Chapitre 1

Attributs des images

1.1 Introduction

Dans tout ce qui suit nous représenterons l'image par une fonction soit continue $A(x, y) = A(s)$ soit discrète $A[m, n] = A[s]$ de deux variables d'espace. Les valeurs prises par A varient typiquement dans $[0, 255]$ et sont appelés niveaux de gris de l'image. Nous allons dans ce chapitre définir les attributs communément utilisés en analyse d'images ainsi que leurs principales propriétés.

La nature des attributs dépend de la modélisation sous-jacente utilisée pour l'image. Ainsi, certains attributs que nous allons décrire tels les moments d'espace, histogramme, coocurrence correspondent à des statistiques calculées sur l'image qui est alors vu comme la réalisation d'un processus stochastique. Par opposition l'analyse spectrale des images suppose que l'on cherche à détecter des structures périodiques dans les images et l'image n'est plus considérée comme un processus stochastique.

1.2 Attributs stochastiques d'une image

Dans ce paragraphe, on considère l'image comme un processus aléatoire discret $A[s]$ à valeur dans $\mathbb{R}$. Si s est fixé, $A[s]$ est alors une variable aléatoire pour laquelle on peut définir différentes grandeurs qui permettent de caractériser le processus $A[s]$ à partir de différentes réalisations.

1.2.1 Moments Statistiques

La première grande catégorie de caractéristiques est celle des moments statistiques.

- *moyenne*

  Pour un signal discret, la moyenne est définie par :

$$\mu[s] = E\{A[s]\} = \sum_i p_i a_i$$

  où a_i est une réalisation de la variable $A[s]$ et $p_i = P(A[s] = a_i)$.

- *moment d'ordre k* $E\{A^k[s]\} = \sum_i p_i a_i^k$
- *moment centré d'ordre k* $\mu_k(s) = E\{(A[s] - \mu[s])^k\}$
- *autocorrélation* $R[s_1, s_2] = E\{A[s_1, s_2]\}$

1.2.2 Moments Spatiaux

Les moments spatiaux sont aussi très souvent employés nous verrons ensuite dans quel cas ils correspondent aux moments statistiques.

- *moyenne spatiale*
$$m_1 = \frac{1}{K} \sum_{s \in R} A[s]$$
- *variance spatiale*
$$m_2 = \frac{1}{K-1} \sum_{s \in R} (A[s] - m_1)^2$$
Remarque : attention normalisation en $K-1$ pour obtenir un estimateur sans biais.
- *moment d'espace centré d'ordre k*
$$m_k = \frac{1}{K-1} \sum_{s \in R} (A[s] - m_1)^k$$
- *autocorrélation*
Pour une translation t du plan, le moment d'espace du second ordre est donné par la formule :
$$M_2(t) = \frac{1}{N_c} \sum_{s \in R, s+t \in R} A[s]A[s+t]$$
où N_c est le nombre de couples $(s, s+t)$ de la région R.

1.2.3 Processus stationnaires

On cherche à établir le lien entre moments statistiques et moments spatiaux, on utilise pour cela la notion de stationnarité. Le processus est stationnaire du premier ordre si $E\{A[s]\}$ est indépendant du site. Il est stationnaire au deuxième ordre si l'autocorrélation ne dépend que de la translation t entre deux sites :

$$R[s_1, s_2] = E\{A[s_1], A[s_2]\} = R[t] = E\{A[s]A[s+t]\}$$

Si de plus le processus est ergodique, on peut écrire :

$$\lim_{R \to \mathbb{N}^2} \tfrac{1}{K} \sum_{s \in R} A[s] = E\{A[s]\}$$
$$\lim_{R \to \mathbb{N}^2} \tfrac{1}{N_c} \sum_{s \in R, s+t \in R} A[s]A[s+t] = E\{A[s]A[s+t]\}.$$

L'étude des seuls moments ne suffit pas pour discriminer les textures (images particulières). Des images différenciables à l'oeil et ayant des moments identiques jusqu'à l'ordre 4 ou 5 ont été synthétisées.

1.2.4 Histogrammes

Pour $\omega \in \Omega = \{0, \cdots, L-1\}$, où L est le nombre de niveau de gris de l'image, on définit :

$$h(\omega) = \frac{1}{K} \sum_{s \in R} \delta(A[s] - \omega) \text{ où } \delta(A[s] - \omega) = \begin{cases} 1 \text{ si } A[s] = \omega \\ 0 \text{ sinon} \end{cases}$$

Pour un très grand nombre de réalisations, on estime $P(A(s) = \omega)$ par $h(\omega)$. Les histogrammes fournissent une caractéristique peu intéressante sur la nature des images car deux images visuellement similaires peuvent avoir des histogrammes de niveau de gris très différents.

1.2.5 Matrice de Coocurrence

Pour une translation t la matrice de coocurrence MC_t d'une région R est définie par :

$$MC_t(a, b) = card\{(s, s+t) \in R^2 | A[s] = a, A[s+t] = b\}$$

La matrice MC_t est de taille $L \times L$. En pratique, on ne considère pas tous les niveaux de gris. MC_t est difficilement manipulable et on extrait de ces matrices de coocurrence différentes caractéristiques dont voici la liste, introduite par Haralick [4] pour la description des textures :

- *homogénéité :*

$$\frac{1}{N_c^2} \sum_a \sum_b (MC_t(a, b))^2$$

  Cet indice est d'autant plus élevé que l'on retrouve souvent le même couple de pixels, ce qui est le cas quand le niveau de gris est uniforme ou quand il y a périodicité dans le sens de la translation.

- *contraste :*

$$\frac{1}{N_c(L-1)^2} \sum_{k=0}^{L-1} k^2 \sum_{|a-b|=k} MC_t(a, b)$$

  Chaque terme de la matrice MC_t est pondéré par sa distance à la diagonale. On obtient un indice correspondant à la notion usuel du contraste, il est élevé quand les termes éloignés de la diagonale sont élevés, c'est à dire quand on passe d'un pixel trés clair à un pixel trés foncé ou inversement.
- *entropie :*

$$1 - \frac{1}{N_c \log(N_c)} \sum_a \sum_b MC_t(a, b) \log(MC_t(a, b)) \chi_{MC_t}(a, b)$$

  avec $\chi_{MC_t}(a, b) = 1$ si $MC_t(a, b) \neq 0$ et 0 sinon. L'entropie est faible si on a souvent le même couple de pixels, forte si chaque couple est peu représenté. Elle fournit un indicateur du désordre que peut présenter une texture.

- *corrélation*
$$\frac{1}{N_c\sigma_x\sigma_y}|\sum_a\sum_b(a-m_x)(b-m_y)MC_t(a,b)|$$
où
$$m_x=\frac{1}{N_c}\sum_a\sum_b aMC_t(a,b)$$
est la moyenne pondérée des lignes de MC_t,
$$m_y=\frac{1}{N_c}\sum_a\sum_b bMC_t(a,b)$$
est la moyenne pondérée des colonnes,
$$\sigma_x^2=\frac{1}{N_c}\sum_a\sum_b(a-m_x)^2MC_t(a,b)$$
est la variance des lignes et
$$\sigma_y^2=\frac{1}{N_c}\sum_a\sum_b(b-m_y)^2MC_t(a,b)$$
est la variance des colonnes.
Cet attribut rend compte de la corrélation entre les lignes et les colonnes de la matrice de cooccurrence, c'est à dire entre le niveau de gris du pixel d'origine et celui d'arrivée de la translation.
- *homogénéïté locale*
$$\frac{1}{N_c}\sum_a\sum_b\frac{1}{1+(a-b)^2}MC_t(a,b)$$
- *directivité*
$$\frac{1}{N_c}\sum_a MC_t(a,a)$$
La directivité est importante s'il y a des pixels de même niveau de gris séparés par la translation t.
- *uniformité*
$$\frac{1}{N_c^2}\sum_a MC_t^2(a,a)$$
Elle est d'autant plus importante qu'un seul niveau de gris apparait souvent dans la direction de la translation.

Les indices liés à la matrice de cooccurrence sont très souvent utilisés pour la discrimination des textures. Ils conservent un intérêt même lorsque l'image n'est pas la réalisation d'un processus aléatoire (ils permettent en effet de mesurer la périodicité dans les images)

1.2.6 Longueurs de plage ("run-length")

Une plage de niveau de gris est un ensemble de pixels consécutifs, dans une direction donnée, ayant le même niveau de gris. La longueur d'une plage est le nombre de pixels dans cet ensemble. A chaque direction, on peut associer une matrice de longueurs de plages $P_\theta = (p_\theta(i,j))$; l'élément $p_\theta(i,j)$ de cette matrice représente le nombre de plages de longueur j (dans la direction θ) constituées de pixels de niveau de gris i. Dans une région, on peut extraire différents attributs de cette matrice de L lignes et n_θ colonnes (L est le nombre de niveaux de gris dans l'image et n_θ est la longueur de la corde de direction θ dans la région) :

- *nombre de longueurs de plage*

$$SLP = \sum_{i=0}^{L-1} \sum_{j=1}^{n_\theta} p_\theta(i,j)$$

- *proportion de petites plages*

$$RF_1 = \frac{1}{SLP} \sum_{i=0}^{L-1} \sum_{j=1}^{n_\theta} \frac{p_\theta(i,j)}{j^2}$$

- *proportion de grandes plages :*

$$RF_2 = \frac{1}{SLP} \sum_{i=0}^{L-1} \sum_{j=1}^{n_\theta} j^2 p_\theta(i,j)$$

- *hétérogénéité des niveaux de gris :*

$$RF_3 = \frac{1}{SLP} \sum_{i=0}^{L-1} \left(\sum_{j=0}^{n_\theta} p_\theta(i,j) \right)$$

  RF_3 mesure la dispersion des plages entre les niveaux de gris.
- *hétérogénéité des longueurs de plage*

$$RF_4 = \frac{1}{SLP} \sum_{j=1}^{n_\theta} \left(\sum_{i=0}^{L-1} p_\theta(i,j) \right)^2$$

- *pourcentage de plages :*

$$RF_5 = \frac{SLP}{K}$$

1.3 Analyse Spectrale des Images

D'autres caractéristiques peuvent être extraite de l'étude de la transformée de Fourier de l'image. La transformée de Fourier permet de passer d'une représentation de l'image dans le domaine spatial (coordonnées m, n) à une représentation dans le domaine fréquentiel (coordonnées u, v). La transformée de Fourier d'une séquence 2D correspondant au signal discret $A[m, n]$ défini sur $\mathbb{Z}^2$ s'exprime sous la forme :

$$F(A)(u, v) = \sum_{m=-\infty}^{m=+\infty} \sum_{n=-\infty}^{n=+\infty} A[m, n] e^{-2i\pi(mu+nv)}$$

En pratique une image $A[s] = A[m, n]$ avec m et n entiers, $0 \leq m \leq M-1$ et $0 \leq n \leq N-1$ est un signal à support borné. Si l'on suppose que $A[m, n]$ correspond à l'échantillonnage d'une fonction continue B aux points $(\frac{mT}{M}, \frac{nT}{M})$. La transformée de Fourier discrète approxime la transformée de Fourier continue selon (cf cours Valérie Perrier 2ème année) :

$$\hat{B}(\frac{k}{T}, \frac{l}{T}) = F(A)[k, l] = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} A[m, n] e^{-2i\pi(\frac{km}{M} + \frac{ln}{N})}$$

La périodicité d'une image se retrouve dans la périodicité du spectre et cette dernière est beaucoup plus facile à extraire que sur l'image elle-même.

1.4 Retour sur le calcul de la fonction d'autocorrélation, théorème de Wiener-Kinchin

La fonction d'autocorrélation possède un rôle central en analyse d'images, pour deux raisons fondamentales : la première est qu'elle exprime de manière très complète les dépendances spatiales des pixels de l'image, la seconde est qu'elle est facilement calculable dans le domaine de Fourier grâce au théorème de Wiener-Kinchin.

Rappelons la formule de la fonction d'autocorrélation dans le cas d'une image finie de taille $M \times N$ supposé périodique dans les deux directions :

$$M_2[k, p] = \frac{1}{NM} \sum_{[m,n] \in [0,M-1] \times [0,N-1]} A[m+k, n+p] A[m, n]$$

Dans le cas de signaux continus, on définit le spectre de puissance comme le module de la transformée de Fourier au carré. Le spectre de puissance pour un signal discret de taille finie est alors défini comme le module de $F(M_2)[u, v]$ (on appelle aussi cette grandeur densité spectrale de puissance).

Ecrivons alors la transformée discrète de Fourier de la fonction d'autocorrélation :

$$F(M_2)[u, v] = \frac{1}{(MN)^2} \sum_{[m,n] \in [0,N-1] \times [0,M-1]} \sum_{[k,p] \in [0,N-1] \times [0,M-1]} A[m+k, n+p] A[k, p] e^{-2i\pi(\frac{mu}{M} + \frac{nv}{N})}.$$

On intervertit l'ordre de calcul des sommes pour obtenir :

$$F(M_2)[u, v] = \frac{1}{MN} \sum_{[k,p] \in [0,N-1] \times [0,M-1]} A[k, p]$$

$$
\begin{aligned}
& \frac{1}{MN} \sum_{[m,n]\in[0,N-1]\times[0,M-1]} A[m+k,n+p]e^{-2i\pi(\frac{(m+k)u}{M}+\frac{(n+p)v}{N})}e^{2i\pi(\frac{ku}{M}+\frac{pv}{N})} \\
= & \ |F(A)[u,v]|^2
\end{aligned}
$$

Autrement dit, la transformée de Fourier discrète de la fonction d'autocorrélation est le module de la tranformée de Fourier discrète du signal au carré.

1.5 Conclusion

Le choix des attributs pour caractériser une région dépend étroitement du type d'image à traiter. On distingue les images qui peuvent être assimilées à des processus aléatoires des images présentant des structures prériodiques ou caractérisées par des niveaux de gris homogènes.

Chapitre 2

Théorie de Shannon, Rééchantillonnage

Dans ce chapitre nous revenons sur la théorie de Shannon et nous discutons le problème du rééchantillonnage.

2.1 La théorie de Shannon

Soit A un signal monodimensionnel dont la transformée de Fourier est de carré intégrable et de support fini. Supposons $Supp(\hat{A}) \subset [-U, U]$ et que $a \leq \frac{1}{2U}$ alors le théorème de Shannon nous dit que l'on peut reconstruire le signal A de manière exacte à partir des échantillons $A(na)$ grâce à la formule suivante :

$$A(t) = \sum_{n \in \mathbb{Z}} A(na) \frac{\sin(\frac{\pi}{a}(t - na)}{\frac{\pi}{a}(t - na)}$$

Ce résultat s'étend évidemment au cas bidimensionnel, en supposant cette fois que A est de carré intégrable sur $\mathbb{R}^2$ et que la transformée de Fourier bidimensionnelle de A est à support compact inclus dans $[-U_1, U_1] \times [-U_2, U_2]$. En considérant alors $a_1 \leq \frac{1}{2U_1}$ et $a_2 \leq \frac{1}{2U_2}$ on peut écrire :

$$A(x, y) = \sum_{(n,p) \in \mathbb{Z}^2} A(na_1, pa_2) \frac{\sin(\frac{\pi}{a_1}(x - na_1))}{\frac{\pi}{a_1}(x - na_1)} \frac{\sin(\frac{\pi}{a_2}(y - pa_2))}{\frac{\pi}{a_2}(y - pa_2)}. \tag{2.1}$$

Ce pose alors la question fondamentale suivante : peut-on considèrer les objets que l'on oberve comme des signaux à spectre limité ? La réponse est non, car un objet quelle que soit l'échelle d'observation présente des structures liées à cette échelle. En revanche, la précision des instruments d'observation font que l'on analyse en fait des objets qui sont à bande limitée.

Le cas limite stipulé par le théorème de Shannon en une dimension nous dit que si $a = \frac{1}{2U}$ et si l'on suppose que le spectre est symétrique (ce qui est le cas pour les signaux réels pairs le support de $\hat{A}$ est exactement $[-U, U]$) alors il y a unicité du filtre de reconstruction. En revanche, si le support est $[-U_1, U]$ avec $U > U_1$ alors il n'y a plus unicité. Tous les filtres satisfaisant la condition suivante peuvent être utilisés pour la reconstruction :

$$W_k(x) = \begin{cases} 1 \text{ sur } [-U_1, U] \\ 0 \text{ en dehors de } [-U, U] \end{cases}$$

Ces filtres W_k sont des transformées de Fourier de filtres w_k et le signal reconstruit s'écrit : $f(x) = f * w_k(x)$ où $*$ désigne la convolution discrète. On voit donc que $(f(na))_{n\in\mathbb{Z}}$ est un sytème générateur mais plus une base. Il est donc nécessaire de trouver un échantillonnage qui soit efficace par rapport au problème posé.

2.2 Améliorer la résolution fréquentielle des images

Si on utilise la transformée de Fourier discrète pour calculer une approximation de la transformée de Fourier continue, les valeurs obtenues correspondent aux fréquences $\frac{k}{T}$. Une technique communément utilisée pour améliorer la résolution fréquentielle est de rajouter des zéros (zero-padding) à la fin de la séquence dont on calcule la transformée de Fourier. En général, on rajoute P zéro de façon à avoir $N+P$ puissance de deux, ce qui permet d'utiliser la transformée de Fourier rapide. En pratique, l'effet de l'ajout de P zéros sur la transformée de Fourier discrète est le suivant :

$$F(A)[k] = \frac{1}{N+P}\sum_{n=0}^{N+P-1} A(n)\exp(-2i\pi k\frac{n}{N+P}) = \frac{1}{N+P}\sum_{n=0}^{N-1} A(n)\exp(-2i\pi k\frac{n}{N+P}).$$

La suite $F(A)[k]$ est maintenant périodique de période $N+P$, et les $F(A)[k]$ founissent une approximation de la transformée de Fourier aux fréquences $\frac{kN}{(N+P)T}$.

2.3 Phénomène d'aliasing

Lorsque l'on échantillonne le signal à la fréquence d'échantillonnage a, pour que la formule soit exacte il faut que le spectre de A soit à support inclus dans $[-\frac{1}{2a}, \frac{1}{2a}]$. En pratique, on est parfois amené à choisir a trop grand ce qui provoque des erreurs appelées erreurs d'aliasing. Nous pouvons calculer cette erreur en utilisant la formule de Poisson. On rappelle qu'un signal echantillonné est de la forme :

$$\Delta_a A = \sum_{n=-\infty}^{+\infty} A(na)\delta_{na}.$$

la transformée de Fourier de δ_{na} est $e^{-2i\pi na\omega}$. On rappelle que :

$$\begin{aligned}
\forall\phi \in \mathcal{S}(\mathbb{R}), < \hat{\delta}_{na}, \phi > &= < \delta_{na}, \hat{\phi} > \\
&= \hat{\phi}(na) \\
&= \int \phi(x) e^{-2i\pi nax} dx \\
&= < e^{-2i\pi nax}, \phi >
\end{aligned}$$

ce qui permet d'écrire

$$\widehat{\Delta_a A}(\omega) = \sum_{n\in\mathbb{Z}} A(na)\exp(-2i\pi na\omega).$$

En utilisant alors la formule de Poisson, on obtient :

$$\widehat{\Delta_a A}(\omega) = \frac{1}{a} \sum_{k \in \mathbb{Z}} \hat{A}(\omega - \frac{k}{a})$$

Supposons alors que le support de $\hat{A}$ ne soit pas compris dans $[-\frac{1}{2a}, \frac{1}{2a}]$ alors la transformée de $\widehat{\Delta_a A}$ ne correspond pas du tout sur l'intervalle $[-\frac{1}{2a}, \frac{1}{2a}]$ à celle de A. L'effet principal de l'aliasing est d'ajouter des hautes fréquences à un signal basse fréquence. Nous présentons maintenant une technique permettant de diminuer l'effet de l'aliasing. On recherche pour cela le signal $\tilde{A}$, dont la transformée de Fourier a comme support $[-\frac{1}{2}, \frac{1}{2}]$, qui approche le mieux A au sens de la norme L^2. On peut donc écrire que :

$$\begin{aligned} \|A - \tilde{A}\|_2^2 &= \int_{-\infty}^{+\infty} |\hat{A}(\omega) - \widehat{\tilde{A}}(\omega)|^2 d\omega \\ &= \int_{|\omega| > \frac{1}{2a}} |\hat{A}(\omega)|^2 d\omega + \int_{|\omega| \leq \frac{1}{2a}} |\hat{A}(\omega) - \widehat{\tilde{A}}(\omega)|^2 d\omega. \end{aligned}$$

La distance est minimale lorsque la seconde intégrale est nulle soit si $\widehat{\tilde{A}}(\omega) = \hat{A}(\omega)\chi_{[-\frac{1}{2a}, \frac{1}{2a}]}$, ce qui se réécrit $\tilde{A} = \frac{1}{a} A * h_{\frac{1}{a}}$. Remarquons que la minisation considère un signal $\tilde{A}$ continu, cependant comme on a $\widehat{\Delta_a \tilde{A}}(\omega) = \widehat{\tilde{A}}(\omega)$, le signal echantilloné a sa transformée de Fourier aussi égale à $\hat{A}$ sur $[-\frac{1}{2a}, \frac{1}{2a}]$.

2.4 Rééchantillonnage des images et des signaux

L'utilisation courante d'applications utilisant des images nécessitent de pouvoir modifier leur taille. Il est facile de voir qu'une diminution de la taille en changeant seulement a conduit inévitablement à des erreurs d'aliasing. L'utilisation de différents algorithmes plus ou moins évolués permet de minimiser le phénomène d'aliasing suivant le type d'image et le facteur de redimensionnement. On peut faire un redimensionnement efficace des images en utilisant la formule de Shannon. Une approximation de A sur des images (de taille finie) peut être obtenue en adaptant la formule (2.1) selon :

$$A(x, y) \approx \sum_{n=-k_1}^{k_1} \sum_{p=-k_2}^{k_2} A(na_1, pa_2) \frac{\sin(\frac{\pi}{a_1}(x - na_1))}{\frac{\pi}{a_1}(x - na_1)} \frac{\sin(\frac{\pi}{a_2}(y - pa_2))}{\frac{\pi}{a_2}(y - pa_2)},$$

on peut alors en déduire la valeur sur tout rééchantillonnage en calculant $A(\frac{x}{\alpha}, \frac{y}{\beta})$ avec la formule d'approximation précédente.

Chapitre 3

La restauration d'images

La restauration d'images est l'opération qui corrige des images dégradées et reconstruit un signal de bonne qualité à partir d'un signal de médiocre qualité.

3.1 Dégradation par application d'un filtre

On considère un signal originel A et un signal dégradé g reliés par la relation :

$$g = \mathcal{D}(A)$$

où $\mathcal{D}$ est un opérateur linéaire et spatialement invariant. $\mathcal{D}$ satisfait donc la définition classique de filtre et peut s'écrire sous la forme d'un produit de convolution :

$$g = d * A.$$

Vu sous cet angle la restauration apparait donc comme une opération de déconvolution. On rappelle que la convolution s'écrit :

$$g(x) = \int_{-\infty}^{+\infty} A(t)d(x-t)dt$$

qui donne par transformée de Fourier :

$$\hat{g}(u) = \hat{A}(u)\hat{d}(u).$$

On peut donc obtenir la transformée de Fourier de A par :

$$\hat{A}(u) = \frac{\hat{g}(u)}{\hat{d}(u)} = \hat{g}(u)W_I(u)$$

puis par transformée de Fourier inverse :

$$A = g * w_I.$$

Le problème est en théorie résolu. Cependant, un des problèmes majeurs de cette approche est que $\hat{d}(u)$ peut-être nul. Au point où $D(u)$ s'annule, on a aucune information fréquentielle. On peut remplacer $W_I(u)$ par le filtre de Bracewell défini par :

$$W_B(u) = \begin{cases} \frac{1}{\hat{d}(u)} \text{ si } d(u) \neq 0 \\ 0 \text{ sinon} \end{cases}$$

3.2 Approche par filtrage de Wiener

On suppose cette fois que le signal est perturbé de la manière suivante :

$$g(x, y) = A(x, y) * d(x, y) + b(x, y)$$

Wiener cherche une solution de la forme $\tilde{A} = g * w_W$ qui minimise ϵ, l'espérance de l'erreur quadratique, intégrale entre le signal et son estimée, espérance estimée sur toutes les réalisations du bruit b :

$$\epsilon =< \int_{\mathbb{R}^2} (A(x) - \tilde{A}(x))^2 dx > .$$

En appliquant le théorème de Parseval, on obtient :

$$\epsilon =< \int_{\mathbb{R}^2} |\hat{A}(u) - \widehat{\tilde{A}}(u)|^2 du > .$$

On remplace alors $\widehat{\tilde{A}}(u)$ par $\hat{g}(u)W_W(u) = (\hat{A}(u)\hat{d}(u) + \hat{b}(u))W_W(u)$, puis on calcule la différentielle par rapport à W_W. La différentielle de l'expression sous l'intégrale est calculée de la manière suivante :

$$\begin{aligned} & |\hat{A}(u) - (\hat{A}(u)\hat{d}(u) + \hat{b}(u))(W_W(u) + h(u))|^2 \\ & = |\hat{A}(u) - (\hat{A}(u)\hat{d}(u) + \hat{b}(u))W_W(u)|^2 \\ & +2 * Re\left(\overline{\hat{A}}(u) - (\overline{\hat{A}}(u)\overline{\hat{d}}(u) + \overline{\hat{b}}(u))\overline{W_W}(u)(\hat{A}(u)\hat{d}(u) + \hat{b}(u))h(u)\right) + O(h^2) \end{aligned}$$

Donc la différentielle de ϵ s'annule si et seulement si :

$$\begin{aligned} & \forall h \quad < \textstyle\int_{\mathbb{R}^2} Re\left(\overline{\hat{A}}(u) - (\overline{\hat{A}}(u)\overline{\hat{d}}(u) + \overline{\hat{b}}(u))\overline{W_W}(u))(\hat{A}(u)\hat{d}(u) + \hat{b}(u))h(u)\right) du > \\ & =< \textstyle\int_{\mathbb{R}^{\not{2}}} Re\Big(|\hat{A}(u)|^2\hat{d}(u) + \hat{b}(u)\overline{\hat{A}}(u) \\ & - \left[|\hat{A}(u)|^2|\hat{d}(u)|^2 + \hat{b}(u)\overline{\hat{A}}(u)\overline{\hat{d}}(u) + \hat{A}(u)\hat{d}(u)\overline{\hat{b}}(u) + |\hat{b}(u)|^2\right]\overline{W_W}(u)h(u)\Big) du >= 0 \end{aligned}$$

Comme l'espérance de $\hat{b}(u)$ est nulle l'expression précédente se simplifie en :

$$< \int_{-\infty}^{+\infty} Re\left(|\hat{A}(u)|^2\hat{d}(u) - \left[|\hat{A}(u)|^2|\hat{d}(u)|^2 + |\hat{b}(u)|^2\right]\overline{W_W}(u)\right) h >= 0,$$

Une condition suffisante d'annulation est donc, en supposant que la densité spectrale de puissance du bruit est constante égale à N_0^2 (Rappel : $|\hat{b}(u)|^2 = \int_{\mathbb{R}^2} R_{bb}(\tau)\exp(-i2\pi < u, \tau >)d\tau = N_0^2$) :

$$W_W(u) = \frac{\overline{\hat{d}}(u)}{|\hat{d}(u)|^2 + \frac{N_0^2}{|\hat{A}(u)|^2}}$$

On remarque alors que si $\hat{d}(u) = 0$, obtient la même valeur que celle fournie par le filtre de Bracewell. En l'absence de bruit, on a un comportement identique à celui du filtre inverse.

3.3 Evaluation du défaut

Nous avons supposé jusqu'ici que le filtre d était connu, ce qui n'est bien sûr pas toujours le cas en pratique ; on procède alors à une evaluation statistique du défaut.

Chapitre 4

Equations aux dérivées partielles et traitement d'images

4.1 Introduction

L'origine du lien entre équations aux dérivées partielles et traitement d'images provient du fait qu'une image ne peut être représentée par ses échantillons que si la fonction sous-jacente est suffisamment lisse (en pratique il faut même que la fonction soit C^∞ pour que sa transformée de Fourier soit à support compact). Cela oblige à faire des convolutions avec des fonctions appelées noyaux en guise de prétraitement.

Dans ce qui suit nous aurons besoin des notations suivantes : $x = (x_1, x_2)$, $|x| = (x_1^2 + x_2^2)^{\frac{1}{2}}$, $x.y$ dénote le produit scalaire sur $\mathbb{R}^2$, $u(x)$ est le niveau de gris au point x. En supposant que u est assez régulière nous notons : $u_x = \frac{\partial u}{\partial x}$, $u_y = \frac{\partial u}{\partial y}$ et $u_{xy} = \frac{\partial^2 u}{\partial x \partial y}$. Nous notons aussi le gradient de u sous la forme $Du = (u_x, u_y)$ et le laplacien est noté $\Delta u = u_{xx} + u_{yy}$.

4.2 Propriété de l'opérateur de moyennisation

On définit l'opérateur de moyennisation sur le disque de centre x et de rayon h par :

$$m_h u_0(x) = \frac{1}{\pi h^2} \int_{D(x,h)} u_0(y) dy$$

On montre que l'opérateur de moyennisation vérifie la propriété :

$$\frac{m_h u_0(x) - u_0(x)}{h^2} = \frac{1}{8} \Delta u_0(x) + \epsilon(h)$$

Dém : Sans perte de généralité plaçons nous en $x = 0$. Rappelons alors (Calcul de différentielles dans les espaces de Banach) :

$$u_0(y) = u_0(0) + Du_0(0).y + \frac{1}{2}((u_0)_{xx}(0)y_1^2 + (u_0)_{yy}(0)y_2^2 + 2(u_0)_{xy}(0)y_1 y_2) + o(h^2).$$

La valeur moyenne calculée alors sur le disque nous donne :

$$(m_h u_0)(0) = u_0(0) + \frac{1}{2\pi h^2}((u_0)_{xx}(0) \int_{D(0,h)} y_1^2 dy_1 dy_2 + (u_0)_{yy}(0) \int_{D(0,h)} y_2^2 dy_1 dy_2) + o(h^2)$$

On obtient alors le résultat souhaité en remarquant que :

$$\frac{1}{2\pi h^2} \int_{D(0,h)} x_1^2 dx_1 dx_2 = \frac{1}{4\pi h^2} \int_{D(0,h)} (x_1^2 + x_2^2) dx_1 dx_2 = \frac{1}{4\pi h^2} \int_0^h 2\pi r^3 dr = \frac{h^2}{8} \blacksquare$$

Nous allons maintenant généraliser cette propriété qui nous permettra de faire le lien entre noyau de convolution et équation de la chaleur.

4.2.1 Convolution par des noyaux à symétrie radiale normalisés

Nous définissons les noyaux à symétrie radiale normalisés de la manière suivante, $g(x) = g(|x|)$ et :

$$\int_{\mathbb{R}^2} g(x)dx = 1 \quad \int_{\mathbb{R}^2} x_1^2 g(x)dx = \int_{\mathbb{R}^2} x_2^2 g(x)dx = 2.$$

Nous nous intéressons alors à la préservation de ces propriétés par changement d'échelle. Regardons les propriétés satisfaites par $ag(\frac{x}{b})$. Pour que la première égalité soit satisfaite, il faut nécessairement que $ab^2 = 1$ c'est à dire $b = \frac{1}{\sqrt{a}}$. Avec cette relation liant a et b, les deux dernières relations sont satisfaites (changement de variables). Il est facile de voir alors que :

$$\int_{\mathbb{R}^2} x_1 g(x)dx = \int_{\mathbb{R}^2} x_2 g(x)dx = \int_{\mathbb{R}^2} x_1 x_2 g(x)dx = 0 \tag{4.1}$$

En effet, la fonction g étant à symmétrie radiale, on a :

$$\int (x_1 - x_2)^2 g(x)dx = \int (x_1 + x_2)^2 g(x)dx$$

Ce qui permet de déduire la dernière propriété. Pour l'une des deux premières propriétés, il suffit de remarquer que :

$$\begin{aligned} \textstyle\int_{\mathbb{R}^2} x_1 g(x)dx &= \textstyle\int_0^{+\infty} x_1 \int_{-\infty}^{\infty} g(x)dx_2dx_1 + \int_{-\infty}^0 x_1 \int_{-\infty}^{\infty} g(x)dx_2dx_1 \\ &= \textstyle\int_0^{+\infty} x_1 \int_{-\infty}^{\infty} g(x)dx_2dx_1 - \int_0^{+\infty} x_1 \int_{-\infty}^{\infty} g(x)dx_2dx_1 = 0. \end{aligned}$$

On considère alors le changement d'échelle $g_h(x) = \frac{1}{h} g(\frac{x}{h^{\frac{1}{2}}})$, et l'on dénote par $g^{n*} = g * g * \cdots * g$ la convolution d'ordre n. On définit de même g_h^{n*}, dont on étudie le comportement lorsque $n \to +\infty$ et $h \to 0$. On étudie tout d'abord les propriétés de la convolution par g_h. Nous avons le théorème suivant :

Théorème 1 *Soit $g(x) \in L^1(\mathbb{R}^2)$ une fonction radiale normalisée, supposons de plus que :*

$$\int_{\mathbb{R}^2} |g(z)||z|^3 dz = C < +\infty$$

alors pour toute fonction C^3 $u \in L_K^\infty$, on a :

$$(g_h * u)(x) - u(x) = h\Delta u(x) + O(h^{\frac{3}{2}})$$

Dém : on peut écrire :

$$\begin{aligned}(g_h * u)(x) - u(x) &= \int_{\mathbb{R}^2} h^{-1} g(\frac{y}{h^{\frac{1}{2}}})(u(x-y) - u(x))dy \\ &= \int_{\mathbb{R}^2} g(z)(u(x - h^{\frac{1}{2}}z) - u(x))dz \\ &= \int_{\mathbb{R}^2} g(z)(-h^{\frac{1}{2}} Du(x).z + \frac{h}{2} D^2u(x)(z,z))dz \\ &\quad -\frac{1}{6}h^{\frac{3}{2}} \int_{\mathbb{R}^2} g(z) D^3u(x - h^{\frac{1}{2}}\theta z)(z,z,z)dz\end{aligned}$$

où $\theta = \theta(x, z, h)$ appartient à $[0, 1]$. Ceci utilise une généralisation de la formule de Taylor Mac-Laurin. Commençons par l'écriture de la différentielle première. L'application u est une application de $\mathbb{R}^2$ dans $\mathbb{R}$. Considérons alors l'application $g(t) = u(x+th) - t(u(x+h) - u(x))$. $g(0) = g(1) = 0$. En utilisant le théorème de Rolle, il vient que :

$$\exists \theta \in]0, 1[\text{ tel que } g'(\theta) = 0 \Leftrightarrow u(x + h) = u(x) + u'(a + \theta h)h.$$

Pour montrer l'égalité avec la différentielle seconde, on pose cette fois $g(t) = u(a+th) - u'(a)th$. On peut alors écrire la formule de Taylor avec reste intégral entre 0 et 1 ce qui donne :

$$\begin{aligned} g(1) &= g(0) + g'(0) + \textstyle\int_0^1 tg''(t)dt \\ u(a + h) - u(a) - u'(a)h &= \textstyle\int_0^1 tu''(a + th)(h, h)\end{aligned}$$

En appliquant alors le théorème de la moyenne à l'intégrale, on obtient :

$$u(a + h) = u(a) + u'(a)h + \frac{1}{2}u''(a + \theta h)(h, h)$$

(Exercice, Généraliser le procédé pour une différentielle d'ordre quelconque). Revenons alors à la démonstration. En utilisant l'information des moments (4.1) et la définition des fonctions radiales normalisées, on obtient :

$$|(g_h * u)(x) - u(x) - h\Delta u(x)| \leq Ch^{\frac{3}{2}} \max_{x\in K} \|D^3u(x)\|.$$

Nous allons alors montrer que $T_h u_0 = g_h * u_0$ permet de définir des itérés $((T_h)^n u_0)(x)$ qui tendent vers $u(t, x)$ où $u(t, x)$ est la solution de l'équation de la chaleur. Ce résultat est démontré dans le théorème suivant :

Théorème 2 *Soit $g(x) \in L^1(\mathbb{R}^2)$ une fonction radiale normalisée, non négative satisfaisant en plus la condition du théorème 1 et posons $g_h(x) = \frac{1}{h}g(\frac{x}{h^{\frac{1}{2}}})$ et $T_h u = g_h * u$. Alors, $((T_h)^n u_0)(x) \to u(t, x) \in L^1(K)$ quand $n \to +\infty$ et $nh \to t$, où $u(t, x) = G_t * u_0 \in L^\infty_K$ est solution de l'équation de la chaleur avec comme solution initiale u_0 :*

$$\frac{\partial u}{\partial t} = \Delta u \qquad \int_K |u(t, x) - u_0(x)|dx \to 0 \qquad \textit{lorsque } t \to 0$$

Ce théorème montre que l'itération de filtres passe-bas, isotropes et correctement normalisés est asymptotiquement équivalent à l'équation de la chaleur.
Dém : Nous étudions tout d'abord les propriétés de T_h appliquée à la solution de l'équation de la chaleur. En utilisant le résultat donné par le théorème précédent, en supposant que $t \in [t_1, t_2]$, nous avons que :

$$T_h u(t,x) - u(t,x) = g_h * u(t,x) - u(t,x) = h\Delta u(t,x) + O(h^{\frac{3}{2}})$$

Par ailleurs, comme u est solution de l'équation de la chaleur, nous avons :

$$u(t+h,x) - u(t,x) = h\Delta u(t,x) + O(h^2)$$

ce qui permet d'écrire :

$$T_h u(t,x) - u(t+h,x) = O(h^{\frac{3}{2}})$$

L'opérateur T_h vérifie le principe du maximum, c'est à dire que si $u \leq C$ alors $T_h u \leq C$. En conséquence, $T_h(O(h^\alpha)) = O(h^\alpha)$. On en déduit donc :

$$(T_h)^2 u(t,x) - T_h u(t+h,x) = O(h^{\frac{3}{2}})$$

En utilisant alors l'équation précédente en remplaçant t par $t+h$, il vient :

$$T_h u(t+h,x) - u(t+2h,x) = O(h^{\frac{3}{2}})$$

En additionnant les équations, on obtient :

$$T_h^2 u(t+h,x) - u(t+2h,x) = 2O(h^{\frac{3}{2}})$$

et en itérant le procédé, on obtient :

$$T_h^n u(t,x) - u(t+nh,x) = nO(h^{\frac{3}{2}})$$

à condition que $t_1 \leq t + nh \leq t_2$. En faisant tendre $n \to +\infty$ et en posant $h = \frac{\tau}{n}$, il vient :

$$T_h^n u(t,x) - u(t+\tau,x) = O(h^{\frac{1}{2}}) \tag{4.2}$$

à condition que $t_1 \leq t + \tau \leq t_2$. Si nous pouvions prendre $t = t_1 = 0$, la démonstration serait terminée. Cependant, nous pouvons seulement le fixer très petit. On fixe $\epsilon > 0$ et on considère t_1 suffisamment petit tel que l'on ait :

$$\|u(t_1) - u_0\|_{L^1(C)} = \int_C |u(t_1,x) - u_0(x)|dx < \epsilon$$

D'après les propriétés de la convolution dans $L^1(\mathbb{R}^2)$, nous avons que pour tout $u \in L^1_K$,

$$\|g_h * u\|_{L^1(C)} \leq \|g_h\|_{L^1(\mathbb{R}^2)} \|u\|_{L^1(C)}$$

Comme $\int g_h = 1$, Nous pouvons déduire que

$$\|T_h^n u(t_1,.) - T_h^n u_0\|_{L^1(C)} \leq \epsilon$$

On intégre alors la relation (4.2) sur le compact C avec $t = t_1$ pour obtenir :

$$\|T_h^n u(t_1,.) - u(t_1+\tau,.)\|_{L^1(C)} = O(h^{\frac{1}{2}}) < \epsilon$$

pour n pris suffisamment grand. En combinant les équations précédentes on obtient :

$$\|T_h^n u_0 - u(t_1+\tau)\|_{L^1(K)} \leq 2\epsilon$$

et par suite en prenant t_1 suffisamment petit et $hn = \tau$, on conclut que

$$\|T_h^n u_0 - u(\tau)\| \leq 2\epsilon$$

ce qui termine la démonstration ∎.

4.3 Détection de bords et représentation espace-échelle linéaire

Une définition simple d'un bord consiste à considérer celui-ci comme une courbe fermée de part et d'autre de laquelle l'intensité de niveau de gris varie fortement. Une caractérisation simple des bords consisterait en un gradient de niveau de gris fort et une direction $Du_0(x)$. Cette représentation des bords pose deux problèmes :

a) les régions de fort gradient correspondent à des régions et non pas à des contours fermés.

b) de forts gradients sont observés en des points ne correspondant pas à des structures intéressantes. Ces forts gradients sont notamment dus à du bruit.

Pour éliminer les points de fort gradient correspondant à du bruit, on effectue classiquement un lissage de l'image à une échelle caractéristique t liée au niveau de bruit observé dans l'image. En géneral, cette étape est effectuée en convoluant avec des Gaussiennes.

Pour résoudre le point a) plusieurs approches sont possibles. La première approche dite de Hildreth-Marr consiste à généraliser l'approche monodimensionnelle où le gradient est maximal là où la dérivée seconde s'annule. En 2D, on définit les bords comme les points d'annulation du laplacien. Une seconde approche, dite approche de Canny, consiste à chercher les maxima locaux du module du gradient dans la direction du gradient. En termes mathématiques, cela revient à considérer une fonction g définie par $g(t) = |Du|(x + tDu)$, puis à calculer les points x tels que $g'(0) = 0$. Pour voir à quoi est égale cette dérivée, remarquons que $g = h(c(t))$ avec $c(t) = x + tDu$ et $h(x) = |Du|(x)$. La différentielle de c est Du tandis que celle de h est un peu plus compliquée à calculer. Nous avons :

$$\begin{aligned} h(x+h) &= |Du|(x+h) \\ &= \sqrt{u_{x_1}^2(x+h) + u_{x_2}^2(x+h)} \end{aligned}$$

$$
\begin{aligned}
&= \sqrt{u_{x_1}^2(x) + Du_{x_1}^2(x).h + u_{x_2}^2(x) + Du_{x_2}^2(x).h + o(h)} \\
&= \sqrt{u_{x_1}(x)^2 + u_{x_2}(x)^2}\left(\sqrt{1 + 2\frac{Du(x).(D^2(u)(x)).h}{|Du(x)|^2}} + o(h)\right) \\
&= |Du|(x) + \frac{Du(x).(D^2(u)(x)).h}{|Du(x)|} + o(h).
\end{aligned}
$$

En appliquant la formule de différentiation des fonctions composées, on obtient que :

$$g'(0) = 0 \Leftrightarrow D^2u(\frac{Du}{|Du|}, \frac{Du}{|Du|}) = 0$$

Nous proposons alors deux implémentations pour les détecteurs de Marr-Hildreth et de Canny :
<u>Algo 1 (Marr-Hildreth) :</u>
- Convoluer u_0 avec un noyau Gaussien de taille croissante
- A chaque échelle, repérer les points correspondant à $Du \neq 0$ et Δu change de signe.

<u>Algo 2 (Canny [1]) :</u>
- Convoluer u_0 avec un noyau Gaussien de taille croissante
- A chaque échelle trouver les points x où $Du(x) \neq 0$ et $D^2u(\frac{Du}{|Du|}, \frac{Du}{|Du|})(x)$ passe par zéro.
- Pour chaque échelle t, parmi les points précédemment séléctionnés, ne retenir que ceux satisfaisant $|Du(x)| > \theta(t)$.

Remarques :

1. une implémentation simple consiste à utiliser les différences finies.
2. les calculs des dérivées secondes de u sont toujours possible car u est C^∞ si u_0 est bornée.

4.3.1 Implémentation de l'algorithme de Marr-Hildreth à l'aide de différence de Gaussienne (DOG)

L'opérateur de Marr-Hildreth correspond à la recherche des zéros du Laplacien de l'image préalablement convoluée avec une Gaussienne. On définit donc $b(x, y) = \Delta(u * g(x, y))$. Cela revient à appliquer directement sur l'image l'opérateur laplacien d'une Gaussienne Δg qui peut être valablement approché par la différence de deux Gaussiennes. En effet, en monodimensionnel, l'opérateur DOG s'écrit :

$$DOG = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}} - \frac{1}{\sigma_i\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma_i^2}}$$

En posant $\sigma_i = \sigma + \delta\sigma$ on peut écrire :

$$\sqrt{2\pi}DOG(x) = \frac{1}{\sigma}e^{-\frac{x^2}{2\sigma^2}} - \frac{1}{\sigma + \delta\sigma}e^{-\frac{x^2}{2(\sigma+\delta\sigma)^2}}$$

En différentiant par rapport à σ, il vient que :

$$\begin{aligned}\sqrt{2\pi}DOG(x) &\approx \delta\sigma\frac{\partial}{\partial\sigma}\left(\frac{1}{\sigma}e^{-\frac{x^2}{2\sigma^2}}\right) \\ &= -\left(\frac{1}{\sigma^2}-\frac{x^2}{\sigma^4}\right)e^{-\frac{x^2}{2\sigma^2}} \\ &= \sigma\frac{\partial^2 g(x)}{\partial x^2}\end{aligned}$$

4.4 Calcul des filtres optimaux pour l'estimation des gradients

L'approche de détection des contours proposée par Canny se fonde sur les gradients de l'image. Le calcul des gradients de l'image se fait grâce à des filtres qui doivent avoir certaines propriétés pour garantir un bon résultat numérique. Ces propriétés introduites par Canny [1] et souvent reprises par la suite sont au nombre de 3 :

1. Garantir une bonne détection, c'est à dire une réponse forte même à de faibles contours.
2. Garantir une bonne localisation.
3. Assurer que pour un contour, il y aura une seule détection.

4.4.1 Détermination du filtre optimal dans le cas monodimensionnel

Ces trois critères permettent de définir le filtre optimal pour la détection d'une marche d'escalier sous l'hypothèse d'un filtre additif indépendant du signal. Si l'on suppose que le filtre a une réponse impulsionnelle $h(x)$. Le signal considéré est un signal monodimensionnel représentant un saut d'amplitude U_0 noyé dans un bruit blanc stationnaire $N(x)$ de moyenne nulle et de densité spectrale de puissance N_0^2 :

$$A(x) = U_0 Y(x) + N(x)$$

où Y est la fonction de Heavyside. Le signal de sortie est donné par : $C(x) = A * h(x)$ tel que $C(x)$ soit maximum au point $x = 0$, en respectant les trois contraintes énoncées précédemment.

Le critère de bonne détection s'exprime à l'aide du rapport signal sur bruit (SNR) défini comme le maximum du rapport de la réponse due au signal seul sur la racine carrée de la puissance du bruit, soit :

$$RSB = \frac{U_0\int_0^{+\infty} h(x-t)dt}{\left(E\left\{\left|\int_{-\infty}^{+\infty} N(t)h(x-t)\right|^2\right\}\right)^{\frac{1}{2}}} = \frac{U_0\int_0^{+\infty} h(x-t)dt}{N_0\left(\int_{-\infty}^{+\infty} h^2(t)dt\right)^{\frac{1}{2}}}$$

afin d'obtenir une réponse nulle pour un signal d'entrée constant h est choisie impaire. L'égalité des deux dénominateurs provient de la relation liant les densités spectrales de puissance. Soit $I(x) = N * h(x)$, en appelant S_{II} la densité spectrale de puissance du signal I, nous avons la relation $S_{II} = |\hat{h}(u)|^2 S_{NN}$. On utilise alors le fait que $E\{|I(x)|^2\} = \int_{-\infty}^{+\infty} S_{II}(u)du = \int_{-\infty}^{+\infty} |\hat{h}(u)|^2 S_{NN}du = N_0^2 \int_{-\infty}^{+\infty} |\hat{h}|^2$, d'où le résultat. En se plaçant en $x = 0$, on obtient donc :

$$RSB = \frac{U_0 \int_{-\infty}^{0} h(t)dt}{N_0 \left(\int_{-\infty}^{+\infty} h^2(t)dt\right)^{\frac{1}{2}}} = \frac{U_0}{N_0} \sum$$

Le critère de bonne localisation est mesuré par l'inverse de la variance de la distance entre le maximum de la réponse et la position réelle de la transition (il faudra donc le rendre maximum). Pour le calculer, on cherche à exprimer $E\{x_0^2\}$, x_0 étant la position calculée de la transition. Cette position x_0 correspond au maximum du signal de sortie, donc à un passage par 0 de sa dérivée première puisque le filtre étudié est un dérivateur. Nous avons donc :

$$\left[\frac{\partial C}{\partial x}\right]_{x_0} = \left[\frac{\partial}{\partial x} A * h(x)\right]_{x_0} = [A * h'(x)]_{x_0} = 0$$

d'où :

$$\begin{array}{rcl} \left[\frac{\partial C}{\partial x}\right]_{x_0} & = & U_0 \left[Y * h'(x)\right]_{x_0} + \left[N * h'(x)\right]_{x_0} \\ & = & U_0 \int_{-\infty}^{+\infty} Y(x_0 - t)h'(t)dt + \left[N * h'(x)\right]_{x_0} \\ & = & U_0 \int_{-\infty}^{x_0} h'(t)dt + \left[N * h'(x)\right]_{x_0} \\ & = & U_0 h(x_0) + N * h'(x_0) \end{array}$$

Ensuite, on exprime le développement limité de la fonction $h(x)$ impaire et continue au voisinage de $x = 0$:

$$h(x_0) = h(0) + x_0 h'(0) \approx x_0 h'(0) \text{ comme } h(0) = 0.$$

On en déduit donc :

$$\left[\frac{\partial C}{\partial x}\right]_{x_0} \approx U_0 x_0 h'(0) + N * h'(x_0) = 0.$$

Donc en considérant les espérances,

$$E\left\{|U_0 x_0 h'(0)|^2\right\} \approx E\left\{|N * h'(x_0)|^2\right\},$$

l'expression se réécrit :

$$U_0^2 h'^2(0) E\{x_0^2\} \approx N_0^2 \left(\int_{-\infty}^{+\infty} h'^2(t)dt\right)$$

ce qui donne :

$$E\{x_0^2\} \approx \frac{N_0^2 \left(\int_{-\infty}^{+\infty} h'^2(t)dt\right)}{U_0^2 h'^2(0)}.$$

Le critère de localisation est la racine carrée de l'inverse de l'expression ci-dessus :

$$\frac{U_0}{N_0} \Delta = \frac{U_0 |h'(0)|}{N_0 \left(\int_{-\infty}^{+\infty} h'^2(t)dt\right)^{\frac{1}{2}}}$$

Le produit $\sum \Delta$ est un critère qui combine une bonne détection et une bonne localisation. Il ne dépend pas du facteur d'échelle K. Pour démontrer cette dernière propriété, on pose :$h_K(x) = h(\frac{x}{K})$ et on obtient :

$$\sum_K = \sqrt{K} \sum \text{ et } \Delta_K = \frac{1}{\sqrt{K}} \Delta$$

d'où $\sum_K \Delta_K = \sum \Delta$.

Pour l'optimisation, Canny propose de maximiser le produit $\sum \Delta$ en utilisant une troisième contrainte qui est la minimisation de la densité des maxima liés à un bord. Cette densité des maxima est liée à la densité du nombre de passage par zéro. En moyenne, la distance entre deux maxima consécutifs est égale à deux fois la distance entre deux passages par zéro consécutifs. La distance entre deux zéros consécutifs peut être vue comme l'inverse de la densité. On connait une expression de la densité d_0 de passages par zéro de la réponse due au bruit. Pour un processus Gaussien $B(x)$ de moyenne nulle, on démontre que :

$$d_0 = \frac{1}{\pi} \left(-\frac{R''_{BB}(0)}{R_{BB}(0)} \right)^{\frac{1}{2}}$$

où $R_{BB}(0)$ est la fonction d'autocorrélation du signal. Posons $B(x) = N * h(x)$, où $N(x)$ est un bruit blanc gaussien, nous avons :

$$R_{BB}(0) = N_0^2 \int_{-\infty}^{+\infty} h^2(t)dt$$

Compte tenu de la relation $R_{B'B'}(\tau) = -R''_{BB}(\tau)$, nous pouvons écrire :

$$R''_{BB}(0) = -N_0^2 \int_{-\infty}^{+\infty} h'^2(t)dt.$$

Nous nous intéressons aux transitions du signal d'entrée qui correspondent aux extrema du signal de sortie du filtre dérivateur du signal étudié. Ces extrema sont aussi les passages par zéro de la dérivé seconde du signal d'entrée, donc de la dérivé du signal de sortie. On exploite la formule précédente en remplaçant $B(x)$ par $B'(x)$, pour obtenir :

$$d_0 = \frac{1}{\pi} \left(\frac{\int_{-\infty}^{+\infty} h''^2(t)dt}{\int_{-\infty}^{+\infty} h'^2(t)dt} \right)^{\frac{1}{2}}$$

et on pose $x_{moy} = \frac{1}{d_0}$.

Canny a choisi de prendre un filtre à réponse impulsionnelle finie de taille M. Ceci revient à remplacer les bornes infinies par $-M$ et M. Ensuite, il a imposé comme troisième contrainte que la distance $x_{\max}$ entre deux maxima consécutifs de la réponse due au bruit seul soit égale à une fraction de M, de façon à contrôler le nombre d'extrémums sur un intervalle de longueur

$2M$ centré sur la singularité. Autrement dit, ce critère correspond à la limitation du nombre de maxima locaux détectés en réponse à un seul contour. $x_{\max}$ est alors donné par :

$$x_{\max} = 2x_{moy} = 2\pi \left(\frac{\int_{-M}^{M} h'^2(t)dt}{\int_{-M}^{M} h''^2(t)dt} \right)^{\frac{1}{2}} = k_m M$$

Le problème à résoudre est donc la maximisation de :

$$\sum \Delta = \frac{\int_{-\infty}^{0} h(t)dt}{\left(\int_{-\infty}^{+\infty} h^2(t)dt\right)^{\frac{1}{2}}} \frac{|h'(0)|}{\left(\int_{-\infty}^{+\infty} h'^2(t)dt\right)^{\frac{1}{2}}}$$

Sous la contrainte du nombre d'extrema constant. Etant donné que le filtre h est anti-symmétrique, on peut remplacer dans les expressions faisant intervenir h et ses dérivés, l'intégrale de $-\infty$ à $+\infty$ par l'intégrale de $-M$ à 0. Le principe de la maximisation de $\sum \Delta$ est de minimiser une des intégrales se trouvant au dénominateur sous la contrainte que les autres soient constantes. On est alors ramenés au problème d'optimisation suivant [2] :

$$\begin{array}{cc} \multicolumn{2}{c}{\int_{-M}^{0} h^2(t)dt} \\ \int_{-M}^{0} h(t)dt = c_1 & \int_{-M}^{0} h'^2(t)dt = c_2 \\ \int_{-M}^{0} h''^2(t)dt = c_3 & f'(0) = c_4 \end{array}$$

ce qui correspond à la minimisation d'une fonctionnelle $\int_{-M}^{0} F(t, h, h', h'')dt$ sous des contraintes $\int_{-M}^{0} G_i(t, h, h', h'')dt = c_i$. Ce problème est équivalent à la minimisation du problème sans contrainte :

$$\min_h J(h) = \min_h \int_{-M}^{0} F(t, h, h', h'') + \sum_i \lambda_i G_i(t, h, h', h'')dt$$

Dans notre cas, cette expression se réécrit :

$$\min_h J(h) = \min_h \int_{-M}^{0} h^2 + \lambda_1 h'^2 + \lambda_2 h''^2 + \lambda_3 h dt.$$

On cherche alors la fonction qui annule la différentielle de J qui s'écrit :

$$\begin{array}{rcc} J(h+k) & = & \int_{-M}^{0}(h+k)^2 + \lambda_1(h+k)'^2 + \lambda_2(h+k)''^2 + \lambda_3(h+k)dt \\ & = & \int_{-M}^{0} h^2 + \lambda_1 h'^2 + \lambda_2 h''^2 + \lambda_3 h + 2hk + 2\lambda_1 h'k' + 2\lambda_2 h''k'' + \lambda_3 hk + k\epsilon(k) \\ & = & J(h) + \int_{-M}^{0}(2h - 2\lambda_1 h'' + 2\lambda_2 h'''' + \lambda_3)k + k\epsilon(k) \end{array}$$

La différentielle doit s'annuler pour tout k donc l'équation vérifiée par h est nécessairement :

$$2h - 2\lambda_1 h'' + 2\lambda_2 h'''' + \lambda_3 = 0.$$

La solution de cette équation est la somme d'une solution particulière de l'équation avec second membre et de l'équation homogène. Une solution particulière de l'équation avec second membre est donnée par $h = -\frac{\lambda_3}{2}$ et la solution générale de l'équation sans second membre est donnée par une fonction du type $Ce^{-\gamma x}$ avec γ qui satisfait $2 - 2\lambda_1\gamma^2 + 2\lambda_2\gamma^4 = 0$. On a donc :

$$\gamma^2 = \frac{\lambda_1}{2\lambda_2} \pm \frac{\sqrt{\lambda_1^2 - 4\lambda_2}}{2\lambda_2}. \tag{4.3}$$

L'équation d'Euler fournit une condition nécessaire à l'existence d'un minimum mais cette condition n'est pas forcément suffisante. En effet, l'annulation de la différentielle première ne nous indique pas s'il s'agit d'un minimum ou d'un maximum. Il faut pour cela étudier la différentielle seconde. On utilise alors le théorème de Taylor MacLaurin, sous la forme :

$$J(h + \epsilon k) = J(h) + \epsilon DJ(h).k + \frac{1}{2}\epsilon^2 D^2 J(h + \rho k)(k, k).$$

Il suffit donc d'étudier la différentielle seconde. On calcule les variations de la différentielle première,

$$DJ(h + q)(k) = DJ(h)(k) + 2\int_{-M}^{0} qk + \lambda_1 q'k' + \lambda_2 q''k''$$

Pour que le point d'annulation de la différentielle corresponde bien à un minimum il faut que :

$$D^2 J(h)(k, k) = 2\int_{-M} k^2 + \lambda_1 k'^2 + \lambda_2 k''^2 \geq 0$$

Comme k est une fonction admissible cela se réécrit :

$$\int_{-M}^{0} g^2 - \lambda_1 gg'' + \lambda_2 g''^2 dx \geq \int_{-M}^{0} (g - \tfrac{\lambda_1}{2} g'')^2 + (\lambda_2 - \tfrac{\lambda_1^2}{4}) g''^2 dx \geq 0.$$

Donc si $4\lambda_2 > \lambda_1^2$, la condition est remplie pour toute fonction g. Par ailleurs, cela prouve que nous n'avons que des racines complexes à l'équation (4.3). Nous pouvons alors en déduire la forme générale $\gamma = \pm\alpha \pm i\omega$ avec α et ω réels. Maintenant que nous avons $\gamma^2 = \alpha^2 - \omega^2 \pm 2i\alpha\omega$, en écrivant l'égalité des parties réelles et imaginaires non obtenons :

$$\alpha^2 - \omega^2 = \frac{\lambda_1}{2\lambda_2} \text{ et } 4\alpha^2\omega^2 = \frac{4\lambda_2 - \lambda_1^2}{4\lambda_2^2}$$

Dans ce cas la forme générale de la solution est la suivante :

$$h(x) = a_1 e^{\alpha x}\sin(\omega x) + a_2 e^{\alpha x}\cos(\omega x) + a_3 e^{-\alpha x}\sin(\omega x) + a_4 e^{-\alpha x}\cos(\omega x) + c$$

Pour trouver l'opérateur $h(x)$ sous la forme d'un filtre à réponse impulsionnelle finie et présentant une pente S à l'origine, Canny a imposé les conditions aux limites suivantes :

$$h(0) = 0, h(M) = 0, h'(0) = S, h'(M) = 0.$$

Ces quatre conditions aux limites permettent de déterminer les coefficients de a_1 à a_4 en fonctions de constantes données par l'utilisateur, h étant impaire, la solution est étendue aux x négatifs. La mise en oeuvre pratique du filtre étant complexe, Canny propose d'approximer le filtre par la dérivée première d'une Gaussienne :

$$h(x) \approx -x\exp(-\frac{x^2}{2\sigma^2})$$

Les résultats de détection-localisation sont néanmoins moins bons avec la dérivée de la Gaussienne (20 % de perte) de plus ce filtre n'est pas à réponse impulsionnelle finie.

4.5 Le filtre de Deriche

On préfère souvent le détecteur de Deriche au détecteur de Canny. Celui-ci répond aux mêmes exigences que le filtre de Canny sauf qu'il n'est pas à réponse impulsionnelle finie. Celui-ci à une expression générale de la forme [3] :

$$h(x) = -cx \exp(-\alpha|x|)$$

avec $h(0) = 0, h(\infty) = 0, h'(0) = S$ et $h'(\infty) = 0$. Le paramètre c est calculé de manière à avoir une réponse maximale égale à 1 en zéro lorsque le signal d'entrée est la fonction de Heavyside, soit : $c \int_{-\infty}^{0} t \exp(-\alpha t) dt = 1 \Rightarrow c = -\alpha^2$. Un tel filtre présente de meilleures performances de détection-localisation que le filtre de Canny. Le paramètre α représente l'inverse de l'écart type σ ($\alpha = \frac{\sqrt{\pi}}{\sigma}$).

4.6 Extension au cas bidimensionnel

Le filtrage 2D est obtenu par l'action d'un filtre selon x et d'un filtre selon y. Pour détecter les contours orientés selon Oy on choisit un filtre passe-bas dans la direction Oy et un filtre passe haut dans la direction Ox, typiquement on choisira :

$$f(x, y) = -x \exp(-\alpha^2 x^2) \exp(-\alpha^2 y^2)$$

Ces filtres ne sont pas isotropes,i.e. ils ne répondent pas de manière identiques à des contours similaires mais d'orientation différentes.

4.7 Restrictions à l'approche proposée par Canny

Le premier obstacle à l'approche précédente réside en ce que la définition d'un bord est nécessairement associée à une échelle. La plupart du temps, les bords associés à des structures importantes perdurent aux grandes échelles. En revanche, la localisation spatiale de ces bords est souvent imprécise. L'idée de Witkin [10] est alors de retrouver la localisation spatiale des bords en remontant les échelles. Cependant, l'application de seuillage des gradients aux différentes échelles ainsi que la présence de fluctuations aux petites échelles impliquent que cette méthode est instable à mettre en oeuvre.

Un autre point majeur est le problème de la non invariance par changement de contraste de l'analyse espace-échelle. En effet, la plupart des images sont définies à un changement de contraste près sensé prendre en compte la sensibilité des capteurs. Un opérateur T invariant par changement de constrastes satisfait la propriété que pour toute fonction g croissante on a :

$$g(Tu) = T(g(u)).$$

Ex : la convolution par un noyau k n'est pas invariant par changement de contraste. En effet, pour avoir $g(k * u) = k * g(u)$, on doit avoir $g \in C^\infty$ en supposant k noyau C^∞. Donc toute fonction g croissante et non C^∞ fera l'affaire.

De la même manière, regardons ce qu'il en est de la solution de l'équation de diffusion :

$$\frac{\partial u}{\partial t} = \Delta u$$

A-t-on

$$\frac{\partial g(u)}{\partial t} = \Delta g(u)$$

pour toute fonction g croissante ? Nous avons bien évidemment $\frac{\partial g(u)}{\partial t} = g'(u)\frac{\partial u}{\partial t}$. Pour le calcul du laplacien, on remarque qu'il s'agit d'une fonction composée, on a déjà $Dg(u) = g'(u)Du$. Il s'agit alors de différentier une fonction composée. On considère la composée d'applications $u \to (g'(u), Du) \to g'(u)Du$ dont la différentielle vaut $g'(u)\Delta u + g''(u)|Du|^2$. On en déduit alors

$$\frac{\partial g(u)}{\partial t} - \Delta g(u) = -g''(u)|Du|^2$$

donc l'équation de la chaleur n'est pas invariante par changement de contrastes.

4.8 Premiers algorithmes de détection de contours

Nous avons vu que les filtres de Canny et Deriche permettent de définir les bords. Les bords sont obtenus sans prétraitement de l'image : nous étudions ici des cas simples où l'on effectue d'abord un prétraitement.

4.8.1 Evolution des formes

Koenderink et Van Doorn [6] définissent une forme sur $\mathbb{R}^2$ comme un ensemble X fermé de $\mathbb{R}^2$. Ils proposent de lisser la forme en appliquant l'équation de la chaleur à la fonction caractéristique de l'ensemble X. L'ensemble dont on étudie l'évolution au cours du temps est

$$X_t = \{x, u(t,x) \leq \frac{1}{2}\}$$

le seuil $\frac{1}{2}$ est choisi de façon à ce que $X = \{(x,y) \in \mathbb{R}^2, \quad x \geq 0\} = X_t$. Cette équation d'évolution présente néanmoins quelques inconvénients. Ce modèle ne tient pas pas compte des interactions locales qui peuvent exister entre différents objets (exemple : prendre deux cercles voisins). De nouvelles structures indésirables peuvent apparaître au cours du traitement.

4.8.2 Evolution des courbes

De la même manière, on peut étudier l'évolution des courbes au moyen de l'équation de la chaleur. Soit une courbe initiale définie par $s \in [0, L] \to x_0(s)$. Si on veut lisser la courbe on peut appliquer l'équation de la chaleur à cette courbe c'est à dire :

$$\frac{\partial x}{\partial t}(t,s) = \frac{\partial^2 x}{\partial^2 s}(t,s) \quad s \in [0, L], \quad t \in [0, +\infty[,$$
$$x(0,s) = x_0(s)$$

Cependant, cette équation présente les inconvénients majeurs suivants :

1. Pour $t > 0$, s ne paramétrise plus la longueur de la courbe
2. Les fonctions $x_1(t, s)$ et $x_2(t, s)$ sont des fonctions C^∞ de s pour tout t positif, cela ne veut néanmoins pas dire que $x(t, s)$ est régulière au sens des courbes dans $\mathbb{R}^2$.

Ex : on peut construire une application $\sigma \in [0, 1] \to x(\sigma) \in \mathbb{R}^2$ de telle sorte que $\{x(\sigma), \sigma \in [0, 1]\}$ soit un carré (á voir, exercice maison).

Le problème de l'équation d'évolution basée sur l'équation de la chaleur est que la paramétrisation s à l'instant initial n'est plus valable aux instants suivants. Il est donc nécessaire de reparamétriser la courbe au cours de son évolution. Mackworth-Moktarian [7] propose un algorithme d'évolution des courbes fondé sur un lissage à petite échelle et une reparamétrisation :

1. Convoluer la courbe initiale $x_0(s_0)$ paramétrisée par sa longueur $s_0 \in [0, L_0]$ avec une Gaussienne G_h, h petit.
2. Soit L_n la longueur de la courbe x_n obtenue aprés n itérations et $s_n \in [0, L_n]$ une paramétrisation de la courbe. Pour $n \geq 1$, $\tilde{x}_{n+1}(s_n) = (G_h * x_n)(s_n)$. Reparamétriser $\tilde{x}_{n+1}(s_n) \in [0, L_{n+1}]$ ce qui conduit à une courbe $x_{n+1}(s_{n+1})$.
3. itérer

Le comportement d'un tel algorithme est expliqué grâce au théorème suivant qui est l'analogue du théorème 1 en dimension 1 :

Théorème 3 *Soit $x_0(s_0)$ une courbe C^2 paramétrisée par sa longueur. Lorsque h tend vers 0, il existe une constante c telle que :*

$$(G_h * x_0)(s_0) - x_0(s_0) = ch^2 \frac{\partial^2 x_0}{\partial^2 s_0} + o(h^2).$$

Une conséquence de ce théorème est que lorsque $nh \to s$, les itérés du processus se comporte comme :

$$\frac{\partial x}{\partial t} = \frac{\partial^2 x}{\partial^2 s} = curv(x(s))$$

4.9 Equations de diffusion non linéaire

En chaque point (x, y) où $Du(x, y) \neq 0$, on considère

$$\eta = \frac{Du}{|Du|} \text{ et } \xi = \frac{Du^\perp}{|Du^\perp|}.$$

Ces deux quantités définissent un nouveau système de coordonnées. Appelons maintenant Φ l'angle formé par le gradient et l'axe Ox. On a alors $\eta = \begin{pmatrix} \cos(\Phi) \\ \sin(\Phi) \end{pmatrix}$ Ecrivons alors

$$\begin{array}{rcl} \frac{\partial u}{\partial \eta} & = & \nabla u.\eta \\ & = & \frac{\partial u}{\partial x} \cos(\Phi) + \frac{\partial u}{\partial y} \sin(\Phi). \end{array}$$

Une seconde dérivation conduit à :

$$\begin{array}{rcl} u_{\eta\eta} & = & \frac{\partial^2 u}{\partial x^2}\cos^2(\Phi) + \frac{\partial^2 u}{\partial x \partial y}\cos(\Phi)\sin(\Phi) + +\frac{\partial^2 u}{\partial^2 y}\sin^2(\Phi) \\ & = & \eta^t D^2 u \eta \\ & = & D^2 u(\frac{Du}{|Du|}, \frac{Du}{|Du|}). \end{array}$$

De la même manière, on a $D^2u(\frac{Du^\perp}{|Du|}, \frac{Du^\perp}{|Du|}) = u_{\xi\xi}$.

Définition 1 *Soit u une fonction C^2 définie autour d'un point x de $\mathbb{R}^2$ tel que $Du(x) \neq 0$. Alors, on appelle courbure de u au point x, le nombre réel défini par :*

$$curv(u)(x) = \frac{D^2u(Du^\perp, Du^\perp)(x)}{|Du|^3} = \frac{u_{\xi\xi}}{|Du|} = div(\eta) = div(\frac{Du}{|Du|}).$$

Pour l'avant dernière égalité on procéde comme suit :

$$div(\frac{Du}{|Du|}) = \frac{u_{xx}u_y^2 - 2u_{xy}u_x u_y + u_{yy}u_x^2}{(u_x^2 + u_y^2)^{\frac{3}{2}}}$$

que l'on peut ensuite réécrire sous la bonne forme.

Parmi les défauts multiples de l'équation de la chaleur, un défaut important est de détériorer les bords des images. Une approche novatrice proposée par Perona et Malik [8] consiste à lisser les régions homogènes tout en préservant les bords. La forme générale de l'équation utilisée est la suivante :

$$\frac{\partial u}{\partial t} = div(g(|Du|)Du)$$

avec $g(s) = \frac{1}{1+(\lambda s)^2}$. On montre facilement que si $|Du| \leq \lambda$, il y a diffusion tandis que si $|Du| > \lambda$ il y a anti-diffusion. Nous allons réécrire l'équation de Perona-Malik dans le repère local défini par η et ξ. On a donc :

$$\begin{array}{rcl} div(g(|Du|)Du) & = & \frac{\partial}{\partial x}(g(|Du|)u_x) + \frac{\partial}{\partial y}(g(|Du|)u_y) \\ & = & Dg(|Du|)Du + g(|Du|)\Delta(u) \\ & = & g'(|Du|)u_{\eta\eta}|Du| + g(|Du|)(u_{\eta\eta} + u_{\xi\xi}) \\ & = & \frac{u_{\xi\xi}}{1+(\lambda|Du|)^2} + \frac{(1-\lambda^2(|Du|^2))u_{\eta\eta}}{(1+\lambda^2|Du|^2)^2}. \end{array}$$

Le premier terme apparait donc comme un terme de diffusion dans la direction orthogonale au gradient tandis que le second est un terme de lissage dans la direction du gradient. Ce modèle permet un analyse de l'image en terme de bords et en plus permet la restauration des images par lissage des zones homogènes bruitées. Cependant, il reste un seuil λ à fixer. De nombreuses équation d'évolution non linéaires ont été testées dans les années 90 pour s'affranchir de cette dépendance en λ.

Chapitre 5

Contours Actifs

5.1 Introduction

Une approche très différente des méthodes antérieures de détection de contours à été proposée en 1987 par Kass [5], appelée contours actifs ou "snakes". Il s'agit d'une méthode semi-interactive dans laquelle l'opérateur place dans l'image, au voisinage de la forme à détecter, une ligne initiale de contour. Cette ligne sera amenée à se déformer sous l'action de plusieurs forces :

- Une énergie propre assimilée à l'énergie mécanique de tension et de torsion d'une ligne matérielle.
- Une énergie potentielle imposée par l'image qui vise à plaquer la courbe sur les contours
- Une énergie externe, introduite par l'utilisateur pour traduire les contraintes spécifiques du problème qu'il se pose.

5.2 Formulation mathématique

Sous ces énergies, le contour actif va évoluer pour rechercher la position d'énergie minimale qui sera un compromis entre les diverses contraintes du problème. L'écriture formelle du problème passe par la définition paramètrique du contour, en fonction d'une variable s, généralement l'abscisse curviligne, la variable temporelle t :

$$v(s,t) = [x(s,t), y(s,t)]^t \quad s \in [a,b], \quad t \in [0,T].$$

L'expression de l'énergie que l'on cherche à minimiser comporte 3 composantes, ces énergies dépendent de t ce que l'on écrit pas pour alléger les notations :

$$E_{totale} = E_{interne}(v) + E_{image}(v) + E_{externe}(v)$$

pour récupérer un contour C^2 [5] :

$$E_{interne} = \int_a^b \alpha(s)\|\frac{dv}{ds}\|^2 + \beta(s)\|\frac{d^2v}{ds^2}\|^2 ds.$$

La première dérivé prend en compte les variations de longueur de la courbe (c'est donc un terme de tension (résistance à la rupture), qui est controlé par l'élasticité que l'on attribue au contour), tandis que la seconde exprime les variations de la courbure (c'est un terme de flexion contrôlé par la raideur du contour). La courbe doit être lisse et rigide sans induire de boucles. Ces deux termes agiront donc pour produire une courbe régulière de classe C^2. Dans certain cas, une telle régularité est insuffisante. Le second terme E_{image} caractérise les lignes que l'on souhaite suivre. Dans le cas de la détection de contours, ce sont les lignes de fort gradient, il vaut généralement :

$$E_{image} = -\int_a^b \|\nabla A(v(s))\|^2 ds$$

Enfin, le dernier terme $E_{externe}$ est donné par l'utilisateur. Le plus souvent ce terme permet de contraindre le contour à resembler à un gabarit donné. En l'absence du terme E_{ext}, l'énergie totale peut alors s'écrire :

$$E_{totale} = \int_a^b \left[-\|\nabla A(v(s))\|^2 + \alpha(s)\|\frac{\partial v}{\partial s}\|^2 + \beta(s)\|\frac{\partial^2 v}{\partial^2 s}\|^2 \right] ds$$

On utilise alors un important résultat de calcul variationnel pour calculer le minimum de l'énergie :

Théorème 1 *Soit s une variable à valeurs dans $[a,b]$ et soit v une fonction dépendant de la variable s. Soit $F(s, v, v', v'', \cdots, v^{(n)})$ une fonction de classe C^2 par rapport à tous ses arguments et soit $J(v)$ la fonctionnelle définie par :*

$$J(v) = \int_a^b F(s, v, v', v'', \cdots, v^{(n)}) ds$$

Alors, trouver v extremum de $J(v)$ sous la condition que les extrémités de v et de ses dérivées jusqu'à l'ordre $n-1$ soient fixées, est équivalent à résoudre l'équation d'Euler donnée par :

$$\sum_{m=0}^{n} (-1)^m \frac{\partial^m F_{v^{(m)}}}{\partial s^m}(s, v, v', v'', \cdots, v^{(n)}) = 0$$

où $F_{v^{(m)}}$ désigne la dérivée partielle de F par rapport à $v^{(m)}$.

dém : une condition nécessaire et suffisante à l'existence d'un extremum est que la différentielle s'annule. On peut calculer cette différentielle simplement :

$$\begin{array}{rcl} J(v+k) & = & \int_a^b F(s, v+k, v'+k', v''+k'', \cdots, v^{(n)}+k^{(n)}) ds \\ & = & \int_a^b F(s, v, v', v'', \cdots, v^{(n)}) ds + \int_a^b F_v k + F_{v'} k' + \cdots + F_{v^{(n)}} k^{(n)} ds + |k|\epsilon(k) \\ & = & \int_a^b F(s, v, v', v'', \cdots, v^{(n)}) ds + \int_a^b (-1)^m \frac{\partial^m F_{v^{(m)}}}{\partial s^m} k + |k|\epsilon(k). \end{array}$$

La dernière égalité étant obtenue par intégration par parties et en utilisant le fait que les expressions considérées sont égales en a et b ■.

Dans notre cas, la variable s joue le rôle de l'abscisse curviligne et la variable x celui de l'une des deux variables $x(s)$ et $y(s)$ qui composent $v(s)$. Par simplification d'écriture, on notera :

$$\begin{array}{cc} x' = \frac{dx(s)}{ds} & y' = \frac{dy(s)}{ds} \\ x'' = \frac{d^2x(s)}{d^2s} & y'' = \frac{d^2y(s)}{d^2s} \\ v' = \frac{\partial v(s)}{\partial s} & v'' = \frac{\partial^2 v(s)}{\partial s^2} \end{array}$$

L'énergie à minimiser est de la forme :

$$E_{totale} = \int_a^b \alpha(s)(x'^2 + y'^2) + \beta(s)(x''^2 + y''^2) - \|\nabla A(v(s))\|^2 ds$$

En appliquant le théorème précédent, la minimisation de E_{totale} s'effectue en calculant les solutions de l'équation d'Euler.

$$\begin{array}{c} E_{totale}(v+u) = \int_a^b \alpha(s)\|v' + u'\|^2 + \beta(s)\|v'' + u''\|^2 - \|\nabla i(v(s))\|^2 ds \\ = E_{totale}(v) + \int_a^b 2\alpha(s) < v', u' > +2\beta(s) < v'', u'' > -2 < D^2 i(v)\nabla i(v), u > ds + |u|\epsilon(u) \\ = E_{totale}(v) + 2\int_a^b < -(\alpha(s)v')' + (\beta(s)v'')'' - 2D^2 i\nabla i, u > ds + |u|\epsilon(u). \end{array}$$

Donc l'équation d'Euler est la suivante :

$$-(\alpha v')' + (\beta v'')'' = D^2 A \nabla A.$$

Lorsque α et β sont des constantes et qu'il n'y a pas de potentiel, cette équation s'exprime sous la forme homogène suivante :

$$-\alpha v'' + \beta v^{(4)} = 0$$

Dans le cas où α et β ne sont pas tous deux nuls, une solution particulière $x(s)$ de cette équation homogène est donnée par :

- si $\beta = 0$ alors $x(s) = c_1 s + c_2$
- si $\alpha = 0$ alors $x(s) = c_1 s^3 + c_2 s^2 + c_3 s + c_4$
- si $\alpha \neq 0$ and $\beta \neq 0$ alors $x(s) = c_1 e^{\lambda s} + c_2 e^{-\lambda s} + c_3 s + c_4$ avec $\lambda = \sqrt{\frac{\alpha}{\beta}}$

Un exemple de conditions aux limites permettant d'avoir une courbe à extrémités libres est donné par :

$$v''(a) = v^{(3)}(a) = v''(b) = v^{(3)}(b) = 0$$

5.2.1 Aspects numériques

5.2.2 Formulation stationnaire

La recherche de la courbe réalisant le minimum d'énergie nécessite la résolution de l'équation aux dérivées partielles et ceci passe par une phase de discrétisation du problème à l'aide des différences finies. On considère la courbe discrétisée $\{v_i = (x_i, y_i), i = 0, \cdots, N-1\}$. Au point v_i, les dérivées première et seconde par rapport à x de la composante x du vecteur v sont respectivement approchées par $\frac{x_i - x_{i-1}}{h}$ et $\frac{x_{i+1} - 2x_i + x_{i-1}}{h^2}$ (de même pour y). Les équations d'évolution de x et y sont découplées. Supposant $\alpha_i = \alpha$ et $\beta_i = \beta$ pour toutes les valeurs de i, l'équation $-(\alpha x')' + (\beta x'')'' = -\alpha x'' + \beta x^{(4)} = f$ devient :

$$\begin{aligned}
&-\alpha\tfrac{1}{h^2}\left(x_{i+1} - 2x_i - x_{i-1}\right) + \beta\tfrac{1}{h^4}\left((x_{i+2} - 2x_{i+1} + x_i) - 2(x_{i+1} - 2x_i + x_{i-1}) + (x_i - 2x_{i-1} + x_{i-2})\right) \\
&= f_{i+1} \\
&\Leftrightarrow \beta x_{i-2} + (-4\beta - h^2\alpha)x_{i-1} + (6\beta + 2h^2\alpha)x_i + (-4\beta - h^2\alpha)x_{i+1} + \beta x_{i+2} \\
&= h^4 f_{i+1}.
\end{aligned}$$

On obtient alors un système linéaire de la forme générale $BX = F$ où :

$$\begin{aligned}
X &= (x_i)_{i=0,\cdots,N-2} \\
F &= (h^4 f_{i+1})_{i=0,\cdots,N-2}
\end{aligned}$$

et B une matrice définie ci-dessous. Dans le cas des snakes fermés, compte tenu du caractère périodique des courbes, la matrice B est circulante et s'exprime sous la forme (on prend le pas d'échantillonnage h égal à 1) :

$$B = \begin{pmatrix}
2\alpha+6\beta & -\alpha-4\beta & \beta & 0 & \cdots & \beta & -\alpha-4\beta \\
-\alpha-4\beta & 2\alpha+6\beta & -\alpha-4\beta & \beta & \cdots & 0 & \beta \\
\beta & -\alpha-4\beta & 2\alpha+6\beta & -\alpha-4\beta & \beta & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\
0 & \cdots & \beta & -\alpha-4\beta & 2\alpha+6\beta & -\alpha-4\beta & \beta \\
\beta & 0 & \cdots & \beta & -\alpha-4\beta & 2\alpha+6\beta & -\alpha-4\beta \\
-\alpha-4\beta & \beta & 0 & \cdots & \cdots & -\alpha-4\beta & 2\alpha+6\beta
\end{pmatrix}$$

Dans le cas des snakes fermés($v_0 = v_{N-1}$), les valeurs propres de la matrice B s'exprime par :

$$\lambda_k = \frac{4}{h^2}\sin^2(\frac{k\pi}{N})(4\beta\sin^2(\frac{k\pi}{N}) + h^2\alpha)\forall k \in \{0, \cdots, N-1\}$$

Donc $\lambda_0 = 0$ est valeur propre, ce qui rend la matrice A non inversible. L'inversion directe de la matrice ne peut donc être mise en oeuvre. Par ailleurs la matrice des f_{i+1} dépend de v, on ne peut pas faire de réolution directe.

5.2.3 Procédé itératif de résolution

Du fait que l'énergie à minimiser n'est pas convexe, on ne peut être certain d'accéder à un minimum global et c'est pour cette raison qu'il importe de se situer près de l'objet à segmenter afin de s'assurer que le minimum local que l'on percevra correspondra au contour cherché.

La résolution du problème se fait de manière itérative en étudiant l'équation aux dérivées partielles suivante :

$$\frac{\partial v}{\partial t} - (\alpha v')' + (\beta v'')'' = \frac{\partial \|\nabla A\|^2}{\partial v} \tag{5.1}$$

dont on cherche un point d'équilibre. Avant de passer à l'étude numérique de cette équation, nous allons en faire l'analyse mathématiques.

Approche variationnelle :

On considère un sous-espace V de $L^2[a,b]$ et une fonction u appartenant à V. Ecrivons alors la formulation variationnelle associée. On note $< .,. >$ le produit scalaire dans $L^2[a,b]$. On peut alors écrire :

$$\frac{\partial}{\partial t} < v, u > + \int_a^b (\alpha v'u' + \beta v''u'')ds - [\beta v''u']_a^b - [\alpha v'u]_a^b + [(\beta v'')'u']_a^b =< f, u > .$$

On note $a(v,u)$ la forme bilinéaire définie par :

$$a(v,u) = \int_a^b (\alpha v'u' + \beta v''u'')ds$$

On se ramène à la recherche d'un espace V satisfaisant aux conditions initiales et dans lequel on résout le problème variationnel suivant :
déterminer v :

$$\begin{cases} v : [0,T] \to V \\ \quad t \to v(t) \end{cases}$$

tel que :

$$\forall u \in V, \frac{\partial}{\partial t} < v(t), u > + a(v(t), u) =< f(t), u >$$

Dans le cas des snakes fermés l'espace V est $H_p^2(]a,b[)$, espace des fonctions de $L^2(]a,b[)$ périodiques sur $]a,b[$ ainsi que leur dérivées jusqu'à l'ordre 2. Cette approche permet une résolution par éléments finis du problème.
Calcul direct :
On peut aussi faire les calculs directement sans passer par la formulation variationnelle. On utilise alors les différences finies pour résoudre l'équation (5.1) : les éléments de la courbe sont réduits à des points auxquels sont attachés les éléments mécaniques (masse, raideur, etc) de la courbe concentrés en ces points.

Dans cette approche, on discrétise la courbe en un nombre n de points :

$$V(t) = [v_0(t) \cdots v_{n-1}(t)]$$

et on pose $f_v = -\frac{\partial \|\nabla i(v)\|^2}{\partial v}$.

Deux schèmas sont alors possibles pour la résolution du problème, un schéma dit implicite et un schéma dit explicite. Le schéma explicite présente l'avantage de ne pas nécessiter d'inversion de matrice mais peut en revanche être instable dans certain cas (cf cour première année sur la stabilité des schémas numériques). Le schéma explicite est de la forme suivante :

$$h^4 \frac{v_i^{t+\Delta t} - v_i^t}{\Delta t} + \beta v_{i-2}^t - (4\beta + h^2\alpha)v_{i-1}^t + (6\beta + 2h^2\alpha)v_i^t - (4\beta + h^2\alpha)v_{i+1}^t + \beta v_{i+1}^t = h^4 f_{i+1}^t$$

Chapitre 6

La segmentation en régions

L'approche duale de la détection de contours pour la décomposition d'une image en ses formes élémentaires est l'approche par région. Elle repose sur la recherche de zones possédant des attributs communs. Nous avons dans le chapitre 1 donné toute une liste d'attributs potentiellement intéressants pour la caractérisation des images. Nous supposons dans cette partie que chaque point est caractérisé par un vecteur d'attributs de taille q. Nous nous intéressons dans un premier temps à la représentation efficace des données correspondant à une zone d'intérêt dans l'image, c'est à dire que l'on va chercher une caractérisation de la zone dans un espace de dimension inférieure à q. C'est l'objet de l'analyse en *composantes principales*. Nous introduisons l'*analyse discriminante* qui servira d'outil primordial pour les méthodes de classification. Nous parlerons ensuite de classification supervisée puis nous nous attaquerons au problème de classification non supervisée. La première classe de problèmes de classification suppose que l'on connait le nombre k de classes et l'affectation à l'une classes de certains échantillons, dit échantillons d'apprentissage. La seconde classe de problèmes suppose a contrario que l'on ne connait rien sur l'affectation des échantillons dans l'une des classes, par ailleurs le nombre de classes peut ne pas être connu mais nous n'entrerons pas dans de telles considérations. Pour la classification supervisée, la méthode que nous présentons consiste en la définition d'une fonction discriminante qui utilise directement l'analyse discriminante faite sur les données. Pour la classification non supervisée, nous présenterons les méthodes de nuées dynamiques et celle des k-plus proches voisins. La plupart des informations concernant l'analyse en composantes principales et l'analyse discriminante sont tirées du livre de Saporta [9].

6.1 Description multidimensionnelle de données numériques : analyse en composantes principales

Dans la plupart des applications on observe souvent un nombre élevé de variables q par individu. L'étude séparée de ces variables conduit à négliger les liaisons qui peuvent exister entre elles. Il faut donc analyser les données en tenant compte de leur caractère multidimensionnel ; l'analyse en composantes principales est alors une méthode particulièrement puissante pour explorer la structure de telles données. Les échantillons sont ici supposés caractériser une seule zone d'intérêt.

6.1.1 Les données et leurs caractéristiques

les observations de q variables sur n individus sont rassemblés en un tableau rectangulaire à n lignes et q colonnes, $X = (x_i^j)_{1\leq i\leq n, 1\leq j\leq q}$. L'indice du bas désigne un numéro d'échantillon tandis que celui du haut désigne un attribut. Dans ce qui suit $\bar{x}$ désigne la moyenne des attributs, et V désigne la matrice de covariance associée aux échantillons. On note par ailleurs $D = \frac{1}{n}I$. Les échantillons sont ici supposés caractéristiques d'une seule zone d'intérêt.

Appelons $D_{\frac{1}{s}}$ la matrice diagonale de l'inverse des écart-types, et Y la matrice des variables centrées. Les variables centrées réduites sont alors définies par $Z = YD_{\frac{1}{s}}$. On note R la matrice de covariance des variables centrées réduites. On a bien sûr la relation $R = Z^tDZ = D_{\frac{1}{s}}Y^tDYD_{\frac{1}{s}} = D_{\frac{1}{s}}VD_{\frac{1}{s}}$.

6.1.2 Rôle de la métrique

Une question essentielle est comment mesurer la distance entre deux individus ? Cette question primordiale doit être résolue avant toute étude statistique car les résultats obtenus en dépendent dans une large mesure. La distance euclidienne n'est jamais appropriée car les attributs ont leurs unités propres. De manière générique, la distance entre deux individus est définie par :

$$d^2(x_i, x_j) = (x_i - x_j)^t M(x_i - x_j)$$

où M est une matrice de taille q symétrique définie positive. L'espace des individus est donc muni du produit scalaire :

$$< x_i, x_j >= x_i^t M x_j.$$

La métrique la plus utilisée est la métrique diagonale des inverses des variances, soit :

$$M = D^2_{\frac{1}{s}} = \begin{pmatrix} \frac{1}{s_1^2} & & & 0 \\ & \frac{1}{s_2^2} & & \\ & & \ddots & \\ 0 & & & \frac{1}{s_q^2} \end{pmatrix}.$$

Cette métrique donne à chaque attribut la même importance quelle que soit sa dispersion.

6.1.3 Inertie

On appelle inertie du nuage de points la moyenne des carrés des distances des points au centre de gravité :

$$I_{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^t M(x_i - \bar{x})$$

Si $\bar{x} = 0$, la matrice d'inertie vérifie la propriété suivante :

$$I_{\bar{x}} = Trace(MV) = Trace(VM)$$

dém : l'inertie étant scalaire elle est égale à sa trace :

$$Trace(\sum_{i=1}^{n} \frac{1}{n} x_i^t M x_i) = Trace(\sum_{i=1}^{n} \frac{1}{n} M x_i x_i^t) = Trace(MV) = Trace(VM)$$

6.1.4 L'espace des variables

Nous avons défini une distance sur l'espace des individus, nous devons aussi définir une distance sur l'espace des variables. La distance que l'on utilise (si les deux variables sont centrées) est la suivante :

$$(x^j)^t D x^k = \frac{1}{n} \sum_{i=1}^{n} x_i^k x_i^j = s_{jk}$$

que l'on note s_j^2 si $k = j$. On peut alors définir l'angle θ_{jk} entre deux variables par :

$$\cos(\theta_{jk}) = \frac{< x^j, x^k >}{\|x^j\| \|x^k\|} = \frac{s_{jk}}{s_j s_k}$$

Dans l'espace des variables on s'intéressera plutôt aux angles car se sont des grandeurs normalisée.

6.1.5 Définition de nouveaux axes

Soit Δ un axe engendré par un vecteur unitaire a (de M-norme égale à 1) et projetons les individus sur cet axe (projection M-orthogonale). On a donc :

$$c_i = a^t M x_i = x_i^t M a =< a, x_i >_M$$

On en déduit alors $c = XMa = Xu$ en posant $u = Ma$. A la variable X sont donc associés trois entités mathématiques :
- un axe Δ de F, espace des individus, de vecteur unitaire a
- un vecteur c de E, espace des variables
- une forme linéaire u appelé facteur

Remarquons que si a appartient à l'espace des individus F, u appartient à son dual F^*, et que si a est M-normé à 1, u est M^{-1} normé à 1 :

$$a^t M a = u^t M^{-1} u \text{ car } u = Ma \Rightarrow a = M^{-1} u$$

La variance de c vaut alors :

$$V(c) = s_c^2 = u^t V u$$

6.1.6 Analyse en composantes principales

Le principe de la méthode est d'obtenir une représentation approchée du nuage de point dans un espace de dimension faible. Ceci s'effectue par projection. Le choix du sous-espace de projection F_k s'effectue de telle façon que l'inertie du nuage projeté sur le sous-espace soit maximale.

Soit P l'opérateur de projection M-orthogonale sur F_k : P est tel que $P^2 = P$ et $P^tM = MP$. Le nuage projeté est alors associé au tableau de données XP^t. La matrice de variance du tableau XP^t est pour des variables centrées :

$$(XP^t)^tD(XP^t) = PVP^t$$

L'inertie du nuage projeté vaut donc : $Trace(PVP^tM)$. Par des opérations élémentaires, on en déduit :

$$\begin{array}{rcl} Trace(PVP^tM) & = & Trace(PVMP) \text{ car } P^tM = MP \\ & = & Trace(VMP^2) \text{ car } Trace(AB) = Trace(BA) \\ & = & Trace(VMP) \text{ car } P \text{ est idempotent} \end{array}$$

Le problème est donc de trouver P, projecteur M-orthogonal de rang k maximisant $Trace(VMP)$ ce qui déterminera F_k. Si F et G sont deux sous-espaces M-orthogonaux alors :

$$I_{F\oplus G} = I_F + I_G$$

Il suffit de remarquer que le projecteur associé à la somme directe de deux sous-espaces M-orthogonaux est la somme des projecteurs associés à chacun des espaces. De ce résultat on déduit le théorème fondamental suivant :

Théorème 1 *Soit F_k un sous-espace portant l'inertie maximale, alors le sous-espace de dimension $k+1$ portant l'inertie maximale est la somme directe de F_k et du sous-espace de dimension 1 M-orthogonal à F_k portant l'inertie maximale : les solutions sont emboitées.*

dém : Soit E_{k+1} un sous-espace de dimension $k+1$. Comme $dim(E_{k+1}) = k+1$ et $dim(F_k^{\perp}) = n-k$, on a :

$$dim(E_{k+1} \cap F_k^{\perp}) \geq 1$$

Soit b un vecteur appartenant à $E_{k+1} \cap F_k^{\perp}$. Posons $E_{k+1} = b \oplus G$ où G est le supplémentaire M-orthogonal de b dans E_{k+1}. G est donc de dimension k et $F = F_k \oplus b$. on a :

$$\begin{array}{rcl} I_{E_{k+1}} & = & I_b + I_G \\ I_F & = & I_{F_k} + I_b \end{array}$$

Comme F_k était le sous-espace de dimension k portant l'inertie maximale $I_G \leq I_{F_k}$, donc $I_{E_{k+1}} \leq I_b + I_{F_k}$, c'est à dire $I_{E_{k+1}} \leq I_F$ et ceci quel que soit E_{k+1}. Le maximum de l'inertie est donc réalisé pour l'espace $F = b \oplus F_k$ et b doit être tel que I_b soit maximal. Pour obtenir F_k, on pourra donc procéder de proche en proche en cherchant d'abord le sous-espace de dimension 1 d'inertie maximale, puis le sous-espace de dimension 1 M-orthogonal au précédent d'inertie maximale, etc.

6.1.7 Axes principaux, composantes principales, facteurs principaux

Axes principaux :

Nous devons chercher la droite de $\mathbb{R}^q$ passant par $\bar{x}$ maximisant l'inertie du nuage projeté sur cette droite. Soit a un vecteur porté par cette droite, le projecteur M-orthogonal sur la droite est alors :

$$P = a(a^tMa)^{-1}a^tM$$

L'inertie du nuage projeté sur cette droite vaut, d'après ce qui précéde :

$$\begin{aligned} Trace(VMP) &= Trace(VMa(a^tMa)^{-1}a^tM) \\ &= \frac{1}{a^tMa}Trace(VMaa^tM) = \frac{Trace(a^tMVMa)}{a^tMa} = \frac{a^tMVMa}{a^tMa} \end{aligned}$$

Pour obtenir le maximum de $\frac{a^tMVMa}{a^tMa}$, il suffit d'annuler la différentielle :

$$\frac{d}{da}\left(\frac{a^tMVMa}{a^tMa}\right) = \frac{(a^tMa)2MVMa - (a^tMVMa)2Ma}{(a^tMa)^2} = 0$$

d'où :

$$MVMa = \left(\frac{a^tMVMa}{a^tMa}\right)Ma$$

soit :

$$VMa = \lambda a$$

car M est régulière, donc a est vecteur propre de VM. S'il en est ainsi le critère $\frac{a^tMVMa}{a^tMa}$ vaut λ. Il faut donc que λ soit la plus grande valeur propre de VM. On a ainsi le résultat suivant :

Théorème 2 *Le sous-espace F_k de dimension k est engendré par les k vecteurs propres de VM associés aux k plus grandes valeurs propres.*

On appelle axes principaux d'inertie les vecteurs propres de VM, M-normés à 1. Ils sont au nombre de q. On montre que les axes principaux sont M et V^{-1} orthogonaux. Notons tout de suite que si a et b sont deux axes principaux M orthogonaux alors comme $Mb = \beta V^{-1}b$, il vient $a^tMb = \beta a^tV^{-1}b$ donc les deux axes sont V^{-1} orthogonaux. Montrons qu'il est possible de construire une base M orthonormée de vecteurs propres. Supposons que Soit λ une valeur propre de VM alors $Ker(VM - \lambda I)$ est stable par VM par définition et $Ker(VM - \lambda I)^\perp$ est stable par VM. En effet,

$$x \in Ker(VM-\lambda I)^\perp \quad \Leftrightarrow \forall y \in Ker(VM-\lambda I) \quad <VMx, y>_M = <x, VMy>_M = \lambda <x, y>_M = 0.$$

Donc, pour construire une base M-orthonormée on construit une base M-orthonormée de $Ker(VM - \lambda I)$ que l'on concatène avec une base M-orthonormée de $Ker(VM - \lambda I)^\perp$. On montre réciproquement que les axes principaux sont le seul système de vecteurs à la fois M et V^{-1} orthogonaux.

Facteurs principaux :

A l'axe a est associée la forme linéaire u coordonnée M-orthogonale sur l'axe défini par a. u est un élément de $\mathbb{R}^{q*}$ (dual de l'espace des individus) qui définit une combinaison linéaire des variables descriptive $x_1, x_2, \cdots, x_q$. A l'axe principal a M-normé à 1 est associé le facteur principal $u = Ma$. Puisque a est vecteur propre de VM :

$$VMa = \lambda a \Rightarrow MVMa = \lambda Ma$$

soit :

$$MVu = \lambda u$$

Les facteurs principaux sont les vecteurs propres M^{-1}-normés à 1 de MV. En effet, on a vu que si $\mathbb{R}^q$ est muni de la métrique M, son dual doit être muni de la métrique M^{-1}. Donc $u^t M^{-1} u = 1$. Les facteurs principaux sont M^{-1} et V orthogonaux.

Composantes principales :

Ce sont les variables c définies par les facteurs principaux : $c = Xu$. c est le vecteur renfermant les coordonnées des projections M-orthogonales des individus sur l'axe défini par a avec a unitaire. La variance d'une composante principale est égale à la valeur propre λ :

$$V(c) = \lambda$$

En effet, $V(c) = c^t Dc = u^t X^t DXu = u^t Vu$. Or, $Vu = \lambda M^{-1} u$, donc :

$$V(c) = \lambda u^t M^{-1} u = \lambda$$

Les c sont les combinaisons linéaires de $x_1, x_2, \cdots, x_q$ de variance maximale sous la contrainte $u^t M^{-1} u = 1$. Les composantes principales sont elles-mêmes vecteurs propres d'une matrice de taille n : en effet, $MVu = \lambda u$ s'écrit $MX^t DXu = \lambda u$, en multipliant à gauche par X et en remplaçant Xu par c, on a :

$$XMX^t Dc = \lambda c$$

D'où pour résumer :

Facteurs principaux u	$MVu = \lambda u$	M^{-1}-orthonormés
Axes principaux a	$VMa = \lambda a$	M-orthonormés
Composantes principales c	$XMX^t Dc = \lambda c$	D-orthogonales
$c = Xu$	$u = Ma$	

En pratique on calcule les u par diagonalisation de MV, puis on obtient les $c = Xu$, les axes principaux n'ayant pas d'intérêt pratique.

6.1.8 Formules de reconstruction

Comme $Xu_j = c_j$, en multipliant à droite les deux membres par $u_j^t M^{-1}$ et en sommant sur j, il vient :

$$X \sum_j u_j u_j^t M^{-1} = \sum_j c_j u_j^t M^{-1}$$

Or $\sum_{j=1}^{p} u_j u_j^t M^{-1} = I$ car les u_j sont M^{-1} orthonormés, il suffit de voir que :

$$\left(\sum_{j=1}^{p} u_j u_j^t M^{-1}\right) u_i = u_i \text{ car } u_j^t M^{-1} u_i = \delta_{ij}$$

donc :

$$X = \sum_{j=1}^{p} c_j u_j^t M^{-1}$$

On peut ainsi reconstituer le tableau de données (centré) au moyen des composantes principales et facteurs principaux.

6.2 Analyse factorielle discriminante

Les données consistent maintenant en n observations réparties en k classes et décrites par q variables explicatives. On suppose que l'on connait la classe d'affectation des n échantillons

6.2.1 Données et Notations

Les n individus x_i de l'échantillon constituent un nuage E de $\mathbb{R}^q$ partagé en k sous-nuages $E_1, E_2, \cdots, E_k$ de centre de gravité $g_1, g_2, \cdots, g_k$, de matrices de variances $V_1, V_2, \cdots, V_k$. On pose $q_1, q_2, \cdots, q_k$ les proportions des échantillons dans les différentes classes. Soit $g = \sum_{j=1}^{k} q_j g_j$ le centre de gravité et V la matrice de variance de E tout entier.
On appelle *matrice de variance interclasse*, la matrice de variance B des k centres de gravité :

$$B = \sum_{j=1}^{k} q_j (g_j - g)(g_j - g)^t$$

et *matrice de variance intraclasse* W la moyenne des matrices V_j

$$W = \sum_{j=1}^{k} q_j V_j.$$

On a la relation suivante :

$$V = W + B$$

qui constitue une généralisation de la relation classique : variance totale = moyenne des variances + variance des moyennes. La dḿonstration de l'égalité précédente est la suivante :

$$\begin{aligned} W + B &= \sum_{j=1}^{k} q_j (g_j - g)(g_j - g)^t + \frac{1}{n} \sum_{x \in C_j} (x - g_j)(x - g_j)^t \\ &= \frac{1}{n} \sum_{x=1}^{n} (x - g)(x - g)^t \end{aligned}$$

Nous supposerons maintenant que $g = 0$, c'est à dire que les variables explicatives sont centrées. Si on considère que le tableau de données à étudier se met sous la forme :

$$\left[\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & \\ & & & & \\ & & A & & X \\ 0 & 0 & \cdots & 1 & \end{array}\right]$$

où X est la matrice des q variables explicatives et A le tableau logique associé à la variable qualitative, les k centres de gravité $g_1, g_2, \cdots, g_k$ sont les lignes de la matrice $(A^tDA)^{-1}(A^tDX)$. A^tDA est la matrice des proportions q_j des sous-nuages :

$$A^tDA \quad = \quad D_q \quad = \quad \begin{bmatrix} q_1 & 0 & \cdots & 0 \\ 0 & q_2 & 0 & \cdots \\ 0 & \cdots & \ddots & 0 \\ 0 & \cdots & 0 & q_k \end{bmatrix}$$

La matrice de variance interclasse s'écrit alors si $g = 0$:

$$\begin{aligned} B &= ((A^tDA)^{-1}A^tDX)^tA^tDA((A^tDA)^{-1}A^tDX) \\ &= X^tDA(A^tDA)^{-1}A^tDX = (X^tDA)D_q^{-1}(A^tDX) \end{aligned}$$

6.2.2 Analyse factorielle discriminante

L'analyse factorielle discriminante consiste à rechercher de nouvelles variables (les variables discriminantes) correspondant à des directions de $\mathbb{R}^q$ qui séparent le mieux possible en projection les k groupes d'observations. Supposons $\mathbb{R}^q$ muni d'une métrique M. On note toujours un axe discriminant a, le facteur associé $u = Ma$, la variable discriminante sera Xu. En projection sur l'axe a, les k centres de gravité doivent être aussi séparés que possible, tandis que chaque sous-nuage doit se projeter de manière groupée autour de son centre de gravité. En d'autres termes, l'inertie du nuage des g_j projetés sur a doit être maximale. La matrice d'inertie du nuage des g_j est BM, l'inertie du nuage projeté sur a est a^tMBMa si a est M-normé à 1. Il faut aussi qu'en projection sur a, chaque sous nuage reste bien groupé, donc que a^tMV_jMa soit faible pour $j = 1, 2, \cdots, k$. On cherchera donc à minimiser la moyenne

$$\sum_{j=1}^{k} q_j a^tMV_jMa$$

soit a^tMWMa. Or la relation $V = B + W$ entraine que $MVM = MBM + MWM$ donc que :

$$a^tMVMa = a^tMBMa + a^tMWMa$$

On prendra alors comme critère de maximiser le rapport de l'inertie interclasse à l'inertie totale. Soit :

$$\max_a \frac{a^tMBMa}{a^tMVMa}$$

On sait que ce maximum est atteint si a est vecteur propre de $(MVM)^{-1}MBM$ associé à sa plus grande valeur propre λ_1 :
$$M^{-1}V^{-1}BMa = \lambda a$$
A l'axe dicriminant a est alors associé le facteur discriminant u, tel que $u = Ma$. On a alors $V^{-1}Bu = \lambda_1 u$. Les facteurs discriminants, donc les variables discriminantes Xu, sont indépendants de la métrique M. On choisira par commodité $M = V^{-1}$ qui donne $BV^{-1}a = \lambda a$ et $V^{-1}Bu = \lambda u$.

6.2.3 Une analyse en composantes principales particulière

D'après les équations précédentes, l'analyse factorielle discriminante, n'est autre que l'ACP du nuage des k centres de gravité avec la métrique V^{-1}. On en déduit que les variables dicriminantes sont non corrélées deux à deux.

6.2.4 Analyse de variance et métrique W^{-1}

Comme il y a q variables, on peut rechercher la combinaison linéaire définie par des coefficients u donnant la valeur maximale pour la statistique de Fisher (rapport entre variance interclasse sur variance intraclasse) ce qui revient à maximiser :

$$\frac{u^tBu}{u^tWu}$$

La solution est donnée par l'équation :

$$W^{-1}Bu = \mu u \text{ avec } \mu \text{ maximal}$$

Les vecteurs propres de $W^{-1}B$ sont les mêmes que ceux de $V^{-1}B$ avec $\mu = \frac{\lambda}{1-\lambda}$. En effet, $Bu = \lambda Vu$ est équivalent à :

$$Bu = \lambda(W + B)u \text{ soit } (1 - \lambda)Bu = \lambda Wu$$

d'où : $W^{-1}Bu = \frac{\lambda u}{1-\lambda}$.

Si $0 \leq \lambda \leq 1$, on a en revanche $0 \leq \mu \leq \infty$ et $\lambda = \frac{\mu}{1+\mu}$. L'utilisation de V^{-1} ou de W^{-1} comme métrique est donc indifférent. La métrique W^{-1} est appelée métrique de "Mahalanobis".

6.2.5 Normalisation des vecteurs propres

La convention usuelle dans la plupart des logiciels est d'avoir des variables discriminantes dont la variance intraclasse vaut 1. On doit donc avoir $u^tWu = 1$. Ce qui revient à $u^tBu = \frac{\lambda}{1-\lambda} = \mu$ et à $u^tVu = \frac{1}{1-\lambda}$ car $u^tBu = u^t\lambda(W + B)u = \lambda u^tVu$.

6.3 Méthodes multidimensionnelles de classification

Les méthodes multidimensionnelles consistent à classifier des individus en fonction d'un ensemble d'attributs. On utilisera dans ce paragrahe les notations suivantes :
E est l'ensemble des individus (card E = n)
C est l'ensemble des classes (card C = k)
Q est l'ensemble des attributs (card Q = q)
A chaque individu est associé un vecteur de q attributs. C'est dans l'espace des attributs ou espace de representation (de dimension q) que la classification s'effectue.

Il existe deux types de méthodes de classification multidimensionnelle. La classification *supervisée* ou classification avec apprentissage qui consiste à construire à partir d'une classification connue a priori d'un certain nombre d'individus, une *fonction d'identification* ou de *discrimination* pour les autres individus. Cette fonction d'identification réalise un découpage de l'espace de représentation. La *classification non supervisée*, consiste à découper l'espace de représentation en zone homogènes selon un critère de ressemblance entre les individus, c'est à dire un critère de proximié dans l'espace des attributs.

Si l'on a des connaissances sur la classe d'appartenance de certains individus, on utilisera le premier type de méthodes. Si ce n'est pas le cas, il faut définir une mesure de proximité entre les attributs et utiliser la deuxième approche.

6.3.1 Classification supervisée

Classification non paramètrique
On se place dans le sous-espace défini pas les r premiers axes discriminants. On normalise ces axes à 1 selon la norme définie par W ou V. On calcule les variables discriminantes associées à la projection des données sur les axes discriminants. On est alors dans un espace de dimension r. On calcule la matrice W associée aux variables discriminantes et l'on utilise comme règle de décision :

$$\mathcal{C}(x) = \underset{j}{Argmin}(c - c_j)^T W^{-1}(c - c_j)$$

Où c est la variable discriminante, et c_j la variable discriminante moyenne sur la classe j.
Approche Bayésienne pour la classification
On considère que chaque classe j est munie d'une densité de probabilité $f_j(v)$, à partir de laquelle on peut définir une probabilité d'appartenance au groupe j donnée par la formule de Bayes :

$$P(v|c = j) = \frac{q_j f_j(v)}{\sum\limits_{j=1}^{k} q_j f_j(v)}$$

La règle consiste alors à affecter l'observation v au groupe qui a la probabilité maximale, c'est à dire, comme les dénominateurs sont égaux, à calculer :

$$\max_j q_j f_j(v)$$

Lorsque v est un vecteur Gaussien qui suit une loi $N(g_j, V_j)$, où g_j est le centre de gravité de la classe j et V_j la matrice de variance de la classe j, la densité f_j est définie par :

$$f_j(v) = \frac{1}{(2\pi)^{\frac{k}{2}}|det(V_j)|^{\frac{1}{2}}|} \exp(-\frac{1}{2}(v - g_j)^t V_j^{-1}(v - g_j)).$$

La règle Bayésienne consiste alors à minimiser :

$$(v - g_j)^t V_j^{-1}(v - g_j) - 2\log(q_j) + \log(|det(V_j)|)$$

L'approche Gaussienne est dans de nombreux cas non satisfaisante pour représenter des distributions d'attributs. De nouveaux modèles d'analyse se fondent sur les descriptions des attributs en terme de distribution Gaussienne généralisée. Pour vérifier l'adéquation du modèle Gaussien on peut utiliser comme test simple le test de la droite de Henry (cours de statistiques de deuxième année) et calculer l'asymétrie et le curtosis définis pour un échantillon de taille n par :

$$\begin{aligned} \text{curt}(x_1, \cdots, x_n) &= \tfrac{1}{n} \sum_{j=1}^{n} \left(\tfrac{x_j - \bar{x}}{\sigma}\right)^4 - 3 \\ \text{asym}(x_1, \cdots, x_n) &= \tfrac{1}{n} \sum_{j=1}^{n} \left(\tfrac{x_j - \bar{x}}{\sigma}\right)^3 \end{aligned}$$

K-plus proches voisins

La méthode des K-plus proches voisins est un moyen simple d'estimation non paramétrique de la classe d'un point x. Elle consiste simplement à affecter à x la classe majoritaire parmi les K-plus proches voisins de x, au sens d'une distance à préciser. Le résultat de discrimination dépend de la valeur de K. Les inconvénients de cette méthode réside dans le cout du calcul des K plus proches voisins car il faut prendre en compte chaque élément de l'ensemble d'apprentissage.

6.3.2 Classification non supervisée

Comme il n'y a pas d'ensemble d'apprentissage, on ne peut pas sélectionner les attributs les plus pertinents. Certains attributs peuvent donc nuire au résultat de classification. Nous présentons ici la méthode des nuées dynamiques. Celle-ci consiste à calculer la partition optimale de l'ensemble des individus en k sous-ensembles, chaque sous-ensemble étant représenté par un noyau (ou centre de la classe). On note :

$\prod$ l'ensemble des partitions π de E en k classes $(C_1, \cdots, C_k)$,

$g = (g_1, \cdots, g_k)$ un ensemble de k noyaux, où g_i est le noyau de la classe C_i.

$W(\pi, g)$ un critère qui mesure l'adéquation entre une partition π et l'ensemble des noyaux g.

On cherche :

$$(\pi^*, g^*) = \underset{(\pi,g)}{argmin} \{W(\pi, g)\}$$

c'est à dire la partition π et la représentation g associée qui optimisent leurs adéquations au sens du critère W.

Nous présentons ici les nuées dynamiques dans le cas où les noyaux ont comme mode de représentation le centre d'inertie de la classe C_i. Le critère d'adéquation est alors :

$$W(\pi, g) = \sum_{i=1}^{Nc} \sum_{x \in C_i} d^2(x, g_i)$$

où d représente la distance euclidienne. W est donc la variance intraclasse (avec des poids identiques pour chacune des classes). On définit alors deux fonctions :

- Une fonction de représentation $h : \pi \rightarrow g$ qui connaissant une partition π, calcule la représentation unique g associée (en calculant chaque centre de gravité g_i de C_i).
- Une fonction d'affectation $f : g \rightarrow \pi$ qui connaissant la représentation g calculera la partition π selon :

$$C_i = \{x \in E, \quad \forall j \neq i, \quad d(x, g_i) < d(x, g_j)\}$$

  et si x est équidistant de g_i et g_j, il est intégré à la classe de plus faible indice.

L'algorithme de calcul est le suivant :

$$\begin{cases} \pi^{n+1} & = & f(g^n) \\ g^{n+1} & = & h(\pi^{n+1}) \\ u_{n+1} & = & W(\pi^{n+1}, g^{n+1}) \end{cases}$$

La suite u_n converge en décroissant (car $u_n = W(\pi^n, g^n) \geq W(\pi^{n+1}, g^n) \geq W(\pi^{n+1}, g^{n+1}) = u_{n+1}$) vers la solution u^* qui est optimal pour g^0 donné.
Remarque : la partition finale dépend des noyaux initiaux (g^0), pour avoir un rsultat fiable il faut appliquer l'algorithme un nombre important de fois.

Bibliographie

[1] J. Canny, *A Computational Approach to Edge Detection*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, 679-698, 1986.

[2] J-P coquerez and S. Philipp, *Analyse d'images : filtrage et segmentation*, Enseignement de la physique,masson,1995.

[3] Deriche R., *Using Canny's criteria to derive recursively implemented optimal edge detector*, International Journal of Computer Vision, vol. 1, pp. 167-187, 1987.

[4] R. Haralick, *Statistical and structural approaches to textures*, Proceedings of the IEEE,vol. 67, no 5, 786-804, 1979.

[5] Kass M., Witkin A., Terzopoulos D., *Active Contour Models*, International Journal of computer Vision,vol. 1, pp. 321-331, 1988.

[6] Koenderink J,van Doorn A.J, *The structure of images*, Biological Cybernetics, vol. 50, pp. 357-393, 1989.

[7] Mackworth A, Mokhtarian F., *A theory of Multiscale, curvature-based shape representation for planar curves*, IEEE Transactions on Pattern Analysis and Machine Intelligence,vol. 14, pp. 789-805, 1992.

[8] Perona. P and malik. J, *Scale Space and Edge Detection using Anisotropic Diffusion*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, pp. 629-639, 1990.

[9] Saporta G., *Probabilités, Analyse de données statistiques*, 1990,Editions Technip.

[10] A. Witkin, *Scale-space Filtering*, International Joint Conference on Artificial Intellignece, pp. 1019-1023, 1983.