la racine de l'arbre de décomposition. Ce sera seulement un fois celle-ci atteinte que les descripteurs de support des formes pourront être centrés et normalisés jusqu'au niveau de la racine de l'arbre. Il en est de même pour les descripteurs de rendus où l'on continuera à employer le moyennage barycentrique pour calculer les coefficients des formes agrégées, mais en tronquant les développements polynomiaux à l'ordre 3.

5. Construction de dictionnaires visuels

5.1. Alphabet visuel des formes simples

Les formes simples issues d'une segmentation régulière par morceaux d'une image comporte:

- un vecteur d'attributs décrivant le support de cette forme, basé sur le calcul des moments généralisés ;
- et pour chaque composante spectrale de l'image, d'un vecteur complémentaire décrivant le rendu de la forme.

Les attributs décrivant le support de chaque forme sont les valeurs :

- $\epsilon = \frac{M(u_2^2)}{M(u_1^2)}$ l'excentricité du support de la forme;
- $\frac{M(u_1^3)}{M(u_1^2)}, \frac{M(u_1^2 u_2)}{M(u_1^2)}, \frac{M(u_1 u_2^2)}{M(u_1^2)}, \frac{M(u_2^3)}{M(u_1^2)}$ les asymétries normalisées .

Ceux décrivant chaque composante spectrale de l'image, le rendu d'une forme est décrit par la suite des valeurs : $\frac{\lambda_v}{\lambda_u}, \frac{a_{u^3}}{\lambda_u}, \frac{a_{u^2 v}}{\lambda_u}, \frac{a_{u v^2}}{\lambda_u}, \frac{a_{v^3}}{\lambda_u}$.

D'un point de vue pratique, il est possible de se restreindre à la restriction convexe du support, c'est-à-dire aux coefficients $\frac{M(u_2^2)}{M(u_1^2)}, \frac{M(u_1^3)}{M(u_1^2)}, \frac{M(u_2^3)}{M(u_1^2)}$ et à un seul coefficient par composante spectrale $\frac{\lambda_v}{\lambda_u}$ suffisant à décrire une quadrique.

Ces valeurs sont indépendantes des transformations linéaires qu'elles peuvent subir dans l'espace visuel où elles ont été capturées. Après quantification, elles se réduisent à des sous-ensembles finis de valeurs normalisées qui peuvent être assimilés à des alphabets visuels permettant d'encoder des images sous la forme de suite de caractères idéographiques comme ce fut le cas avec les premiers systèmes d'écriture idéographiques développés par l'humanité. En fonction de la précision employée pour quantifier les formes simples, l'alphabet ainsi produit peut être plus ou moins riche.

Que se soient les descripteurs des supports de forme comme ceux de leurs rendus, la quantification est mise en œuvre en construisant leurs arbres dans leurs repères propres à la précision souhaitée pour les reconnaître, puis les reconstruire lors de leur synthèse.

Comme l'espace est unitaire, l'algorithme présenté dans le chapitre précédent s'applique non seulement pour encoder les valeurs de ces descripteurs, mais aussi les points décrivant la forme de chaque support de manière à être réutilisée lors de la synthèse d'une image. Ainsi chaque support de forme est discrétisé dans le plan unitaire et se trouve ainsi représenté par un arbre quaternaire, dont il a pu être montré dans le passé qu'il est moins encombrant qu'un codage à longueur de trains ([28]). Comme l'algorithme est additif, plusieurs occurrences d'un même support peuvent être cumulés en phase d'apprentissage de manière à obtenir une description visuelle riche et stable comme cela a été évoqué pour des objets en mouvement dans une séquence vidéo.

En ce qui concerne les descripteurs de rendu des formes, il n'est pas nécessaire de les discrétiser au-delà de la quantification de leurs coefficients, car il suffira de discrétiser l'expression de ses coefficients sur le support de la forme pour régénérer la forme telle qu'elle a pu être acquise avant tout traitement.

La description des formes simples telle qu'elle vient d'être décrite peut être aussi comparée au système phonétique employé par les alphabets actuels composé de consonnes et de voyelles dont la conjonction permet de mettre en œuvre une écriture syllabique : chaque forme simple est la conjonction du support d'une forme géométrique avec une expression analytique précisant le rendu visuel à produire pour régénérer la forme simple initiale capturée dans le contexte d'une image. Le support s'assimilerait alors à une consonne et le rendu à une ou plusieurs voyelles. C'est pour cette raison que cet ensemble visuel primaire a pu être appelé visème par analogie au terme de phonème que l'écriture syllabique tente de représenter : le phonème est décrit comme la plus petite unité discrète que l'on peut distinguer par segmentation de la chaîne parlée ([3]).

5.2. Dictionnaire visuel des formes complexes

Le regroupement de formes simples en formes composées est à la base de la technique de reconnaissance de formes en parties cachées proposée ici-même et permettant d'identifier des objets réels tridimensionnels observés en géométrie projective par un capteur plan. En fonction de la position et de l'attitude de l'observateur face à l'objet observé, il sera produit différentes séquences d'assemblage de formes simples dont les arborescences décriront l'organisation et permettront d'identifier des séquences de formes communes comme des branches ou sous-branches identiques. Chacune de ces arborescences de formes simples décrivant une forme composée donnée constitue un mot bâti sur l'alphabet des visèmes précédemment décrit. Ainsi un même objet tridimensionnel observé selon différents points de vue produira autant de mots sur cet alphabet qu'il pourra y avoir autant d'occurrences pour le décrire : au point de vue lexicographique, ils sont synonymes les uns des autres.

Ainsi dans l'exemple de la reconnaissance faciale, le tableau d'attributs constitue de cette manière l'un des mots permettant d'indexer le visage qui a été segmenté en vue d'être étiqueté avec le nom de la personne auquel il appartient. Pour être plus précis, il décrit uniquement les consonnes dont il est

composé, les voyelles étant masquées dans la segmentation qui a été mise en œuvre dans cette application (sans prise en compte du rendu).

L'ensemble des formes composées présentes dans une image permet alors de construire le dictionnaire des formes de l'image : il possède à la base une structure similaire au résultat de la segmentation régulière par morceaux de cette même image, mais il diffère par le fait qu'il est indexé sur un alphabet de formes simples au lieu de morceaux surfaciques d'ordre donné. Ce dictionnaire peut être enrichi par les formes présentes non plus dans une seule image, mais aussi dans une séquence d'images en identifiant les regroupements qui suivent un mouvement commun.

Il peut encore prendre en compte des images ou des séquences d'images multiples pour consolider sa construction et permettre de reconnaître les séquences similaires et identifier les synonymes visuels.

Il s'agit donc de construction de bases de reconnaissance par apprentissage :

- apprentissage supervisé en nommant les objets observés dans la première image d'une séquence correspondant à un plan fixe vidéo, puis en approfondissant celui-ci en suivant les objets fixes ou en mouvement d'image en image;
- apprentissage non supervisé en laissant la procédure d'enrichissement identifier par elle-même les objets similaires et en les nommant à l'issue du traitement.

Ainsi en étiquetant les objets similaires de la base d'apprentissage, le dictionnaire prend une dimension ontologique en nommant les objets observés. La richesse du vocabulaire enregistré dans un dictionnaire visuel varie avec la profondeur de l'arbre d'indexation des formes composées qu'il contient.

5.3. Synthèse des formes simples et complexes

A l'aide du dictionnaire des formes complexes présents dans une image, il faut décomposer chaque forme complexe en série de formes simples que l'on positionnera en lieu et place, tel que cela a été enregistré :

- en positionnant le support des formes là où elle ont été localisées, c'est-à-dire en (X_G, Y_G, θ) dans le plan image après mise à l'échelle d'un rapport $M(u_1^2)$;
- en numérisant le rendu de chaque composante dans le repère polaire de son expression analytique $(\theta_{Xu}, \theta_{XZ}, \theta_{YZ})$ centré en $\begin{pmatrix} \overline{z} \\ \overline{x} \\ \overline{y} \end{pmatrix}$ sachant que $\begin{pmatrix} \overline{x} \\ \overline{y} \end{pmatrix} = \begin{pmatrix} X_G \\ Y_G \end{pmatrix}$, et en effectuant une homothétie d'un rapport λ_u .

Pour combler les trous et lever les duplications ponctuelles de valeurs, il sera préférable d'appliquer un filtrage médian sur l'image des étiquettes des supports de formes avant de discrétiser les expression analytiques de chaque composant spectrale.

6. Codage d'images et de vidéos

6.1. Parcours optimal des blocs d'une image régulièrement divisée

Une image plane peut être régulièrement divisée en blocs de même taille à l'aide de représentation arborescentes comme par exemple les quadtrees ou arbres quaternaires. Le courbe de Hilbert ou de Peano-Hilbert permet de parcourir chacun de ces blocs en se déplaçant au plus proche voisin en ne passant qu'une fois et une seule par un même bloc. Disposant d'une représentation de l'image par un quadtree, il est aisé de mettre en œuvre un parcours récursif qui produise cette courbe pour visiter chacun des points d'une image donnée ([8]).

L'algorithme proposé est le suivant :

- il s'agit d'un algorithme récursif visitant les nœuds d'un quadtree en profondeur d'abord ;
- à chaque nœud de l'arbre, on associe une orientation qui spécifiera dans quel ordre visiter ses fils ;
- au moment de leur visite, l'orientation de chaque fils est recalculée à partir de celle du père de manière que la visite des petits-fils soit homothétique de celle des fils à une rotation près ;
- quatre orientations différentes peuvent être spécifiées correspondant chacune à un motif de base de la courbe de Hilbert qui se déduisent les uns des autres par une rotation de 90°.

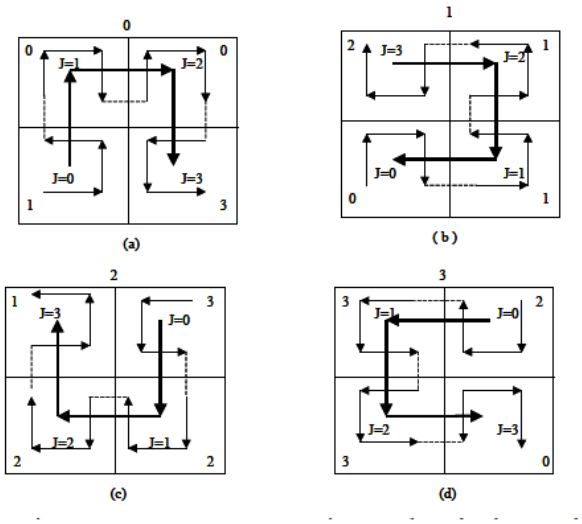


Figure 5: Quatre motifs de base d'une courbe de Hilbert dans le plan

On trouvera présenté ci-dessus chacun de ces quatre motifs ([27]) et par la suite le développement d'un motif à plusieurs résolutions successives.

6.2. Parcours de visite des objets d'une image

Nous proposons d'utiliser un tel parcours pour visiter les différences morceaux et leurs regroupements dans une image pour réaliser son encodage. Comme il y a identité de structure entre la décomposition régulière par morceaux de l'image et l'arbre des centres de gravité de ces mêmes morceaux, il sera recherché de visiter l'ensemble de ces morceaux en se déplaçant au plus proche voisin selon la position de leurs centres de gravité. Cela équivaudra encore à minimiser les mouvements oculaires d'un observateur qui souhaiterait examiner consécutivement chaque forme présente dans l'image en les observant une fois et une seule.

Voici par exemple l'un des parcours possibles pour se déplacer entre plusieurs grandes villes d'un pays donné, en minimisant les déplacements d'une ville à l'autre ([29]). Leurs coordonnées sont spécifiées de manière cartographique (en latitude et longitude sur la planisphère terrestre).

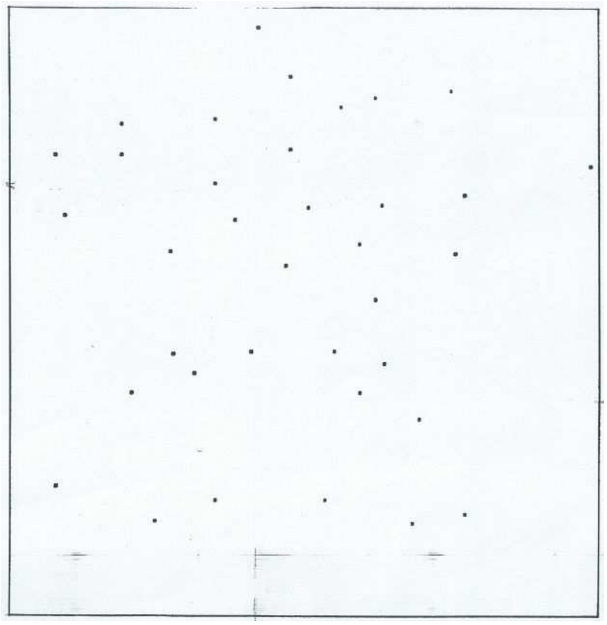


Figure 6 : Carte des principales villes de France en coordonnées polaires

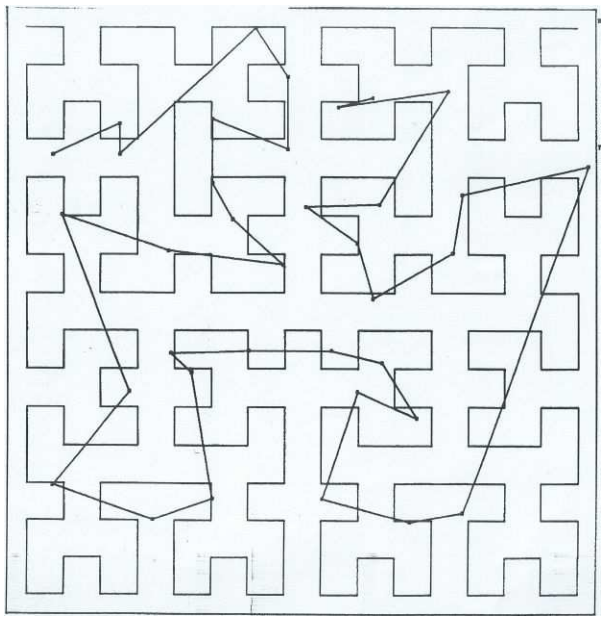


Figure 7 : Recherche d'un parcours de visite par balayage de Hilbert

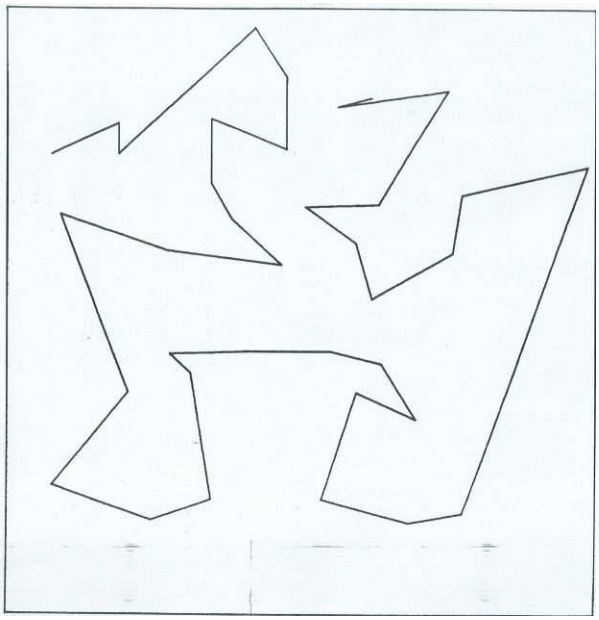


Figure 8 : Parcours de visite au plus proche voisin des principales villes de France

6.3. Syntaxe et phraséologie visuelles

Il s'agit de ce parcours qui est envisagé pour encoder les formes simples comme les formes complexes présentes dans une image. Il permet de lister les différentes formes qui constituent une image en se déplaçant de forme en forme au plus proche voisin. Du point de vue syntaxique, il permet de représenter une image comme la phrase d'une suite de mots enregistrés dans le dictionnaire des formes complexes qui eux-mêmes s'écrivent sur l'alphabet des formes simples dont les consonnes sont les supports de ces formes et les expressions analytiques du rendu en sont les voyelles.

Comme le parcours va constituer un balayage de l'image en commençant à peu près toujours au même endroit et en suivant les mêmes règles de déplacement, il est envisageable que ces phrases puissent être catégorisées de manière à pouvoir décrire :

- la composition de la scène en se référant à des scènes conventionnelles (paysage de mer, de campagne, de montagne, vue de groupe, portrait, etc.);
- la nature de la prise de vue (plan large, plan rapproché, vue aérienne, vue en contre-plongée, etc.).

6.4.Tri multidimensionnel d'une base de formes

La modélisation d'espaces de dimension k régulièrement décomposés par des arbres d'ordre 2^k paginés dans des arbres binaires montre que ce modèle de représentation a la puissance du continu, c'est-à-dire qu'il est possible de trouver une transformation continue qui pagine des sous-ensembles de R^k directement dans R .

Cela a pour conséquence que l'on doit pouvoir trouver des parcours permettant de visiter de manière continue une fois et une seule l'ensemble des données d'une base : dans le cas des 2^k -arbres en se déplaçant au plus proche voisin au sens de la distance de Hausdorff.

La courbe de Hilbert n'est pas restreinte aux seuls balayages du plan, mais s'étend aussi à des espaces multidimensionnels en généralisant le système de visite des quadrants d'un arbre quaternaire aux 2^k -ants d'un 2^k -arbre.

Pour une base de données multidimensionnelles, cela revient à trouver un parcours de visite maximisant de proche en proche l'indice de similarité des données enregistrées, c'est-à-dire à trier les données de la base en fonction de leur similarité.

Dans le projet de recherche évoqué en introduction, il était envisagé de proposer un module d'édition d'image où des bibliothèques de formes puissent être construites par apprentissage et indexées à l'aide de leurs descripteurs de manière à pouvoir y accéder de manière triée en fonction de leur similarité dans l'espace multidimensionnel de description, le parcours s'effectuant au plus proche voisin à précision variable.

Conclusion

Le programme scientifique qui vient d'être décrit permet de se rapprocher du concept d'image généralisée tel que D.H. Ballard et C.M. Brown l'ont présenté dans leur ouvrage « Computer Vision » ([2]):

- l'image numérique correspond au niveau de description d'un tableau de luminescences;
- l'image segmentée permet de définir des objets homogènes (segmentation du support en composantes partageant des propriétés communes dans l'espace fonctionnel);
- la structure géométrique : décomposition des supports segmentés sur une base de primitives géométriques, niveau de description d'une image équivalent à de l'information graphique;
- la structure relationnelle : regroupement de sous-ensembles de la partition originelle de l'image par découverte des relations spatiales ou de similarité avec des modèles d'organisation connus.

Il fait écho à l'article que publie au même moment J. Mariani dans un numéro spécial de la revue « La Recherche » portant sur « L'Intelligence Artificielle » ([3]) et dans lequel il décrit la reconnaissance de la parole comme un processus décomposé en quatre niveaux plus un spécifique à l'application visée:

- le niveau acoustique : niveau où le signal phonique est digitalisé et dont on extrait les traits acoustiques ;
- le niveau phonétique : dont on extrait les phonèmes ;
- le niveau lexical : où l'on compose les mots ;
- le niveau syntaxique et sémantique : où l'on compose les phrases ;
- et enfin le niveau pragmatique : niveau lié à l'application, valable dans un univers fini et fixé à l'avance, sans lequel les étapes précédentes ne peuvent être validées .

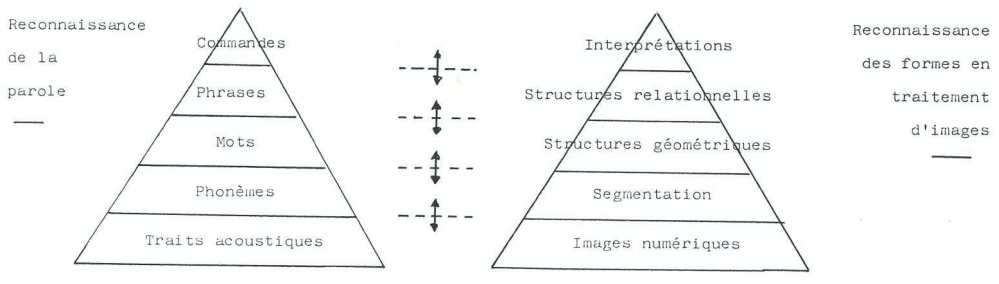


Figure 9 : Emboîtement pyramidal des informations dans un système perceptuel

Comme l'illustre les développements proposés dans le programme de recherche qui vient d'être décrit, on ne peut s'empêcher de faire le lien entre ces deux approches en perception artificielle pour établir le parallèle présenté ci-dessus sous la forme d'un diagramme à l'organisation pyramidale.

En reprenant les résultats attendus décrits dans la proposition de recherche soumise à l'initiative Open-FET du 6ième PCRD, le projet SDVC tente de déployer les fondations d'une solution proposant une représentation innovante pour décrire l'information visuelle qui doit permettre de définir des ontologies visuelles et de gérer du contenu visuel de manière similaire à du contenu textuel. Grâce à ce schéma de modélisation, SDVC devrait fournir un outil efficace pour la recherche rapide de contenu mixte (textuel et visuel) et devrait ouvrir une voie semblable pour prendre en compte les contenus audios dans l'avenir.

Dans cette optique, le projet SDVC vise à mettre en place des activités de recherche innovantes et devrait permettre de faire progresser l'état de l'art sur le sujet:

- en proposant une description de l'information de niveau intermédiaire permettant de construire des systèmes de recherche de l'information image et vidéo plus efficace ;
- en prenant en compte les lois de la perception pour réduire le fossé entre les objets numériques et les vues en perspective des objets du monde réel ;
- en gérant les objets visuels indépendamment de leur localisation dans le plan de vue et d'un ensemble de transformations géométriques données ;
- en proposant un schéma d'indexation à partir duquel des ontologies visuelles peuvent être inférées et en permettant de définir une sorte d'alphabet visuel pour les formes simplement connexes et un dictionnaire visuel intégrant les regroupements perceptuels ;
- et par conséquent en fournissant un moyen pour réduire le fossé sémantique existant au cœur de l'information visuelle ([26]).

Le projet présente une forte perspective exploratoire et ouvre la voie pour de futures investigations à propos de la représentation des contenus et des technologies du codage auto-descriptif de l'information.

Le projet a été sélectionné lors de la première étape de l'appel à projets Open-FET (projets ouverts en technologies futures et émergentes) sous le nom de GROOVIES, mais n'a malheureusement pas été financé. Il a été soumis à un moment où les recherches sur les représentations multi-échelles et les ondelettes géométriques étaient intenses ([17]-[23]). Ces mêmes représentations sont actuellement mobilisées pour tenter d'expliquer l'efficacité de l'apprentissage par réseaux convolutifs multi-étages ([24]). Les travaux qui viennent présentés pourraient peut-être lever un coin du voile sur certaines questions, notamment pour savoir combien d'étages sont nécessaires au bon fonctionnement d'un tel réseau.

Références bibliographiques

1. D. Lecomte, D. Cohen, Ph. De Bellefonds, J. Barda, Les normes et les standards du multimédia – *Série Internet et Intranet, Dunod 1999*
2. D. H. Ballard, C. M. Brown – Computer Vision - *Prentice Hall ,1985*
3. J. Mariani - La Reconnaissance de la Parole *in "L'Intelligence Artificielle", Numéro spécial La Recherche, Octobre 1985*
4. Th.. Pavlidis - Algorithms For Graphic and Image Processing. *Computer Science Press, 1982*
5. F. P. Preparata, M .I. Shamos - Computational Geometry: An introduction. *Texts and Monographs in Computer Science - Springer Verlag, 1985*
6. A. Avez – Calcul différentiel – *Maîtrise de Mathématiques Pures, Masson 1983*
7. J.-J. Risler - Méthodes Mathématiques pour la CAO – *Collection Recherches en Mathématiques Appliquées - Masson 1991*
8. J.-C. Simon – La reconnaissance des formes par algorithmes- *Études et Recherche en Informatique, Masson 1984*
9. G. Gaillat – Méthodes statistiques de reconnaissance des formes – *Département informatique – automatique – E.N.S.T.A. 1983*
10. L. Miclet – Méthodes structurelles pour la reconnaissance des formes – *Collection technique et scientique des télécommunications, Eyrolles 1984*
11. J. L. Bentley - Multidimensional Binary Search Trees Used for Associative Searching - *CACM, Vol. 18, N° 9, September 1975*
12. J. L. Bentley - Multidimensional Divide-and-Conquer - *CACM, Vol. 23, N° 4, April 1984*
13. P. Fiche, V. Ricordel, C. Labit - Etude d'algorithmes de quantification vectorielle arborescente pour la compression d'images fixes - *Programme 4 Robotique, image et vision - Projet Temis - Publication IRISA n° 807 - Janvier 1994*
14. B. S. Manjunath, P. Salembier, T. Sikora (Eds.), Introduction to the MPEG-7 Multimedia Content Description Language, *Wiley & Son, 2002*
15. J. M.Buhmann,J. Malik, P. Perona, Image recognition: Visual grouping, recognition, and learning, 5th annual German-American Frontiers of Science symposium, Postdam, June 10-13, 1999
16. J. Luo, C.-E. Guo, Perceptual Grouping of Segmented Regions in Color Images. *Pattern Recognition 36(2003) 2781-2792*
17. Ch. S. Sastry, A. K. Pujari, B. L. Deekshatulu, C. Bhagvati, A Wavelet Based Multiresolution Algorithm for Rotation Invariant Feature Extraction, *Pattern Recognition Letters, Vol. 25, Issue 16 (Dec. 2004)*
18. T. Lindeberg, Scale-Space Theory in Computer Vision, *Kluwer Academic Publishers, Dordrecht, Netherlands, 1994*
19. F. Mokhtarian, S. Abbasi, J. Kittler, Efficient and Robust Retrieval by Shape Content through Curvature Scale Space, *Image and Databases and Multi-Media Search, pp 51-58, World Scientific Publishing, 1997*
20. D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision, vol. 60, no. 2, 2004*
21. J. Romberg, M. Wakin, and R. G. Baraniuk, "Multiscale Geometric Image Processing," *presented at SPIE Visual Communications and Image Processing, Lugano, Switzerland, 2003*
22. S. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.*

11, pp. 674-693, 1989

23. E. Le Pennec and S. Mallat, "Sparse Geometric Image Representations with Bandelets," *IEEE Transactions on Image Processing, vol. 14, pp. 423-438, 2005*
24. S. Mallat, Understanding Deep Convolutional Networks*, Philosophical Transactions A, The Royal SocietyPublishing March 2017*
25. B. J. Wielingu, A. Th. Schreiber, J. Wielemaker, J. A. C. Sandberg, From Thesaurus to Ontology, *International Conference on Knowledge Capture, Victoria, Canada, October 2001*
26. P. Enser, C. Sandom, Towards a Comprehensive Survey of the Semantic Gap in Visual Image Retrieval*, Conference on Image and Video Retrieval 2003*
27. F-Z.N. Nacer, A. Zergaïnoh, A. Merigot, Décomposition en quadtree de la DCT globale pour la compression des images, *Coresa 2001*

28. *https://hal.archives-ouvertes.fr/hal-01185357* O. Guye, Modélisation Hiérarchique de Données Multidimensionnelles dans des Espaces Régulièrement Décomposés: Principes de Base;
29. *https://hal.archives-ouvertes.fr/hal-01185361* O. Guye, Modélisation Hiérarchique de Données Multidimensionnelles dans des Espaces Régulièrement Décomposés: Implémentation sur Calculateur;

30. *https://hal.archives-ouvertes.fr/hal-01185368* O. Guye, Modélisation Hiérarchique de Données Multidimensionnelles dans des Espaces Régulièrement Décomposés: Applications en Analyse d'Images.