

# Chapitre I

## Champs de Markov en analyse d'images

### 1. Introduction

La recherche sur les machines de vision est menée parallèlement sur deux fronts : le premier est fortement pluridisciplinaire, et l'on trouve des spécialistes en mathématiques appliquées travaillant en étroite collaboration avec des neurobiologistes et des cognitivistes pour tenter d'appréhender les mécanismes de vision biologiques résultant d'une très longue évolution naturelle. Des ingénieurs tentent ensuite de reproduire artificiellement - avec plus ou moins de bonheur - quelques fonctions du système de vision biologique (on parle de systèmes *neuromimétiques*). La deuxième approche est plus abstraite et mathématique. Elle est généralement désignée par l'expression *vision artificielle* et consiste à considérer les images comme des *signaux* pour en extraire des *primitives* à l'aide d'algorithmes appropriés.

Etant donnée l'efficacité des systèmes de vision naturels, la première démarche semble a priori très prometteuse. Cependant, malgré la recherche en neurobiologie qui avance à pas de géant, la compréhension des mécanismes intervenant dans la vision des vertébrés est loin d'être complète (voir *fig.I.1*). Les espoirs sont finalement déçus : à l'heure actuelle, on peut dire que seule la fonction de capteur d'images est très correctement reproduite artificiellement ; mais même dans ce cas, c'est sans compter avec le système d'asservissement automatique de la focale du cristallin (*accommodation*), ou le mécanisme de compression logarithmique qui vaut à l'œil humain l'incroyable exploit d'une dynamique pouvant couvrir de 8 à 10 ordres de grandeur. Par ailleurs, les composants biologiques à la base des systèmes de vision biologiques (ou *wetware*) ne trouvent pas d'équivalent simple (et/ou bien maîtrisé) en technologie électronique, et surtout pas en électronique *numérique*<sup>6</sup>.

---

<sup>6</sup> Le traitement de l'information par le système nerveux se fait selon des principes totalement étrangers aux techniques numériques familières aux ingénieurs. Le signal n'est pas traité de manière digitale et les

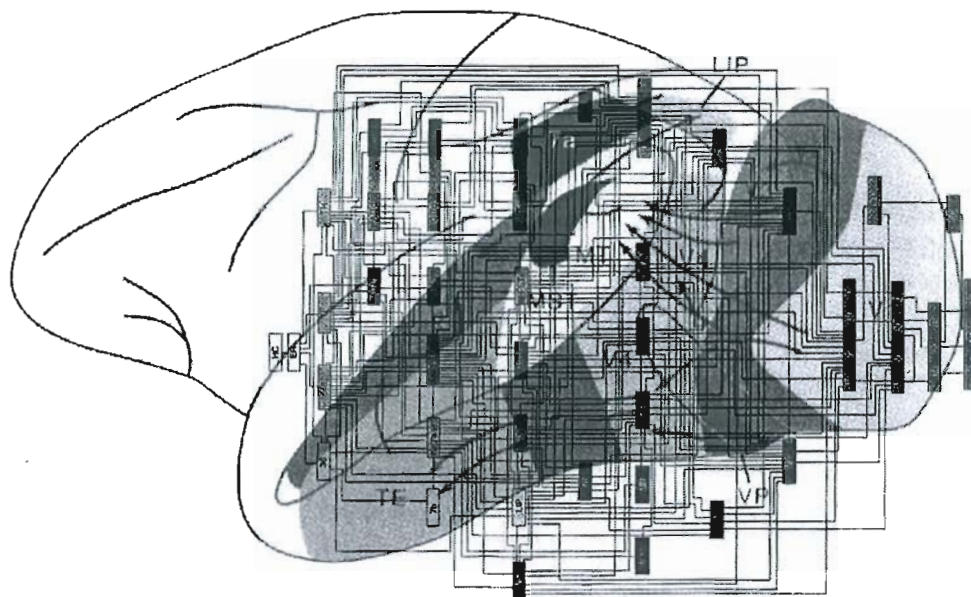


Fig.I.1 : Une bonne partie du cerveau des vertébrés supérieurs - un tiers du cortex chez l'humain - est consacrée à des tâches relatives à la vision (aires grisées dans la figure, représentant un cerveau de macaque). Bien que la "cartographie" des circuits neuronaux impliqués dans la vision soit aujourd'hui extrêmement précise (circuit superposé), la compréhension des principes et mécanismes sous-jacents est loin d'être complètement élucidée. Représentation simplifiée et volontairement 'artistique' à partir de [Essen92].

En revanche, la recherche en vision artificielle évolue plutôt au rythme des progrès de l'informatique et de la recherche en traitement du signal. De nombreux algorithmes qui ont fait leurs preuves ont vu le jour depuis quelques années. L'avantage de cette approche réside surtout dans le fait que l'on peut procéder par étapes, en posant à chaque fois le problème en termes mathématiques. La puissance de ces algorithmes reste cependant liée à la rapidité des machines plus ou moins classiques sur lesquelles ils s'exécutent. Une des étapes *critiques* de la vision artificielle - extrêmement gourmande en calcul - est celle que l'on appelle vision *bas-niveau* (*early vision* en anglais). Dans cette étape on s'intéresse à l'extraction des toutes premières primitives de l'image brute : le programme est confronté à un flot de données énorme et doit faire le tri très vite entre le pertinent et l'accessoire pour que l'information soit accessible en *temps-réel*<sup>7</sup>. Tout comme dans les systèmes

---

opérations de base ne sont pas booléennes : le système nerveux exploite *directement* les phénomènes physico-chimiques élémentaires de ses "composants" comme primitives de calcul, alors que l'ordinateur conventionnel ne profite des non-linéarités des transistors que de façon très frustrante, par l'opération de seuillage [Mhani97].

<sup>7</sup> On dit qu'un système de traitement est *temps-réel* (*real-time*) quand les résultats sont obtenus en un temps déterminé, indépendant des caractéristiques particulières de l'échantillon de données à traiter - autrement dit, le flux de données de sortie correspond au flux de données d'entrée à un décalage temporel constant près.

biologiques, il s'opère dans cette étape une réduction considérable de l'information<sup>8</sup> qui aboutit à des représentations abstraites très simples du contenu de l'image ou *primitives*, représentations fragmentaires et pas immédiatement reliées entre elles (contours [Koch86], régions et textures [Geman84], champs de mouvement [Poggio85], [Heitz92], profondeur des objets en stéréovision [Marroq87], etc.). A partir de là, des fonctions cognitives supérieures prennent tour à tour le relais, en fusionnant ces données à l'aide de modèles a priori plus ou moins complexes dans le but ultime de construire une représentation *haut niveau* du monde extérieur.

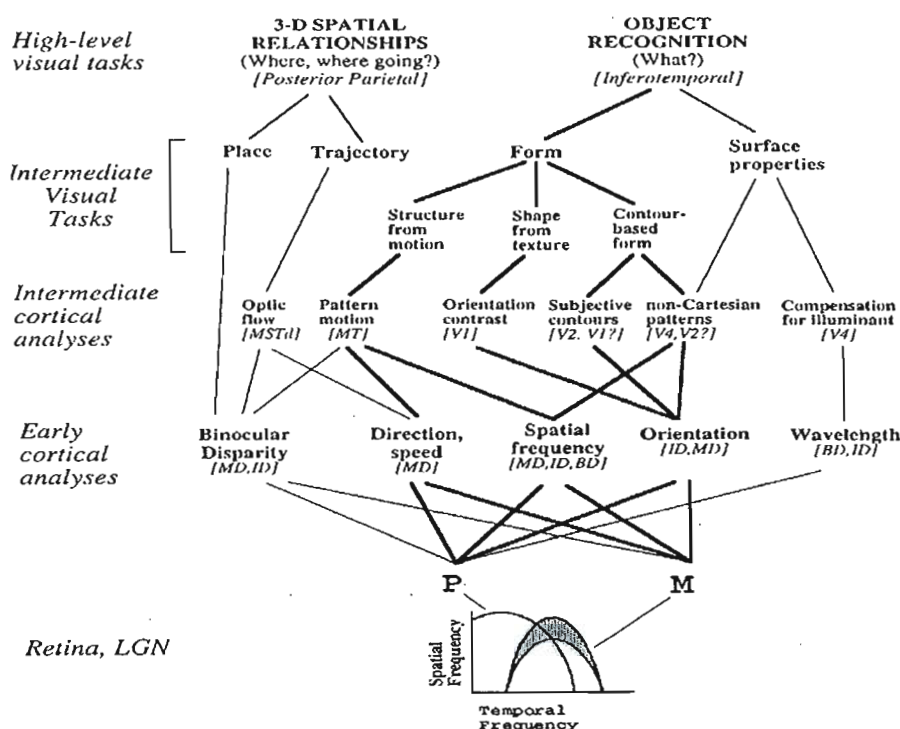


Fig.I.2 : Modularité et hiérarchie dans le système de vision biologique, et relation avec les aires visuelles du cerveau humain. Pour détails, voir [Essen92].

Remarquons enfin que l'approche biologique recoupe l'approche mathématique au moins en deux points. Tout d'abord, le système de vision des vertébrées est lui aussi structuré de façon hiérarchique et modulaire (voir fig.I.2). Par ailleurs, certaines approches statistiques en vision artificielle *bas niveau* conduisent à des modèles tout à fait équivalents aux systèmes connexionnistes biologiques (réseaux de neurones d'un côté et régularisation bayésienne par champs de Markov de l'autre). C'est à ces modèles statistiques que seront consacrés les deux premiers chapitres de cette thèse.

Ce temps dépend uniquement du domaine d'application (1 ms pour l'optique adaptative, 10 ms pour le suivi de missiles, 100 ms pour l'aide à la conduite, etc.).

<sup>8</sup> Il est intéressant de remarquer que la rétine (de l'homme) possède plus d'une centaine de photorécepteurs (cônes et bâtonnets) pour une seule fibre nerveuse de sortie dans le nerf optique.

## 2. Régularisation Bayésienne

### 2.1 Introduction et notations

L'extraction de primitives de l'image en vision bas niveau est en général un problème *mal posé* au sens d'Hadamard [Berter88], i.e., pour une donnée unique du problème, il existe plusieurs solutions possibles compatibles avec l'observation. L'introduction d'un certain nombre de contraintes permet de restreindre l'ensemble des solutions. Cette technique de tri *a priori* porte le nom de *régularisation*. Par rapport à la régularisation standard, qui met définitivement de côté un sous-ensemble de solutions, la régularisation bayésienne est une approche *probabiliste* : le problème est modélisé de telle sorte qu'une distribution de probabilités pondérant chacune des solutions potentielles est construite sur l'ensemble des solutions possibles.

Les notations suivantes seront appliquées tout au long de la thèse. Leur interprétation concrète dépendra du problème traité :

Le support de l'image (grille de pixels ou sites) :

$S = \{0 \dots n-1\} \times \{0 \dots m-1\}$  ensemble de sites de l'image de taille  $n \times m$ .  $N = mxn$  est donc le nombre de pixels ou sites de l'image.  $q, r$  ou  $s \in S$  sera la notation usuelle pour un site particulier de l'image. Enfin,  $V(s)$  dénotera un ensemble de sites autour de  $s \in S$  appelé *voisinage* de  $s$ .

L'image (image "réelle" observée par le capteur) :

$I = \{I(s) \in G_I, s \in S\}$  est le champ *image* (considéré comme un champ aléatoire).  $G_I$  est un ensemble continu ou discret, fini ou infini de niveaux de gris de l'image réelle. Le champ (déterminé)  $i = \{i(s) \in G_I, s \in S\}$  représente une image particulière (réalisation du champ image). On notera  $\Omega_I$  l'ensemble des réalisations possibles du champ image :  $\Omega_I = G_I^{n \times m}$ .

L'observation (résultat de l'observation de l'image "réelle") :

$O = \{O(s) \in G_O, s \in S\}$  est le champ *d'observation* associé à l'image.  $G_O$  est un ensemble continu ou discret, fini ou infini des valeurs possibles pour l'observation. Le champ  $o = \{o(s) \in G_O, s \in S\}$  est une réalisation particulière du champ  $O$ . On notera  $\Omega_O = G_O^{n \times m}$  l'ensemble des réalisations possibles du champ d'observation.

Le champ de primitives ou étiquettes de l'image (choses "à voir") :

$E = \{E(s) \in G_E, s \in S\}$  est le champ *d'étiquettes* ou *primitives* associé à l'image.  $e = \{e(s) \in G_E, s \in S\}$  est une réalisation particulière du champ d'étiquettes.  $e(s)$  est l'étiquette associée au site  $s$  à valeurs dans  $G_E = \{a_1, a_2, \dots, a_p\}$ , ensemble *discret*

et fini des valeurs possibles pour les étiquettes. L'ensemble des réalisations possibles du champ d'étiquettes sera noté  $\Omega_E = G_E^{n \times m}$ .

On utilisera indifféremment les notations  $E(s)$  ou  $E_s$  pour caractériser le champ (d'étiquettes en l'occurrence) au site  $s$  de la grille  $S$ . Dans le cas de traitement spatio-temporels (détection du mouvement par exemple), les champs aléatoires précités doivent être remplacés par des *processus aléatoires* à dimension temporelle (cf. Chap. IV).

## 2.2 Règle de Bayes

La règle de Bayes est à la base du formalisme. Soit le champ d'observations  $O$  issu des données observées et  $E$  le champ d'étiquettes que l'on cherche à estimer. La distribution de probabilités sur l'ensemble  $\Omega_E$  de toutes les réalisations possibles de  $E$  connaissant  $O=o$  s'appelle *distribution a posteriori* du champ à estimer et s'obtient par la formule de Bayes :

$$\Pr(E = e / O = o) = \frac{\Pr(E = e) \cdot \Pr(O = o / E = e)}{\Pr(O = o)} \quad (\text{Eq.I.a})$$

L'application de la règle de Bayes implique d'une part (a) la possibilité de décrire statistiquement le processus d'observation par la distribution de probabilités  $\Pr(O/E)$ , et d'autre part (b) la connaissance de la distribution de probabilités *a priori*  $\Pr(E=e)$  du champ d'étiquettes. Si ces deux conditions sont remplies, on peut construire à partir de la distribution a posteriori plusieurs estimateurs du champ d'étiquettes recherché ; un *estimateur*  $\hat{e}$  est une application associant à toute réalisation du champ d'observation  $o$ , une configuration d'étiquettes estimée  $\hat{e}(o)$  :

$$\begin{aligned} \hat{e} : \Omega_o &\rightarrow \Omega_E \\ o &\mapsto \hat{e}(o) \end{aligned}$$

La *qualité* d'un estimateur est définie par rapport à une distance  $C$  sur l'ensemble  $\Omega_E$ , aussi appelé *fonction de coût*. On appelle *estimateur bayésien optimal* relativement à la fonction de coût  $C$ , un estimateur  $\hat{e}$  pour lequel l'espérance mathématique de la fonction de coût (ou *risque*) est minimale<sup>9</sup>. Puisque la distribution de probabilité du champ  $O$  est

<sup>9</sup> la moyenne porte sur toutes les réalisations *conjointes* possibles de  $E$  et de  $O$  (le *risque* ainsi défini est une mesure du coût moyen "d'erreur" de l'estimateur) :

$$\begin{aligned} \text{Risque}(\hat{e}) &= \langle C(\hat{e}(O), E) \rangle_{\Omega_o, \Omega_E} = \int_{o \in \Omega_o} \int_{e \in \Omega_E} C(\hat{e}(o), e) p_{E|O}(o, e) \mu(de) \mu(do) \\ &= \int_{o \in \Omega_o} p_o(o) \mu(do) \int_{e \in \Omega_E} C(\hat{e}(o), e) p_{E|O}(e|O=o) \mu(de) \end{aligned}$$

(on appelle *risque conditionnel* la valeur de l'intégrale portant sur  $\Omega_E$  dans l'expression ci-dessus).

positive, l'estimateur bayésien optimal doit minimiser le *risque conditionnel* pour chaque réalisation  $o \in O$  :

$$\forall o \in O, \hat{e}(o) = \arg \min_{\hat{e} \in \Omega_E} \int_{e \in \Omega_E} C(\hat{e}, e) \cdot p_{E|O}(e|O=o) \mu(de)$$

Le choix de l'estimateur dépend du problème considéré. Voici une liste des estimateurs le plus couramment utilisées en traitement des images [Marro85a] :

Nom	Fonction de coût	Expression résultante
Maximum A Posteriori (MAP)	$\forall (a, b) \in \Omega_E \times \Omega_E$ $C(a, b) = 1 - \delta(a, b)$	$\hat{e}^{MAP}(o) = \arg \max_{e \in \Omega_E} [\Pr(E = e   O = o)]$
Maximum des distributions a Posteriori Marginales (MPM)	$C(a, b) = \sum_{s \in S} (1 - \delta(a_s, b_s))$	$\hat{e}^{MPM}(o) = \{\hat{e}_s^{MPM}, s \in S\}$ , avec $\forall s \in S$ : $\hat{e}_s^{MPM}(o) = \arg \max_{g \in G_E} [\Pr(E_s = g   O = o)]$
Estimateur du Champ Moyen (MF)	$C(a, b) = \sum_{s \in S} (a_s - b_s)^2$	$\hat{e}^{MF}(o) = \{\hat{e}_s^{MF}(o), s \in S\}$ , avec $\forall s \in S$ : $\hat{e}_s^{MF}(o) = \langle E_s   O = o \rangle_{G_E}$ .

### 2.3 Lien statistique entre observation et étiquettes.

Le terme de probabilité conditionnelle des observations par rapport aux étiquettes  $\Pr(O=o/E=e)$  dépend de la modélisation du processus d'observation. Le champ observé est censé contenir toute l'information disponible sur l'image (ou la séquence d'images dans le cas de la détection du mouvement). Selon les cas, l'observation est soit directement accessible, soit elle résulte d'un pré-traitement plus ou moins complexe de l'image originale. Il est courant de modéliser le lien statistique entre observation et étiquettes par la relation [Geman84] :

$$\forall s \in S, O(s) = \Phi[F_s(e^*), B(s)], \quad (\text{Eq.I.b})$$

où  $s$  appartient à un sous-ensemble de sites de l'image  $s \in A \subseteq S$ . Dans la formule,  $e^*$  représente le champ d'étiquettes réel que l'on cherche à estimer, et  $O(s)$  est l'observation (considérée comme une variable aléatoire) au site  $s$ . Dans le cas de la restauration d'images,  $e^*$  n'est autre que l'image réelle  $i \in \Omega_I$  avant d'être corrompue par le bruit, et  $o$  est l'image dégradée récupérée par le système d'imagerie –voir exemple plus loin. Le fait que les observations soient définies pour un sous-ensemble  $A$  de l'image permet de modéliser le cas de données incomplètes. La fonction  $F_s$ , qui est typiquement non inversible représente par exemple la réponse impulsionnelle du système d'imagerie. La fonction  $\Phi$

modélise quant à elle la réponse du détecteur : il s'agit d'une fonction inversible qui couple localement aux observations un bruit de nature aléatoire, noté  $B(s)$ , avant d'appliquer une transformation locale sans effet mémoire. La fonction  $\Phi$  peut être linéaire (cas des caméras CCD) ou non linéaire (logarithmique pour l'œil ou la plaque photo). Le bruit peut être couplé aux mesures de manière additive ou multiplicative. Sa statistique conditionnera directement celle du terme de vraisemblance ; elle peut être, par exemple poissonnienne ou gaussienne, selon la nature du phénomène observé.

A partir de ce modèle, on peut remonter au terme de vraisemblance recherché  $\Pr(O/E)$ . On a tout d'abord  $B(s) = \Phi^{-1}[F_s(e^*), O(s)]$  ; si le bruit est décorrélié spatialement (i.e. si les variables  $B(s)$ ,  $s \in S$  représentent des variables aléatoires indépendantes) alors on peut écrire directement le terme de vraisemblance :

$$\Pr(O = o / E = e) = \prod_{s \in A} \Pr[B(s) = \Phi^{-1}(F_s(e), o(s))]$$

Par la suite on préférera la forme exponentielle :

$$\Pr(O = o / E = e) = \frac{1}{Z} \exp \left[ - \sum_{s \in S} \Psi_s(e, o(s)) \right],$$

où l'on a posé de façon formelle :

$$\Psi_s(e, o(s)) = -Z \cdot 1_A(s) \cdot \ln \Pr[B(s) = \Phi^{-1}(F_s(e), o(s))]$$

( $Z$  est une constante arbitraire et  $1_A$  est la fonction caractéristique de  $A \subseteq S$ ). Cette formulation s'avérera utile par la suite car elle fait apparaître la *fonction de coût ou énergie de contrainte aux données observées*  $U_{obs}(e, o) = \sum_{s \in S} \Psi_s(e, o(s))$  lors de

l'intégration dans la formule de Bayes (cf. §4.1) :

$$\Pr[O = o / E = e] = \frac{1}{Z} \exp[-U_{obs}(e, o)] \quad (\text{Eq.I.c})$$

**Exemple** : Restauration d'images dégradées sans prise en compte de bords. Pour fixer les idées, prenons l'exemple plutôt simple de la restauration d'images dégradées sans prise en compte des bords [Geman84]. Dans ce cas, le champ d'observation  $O$  s'identifie avec le champ de niveaux de gris de l'image issu d'un système d'acquisition (objectif et caméra CCD par exemple). Le champ d'étiquettes  $E$  à reconstruire est tout simplement l'image originale non dégradée.

La *fig.I.3* représente les diverses transformations de l'image originale pour aboutir au champ d'observation  $O$ . Le processus d'observation est bien modélisé par la formule *eq.I.b*. Dans le cas très simple d'un système d'imagerie sans flou ( $F=I$ ) et d'une réponse parfaitement linéaire du capteur, l'*eq.I.b* devient tout simplement :  $o = e + n$ . Si l'on suppose en outre que le bruit est décorrélié spatialement, et qu'il

est sur chaque site gaussien, centré et de variance  $\sigma^2$ , alors sur l'image entière le terme de vraisemblance s'écrit :

$$\Pr(O = o / E = e) = \prod_{s \in S} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(o(s) - e(s))^2}{2\sigma^2}\right).$$

La fonction d'énergie à minimiser est alors (à une constante additive près) :

$$U_{obs}(e, o) = \sum_{s \in S} \Psi(o, e_s) = \frac{1}{2\sigma^2} \sum_{s \in S} [e(s) - o(s)]^2$$

La fonction  $\Psi(o, e_s) = \frac{1}{2\sigma^2} (e(s) - o(s))^2$  peut être vue comme un potentiel de clique binaire liant le champ d'observation avec le champ d'étiquettes (voir remarque au §4.1).

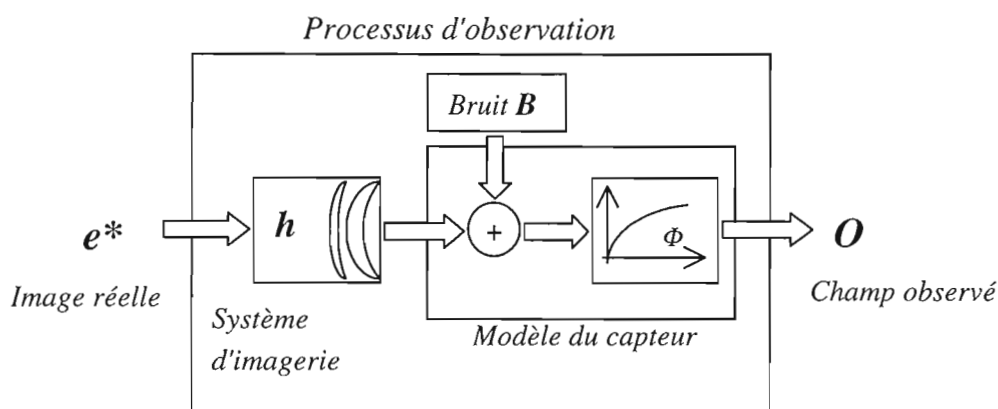


Fig.1.3 : Modèle du processus d'observation (acquisition d'images bruitées).

## 2.4 Probabilité *a priori* : modèle Markovien spatial.

Si aucune information supplémentaire *a priori* n'est introduite,  $\Pr(E=e)$  est une constante par rapport à  $e$ , et la distribution *a posteriori* issue de la formule de Bayes s'identifie avec le terme de vraisemblance  $\Pr(O/E)$ . Le critère du MAP devient alors tout simplement le critère du *Maximum de Vraisemblance* (MV)<sup>10</sup>. Puisque aucune contrainte n'est imposée quant à l'ensemble des solutions possibles du problème, la solution optimale correspond simplement à celle qui s'approche le mieux des données observées : il sera inutile par exemple de traiter par cette méthode la reconstruction d'images dégradées, car l'image observée est elle-même une solution acceptable du problème - et qui plus est, c'est la *meilleure* au sens du MV. Il faut donc introduire des contraintes *a priori* sur les solutions. Un *champ de Markov* en traitement d'images est un champ d'étiquettes à valeurs continues ou discrètes destiné à caractériser les informations sous-

<sup>10</sup> L'estimateur du maximum de vraisemblance maximise la probabilité conditionnelle  $\Pr(O/E)$ , considéré comme fonction de  $E=e$  (celle-ci est appelé alors *vraisemblance*).



jaçentes de l'image ou *primitives* (champ de vitesses dans le cas de l'estimation du mouvement, numéro de région dans le cas de la segmentation, etc.). La modélisation par champ de Markov rend compte des propriétés spatiales *a priori* (et/ou spatio-temporelles dans le cas d'une séquence d'images) des images observées. La propriété principale de la régularisation par champs de Markov est de conduire à des primitives *homogènes par zones* et automatiquement *débarrassées du bruit* [Geman84]. D'un point de vue formel, la modélisation par champ de Markov présente deux intérêts majeurs : (1) tout d'abord, il s'agit d'un modèle facilement paramétrable ; (2) comme on le verra au paragraphe suivant, à cette souplesse s'ajoute l'avantage d'une formulation immédiate du terme de probabilité conjointe du champ d'étiquettes  $Pr(E)$  grâce au théorème de *Hammersley-Clifford* qui relie propriétés locales (marginales) et probabilités globales du champ.

Ce qui suit est destiné à rappeler la terminologie et les fondements de la théorie des champs des Markov aléatoires ou MRF (pour *Markov Random Fields*). Pour plus de détails, voir [Azenco87].

### 3. Champs de Markov.

#### 3.1 Définitions.

Le but de cette partie est de rappeler les principales définitions et propriétés relatives aux MRF utiles dans le cadre de la régularisation bayésienne. La caractéristique principale des MRF est de fournir un modèle probabiliste *a priori* du champ des primitives à estimer.

**Définition.** Un champ aléatoire<sup>11</sup>  $E$  est un *champ de Markov* relativement à un voisinage  $V$  si et seulement si la propriété suivante est vérifiée :

$$Pr [E(s)=e(s) / E(r)=e(r), \forall r \in S, s \neq r] = Pr [E(s)=e(s) / E(r)=e(r), r \in V(s)].$$

Ainsi, les propriétés statistiques locales d'un champ de Markov au site  $s$  ne dépendent que de l'état d'un voisinage plus ou moins étendu défini autour de ce site et non de l'état du champ tout entier. En traitement d'images bas-niveau, on est en général amené à se contenter de considérer des voisinages d'ordre 1 (4 voisins), d'ordre 2 (8 voisins) ou d'ordre 3 (12 voisins) - voir *fig.I.4*.

---

<sup>11</sup> En toute rigueur, un champ aléatoire est défini par le triplet  $(\Omega_E, \mathcal{I}, P)$ , où l'espace des configurations du champ  $\Omega_E$  est muni d'une tribu  $\mathcal{I}$  et d'une mesure de probabilité  $P$  (mesure positive et de masse 1).

**Remarque** : Cette propriété de localité rend compte des propriétés intrinsèques de l'image, mais elle sera également mise à profit pour le calcul d'optimisation qui pourra être lui aussi local (et donc parallélisable) puisque toute l'information contextuelle à la prise de décision en un point est contenue dans un voisinage assez réduit (cf. Chap.II,§2.2.7).

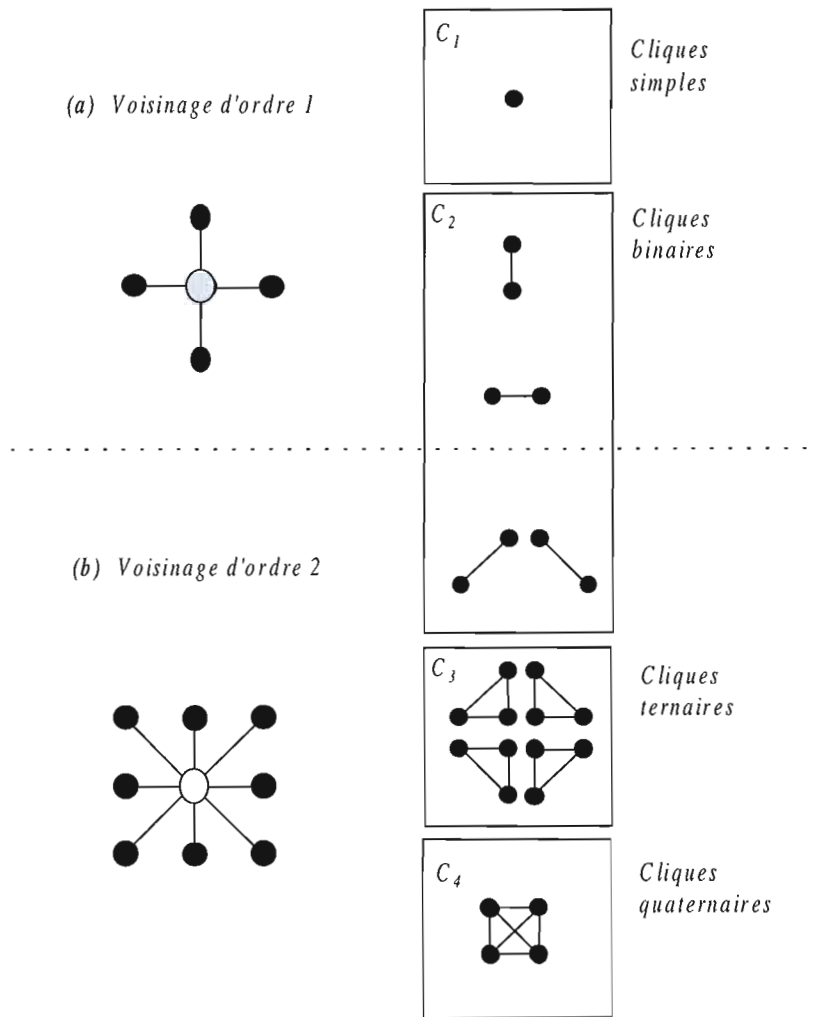


Fig.I.4 : Voisinages (a) d'ordre 1 et (b) d'ordre 2 et cliques correspondantes. Les sept dernières cliques sont exclusives du voisinage d'ordre 2.

### 3.2 Théorème d'équivalence

La définition du champ de Markov à partir des seules distributions marginales ne semble offrir aucune issue permettant d'aboutir à la distribution conjointe recherchée (ou terme d'*a priori*)  $Pr(E)$ . C'est pourquoi l'utilisation des champs de Markov en traitement

d'images n'a été effective qu'à la découverte du théorème de Hammersley-Clifford qui établit l'équivalence entre champ de Markov et distribution de Gibbs :

**Théorème de Hammersley-Clifford** (ou théorème d'équivalence). Le théorème de Hammersley-Clifford [Besag72] stipule que  $E$  est un champ de Markov si et seulement si la distribution de probabilités globale du champ  $\Pr[E=e]$  peut se mettre sous la forme *Gibbsienne* suivante :

$$\Pr[E=e] = \frac{1}{Z} \exp(-U_{reg}(e)) \quad (\text{Eq.I.d})$$

avec  $Z = \sum_{e \in \Omega_E} \exp(-U_{reg}(e))$ , et  $U_{reg}(e) = \sum_{c \in C} V_c(e)$ , où  $C$  désigne l'ensemble des cliques

définie sur la grille<sup>12</sup>  $S$  et  $V_c$  est une fonction de potentiel associé à la clique<sup>13</sup>  $c \in C$ .

Par la suite, la fonction  $U_{reg}$  sera indistinctement appelée *énergie spatiale*, *énergie a priori*, *énergie de régularisation* ou encore *de lissage spatial*. Elle ne dépend que des variables du champ à estimer et non du champ d'observation.

Un champ de Markov peut donc être défini en spécifiant les potentiels des cliques  $V_c$  ; les probabilités conditionnelles locales du champ s'écrivent alors :

$$\Pr[E(s) = e(s) / E(r) = e(r), \forall r \in S, r \neq s] = \frac{1}{Z_s} \exp[-U_s(e)],$$

où  $Z_s = \sum_{g \in G_E} \exp(-U_s(e)|_{e(s)=g})$  et  $U_s(e) = \sum_{c \in C: s \in c} V_c(e)$ .

La définition des potentiels élémentaires sera spécifique au problème à résoudre ; la force de l'approche markovienne réside justement dans le fait qu'il est beaucoup plus aisé de décrire et interpréter les interactions par le biais d'un potentiel que par l'écriture des distributions conditionnelles locales. De plus, la *seule* contrainte à respecter est que chaque potentiel  $V_c$  ne dépende que des pixels de la clique correspondante.

<sup>12</sup> Si l'on distingue la cardinalité des cliques, alors on peut poser :

$$U(e) = \sum_{c \in C_1} V_c(e) + \sum_{c \in C_2} V_c(e) + \sum_{c \in C_3} V_c(e) + \sum_{c \in C_4} V_c(e)$$

(dans le cas d'un champ de Markov défini sur un voisinage d'ordre 2).

<sup>13</sup> Par définition, l'ensemble de cliques associés à un voisinage donné correspond à l'union des sites singletons et de tous les sous-ensembles du voisinage composés de sites voisins deux-à-deux (voir *fig.I.4*).

Ainsi, le terme *a priori*  $\Pr(E)$  que l'on cherchait pour compléter la règle de Bayes est tout simplement explicitée par la distribution de Gibbs de l'éq.I.d ci-dessus.

**Exemple (suite)** : si l'on s'intéresse toujours à la restitution d'images bruitées et si l'on modélise l'image par un MRF défini par rapport à un voisinage d'ordre 1, alors il est classique de modéliser la contrainte d'homogénéité spatiale de l'image par un potentiel portant sur les cliques binaires de la forme :

$$V_c(e) = V_c[e(s), e(r)] = \beta_s \cdot (e(s) - e(r))^2, \text{ avec } \beta_s > 0$$

où  $c$  est la clique binaire comprenant les sites voisins  $s$  et  $r$ . Les configurations localement favorisées sont celles pour lesquelles les sites ont le même niveau de gris [Murray87], [Derin87], [Lalan90b]. La forme quadratique du potentiel spatial ressemble dans ce cas à celle du potentiel de contrainte  $\Psi$  trouvé lors de la modélisation du processus d'observation (dans le cas de variables binaires, le potentiel ainsi défini devient un potentiel à deux niveaux  $0, \beta_s$  - symétrisé en  $\pm\beta_s$  par addition d'une constante). L'énergie spatiale (ou terme de régularisation, ou encore de lissage spatial) s'écrit finalement :

$$U_{reg}(e) = \sum_{c \in C_2} V_c(e) = \frac{\beta_s}{2} \sum_{s \in S} \sum_{r \in V(s)} (e(s) - e(r))^2$$

La valeur du paramètre  $\beta_s$  dépend des caractéristiques spatiales de l'image réelle non bruitée<sup>14</sup>. Par exemple, le paramètre sera plutôt faible pour une image ayant une forte granularité ou des textures fines, et plutôt fort pour une image homogène. En fait  $\beta_s$  peut être estimé pour une image et un système de voisinage données, à condition de définir un critère d'erreur (de reconstruction) [Marro85a]. On peut également estimer la valeur de  $\beta_s$  sur un échantillon d'images "réelles" (voir remarque au §5.1).

### 3.3 Le problème de l'estimation des paramètres.

La formulation correcte d'un problème particulier en termes de fonction d'énergie à minimiser passe par le choix d'un certain nombre de paramètres définissant aussi bien les potentiels des cliques spatiales que le modèle d'observation. Dans la majorité des cas, ces paramètres sont fixés empiriquement ; ils peuvent être déterminés une fois pour toutes ou être réactualisés périodiquement en fonction des nouvelles images acquises.

Certains auteurs proposent encore des techniques *d'estimation* des paramètres du modèle. Azencott [Azenco88] propose par exemple une méthode de calibration consistant à fixer conjointement des configurations locales particulières du champ d'étiquettes *et* du champ d'observation, et d'imposer ensuite à chacune d'entre elle une probabilité conditionnelle locale. Ces probabilités sont proches de 0 ou 1 car les configurations sont volontairement exemplaires et donc extrêmes. Il en résulte un

---

<sup>14</sup> En toute généralité, c'est le choix du *système de voisinage du champ* tout entier (topologie, forme et valeurs des potentiels) qui en dépend ; on ne s'intéresse ici qu'à un voisinage d'ordre 1 ou 2 invariant par translation, et pour lequel le potentiel des cliques peut prendre uniquement deux valeurs  $\pm\beta_s$ .

ensemble d'équations liant les paramètres du modèles. Une illustration de cette méthode est développée dans [Letang93] ; nous avons testé la méthode dans un cas très simple (débruitage d'images binaires, cf.§5.1). Une autre technique consiste à estimer les paramètres conjointement à l'estimation bayésienne [Besag86], [Derin89].

## 4. Estimateur MAP et minimisation de l'énergie

### 4.1 Construction de la fonction de coût.

En remplaçant dans la formule de Bayes (éq.I.a) le terme de vraisemblance par son expression dans l'éq.I.c et le terme a priori par la formule de l'éq.I.d, on obtient finalement l'expression de la probabilité a posteriori :

$$\Pr(E = e / O = o) = \frac{1}{\Pr(O = o)} \cdot \frac{1}{Z} \cdot \exp \left[ - \left( U_{obs}(e, o) + U_{reg}(e) \right) \right]$$

L'introduction de la modélisation markovienne dans le cadre de l'estimation bayésienne permet, comme annoncé, d'avoir recours à des distributions gibbsiennes. Le critère du MAP, c'est à dire la maximisation de la probabilité a posteriori par rapport au champ des labels  $e$  étant donné un champ d'observations  $o$ , est alors équivalent à la minimisation d'une fonction de *coût* ou *énergie* totale comprenant deux termes :

$$U(e, o) = U_{obs}(e, o) + U_{reg}(e)$$

- Le premier terme est un terme de *rappel* aux données qui provient du modèle du processus d'observation ;
- Le deuxième terme est un terme de *régularisation* qui intègre les connaissances *a priori* dans le modèle (champ de Markov).

La fonction d'énergie est une fonction à grand nombre de variables (i.e. le champ d'étiquettes). La nature des interactions et du processus d'observation peut conduire à des fonctions qui ne sont *ni quadratiques ni convexes* par rapport aux variables ; l'énergie peut comporter alors plusieurs minimums locaux. Par contre, dans le cas d'une fonction quadratique et convexe, on est assuré de l'existence d'un minimum global. Ce minimum peut être obtenu par des méthodes d'optimisation déterministes classiques de type *descente de gradient* (plus forte pente à pas optimal, gradients conjugués, méthode de Newton).

**Remarque** : La distribution de probabilités *conjointe* du champ d'étiquettes et du champ d'observation  $\Pr[E=e, O=o]$  s'écrit :

$$\Pr(E = e, O = o) = \frac{1}{Z} \cdot \exp[-U(e, o)].$$

Or, la fonction d'énergie globale  $U(e, o)$  peut s'écrire comme somme de potentiels de cliques "généralisées"  $c \in C_X$  reliant le champ d'étiquettes et le champ d'observation :

$$U(e, o) = \sum_{c \in C_X} V_c(e, o) = \sum_{s \in S} V_s(e, o) + \sum_{c \in C} V_c(e),$$

où  $V_s(e, o) = \Psi_s(e, o(s))$  (voir §2.3). Autrement dit, et en vertu du théorème de Hammersley-Clifford, le champ *couplé*  $X = \{E, O\}$  est un champ de Markov.

**Exemple (suite)** : Si l'on remplace dans la formule de Bayes le terme de vraisemblance et le terme d'a priori par les expressions correspondant au problème de la restauration d'images, alors l'énergie à minimiser prend la forme finale :

$$U(e, o) = \frac{1}{2\sigma^2} \sum_{s \in S} (e(s) - o(s))^2 + \frac{\beta}{2} \sum_{s \in S} \sum_{k \in V(s)} (e(s) - e(k))^2,$$

que l'on a coutume d'exprimer de la façon suivante :

$$U(e, o) = \sum_{s \in S} (e(s) - o(s))^2 + \lambda \sum_{s \in S} \sum_{k \in V(s)} (e(s) - e(k))^2$$

Le facteur  $\lambda$  permet de pondérer le terme de rappel aux données par rapport au terme de régularisation [Blake87]. Remarquons enfin que  $U$  est quadratique *et* convexe par rapport aux variables du champ (dans le cas de niveaux de gris continus) : on est alors assurés qu'il existe un minimum global de l'énergie. Qui plus est, celui-ci peut être trouvé en un temps polynomial par des algorithmes *variationnels* déterministes de type descente de gradient.

## 4.2 Remarque : estimateurs MPM et MF.

L'estimation au sens du MPM ou du MF n'est pas immédiate<sup>15</sup> : pour le premier, il manque les distributions marginales *a posteriori*, qui ne sont pas directement accessibles en général. La difficulté du MF provient aussi du fait que son calcul passe par la connaissance de la fonction de partition de la distribution *a posteriori*. Cependant, en raison du caractère moins brutal de la fonction de coût employée<sup>16</sup>, l'estimateur MPM semble se comporter mieux que le MAP lorsque le processus de formation des observations introduit un bruit important [Prémon99]. Par ailleurs, les deux estimateurs pourraient être implantés facilement dans notre démonstrateur optoélectronique (cf. Chap.III). En effet, et comme on le verra au chapitre suivant, le noyau de l'algorithme d'optimisation stochastique utilisé pour résoudre la minimisation de la fonction d'énergie (dans le cadre de l'estimateur MAP) nécessite la mise en place d'*échantillonneurs*

<sup>15</sup> ...pour le moins en ce qui concerne la *formulation* du problème d'optimisation : en général, le calcul de l'estimateur MAP n'est *pas* immédiat non plus (voir chapitre suivant).

<sup>16</sup> La fonction de coût associé à l'estimateur MAP pénalise de la même façon toutes les configurations non optimales, alors que la fonction de coût du MPM pénalise chaque configuration *par rapport au nombre de pixels différents de la configuration optimale*.

Gibbsiens ; une stratégie pour obtenir les estimés du MPM ou du MF peut consister alors à *estimer* les propriétés statistiques *a posteriori* de l'image à partir d'un large échantillon d'images (loi des grands nombres) issu d'une procédure d'échantillonnage (cf. Chap. II, §2.2.2). Ainsi, l'estimé du MF est tout simplement approché par la moyenne de tous les échantillons, et l'estimé du MPM (en chaque site) est approché par l'étiquette apparaissant le plus fréquemment dans les échantillons. Nous n'avons pas exploré ces possibilités ; seul l'estimateur du MAP a retenu toute notre attention, la raison principale étant la formulation mathématique rigoureuse du problème comme un problème *d'optimisation*.

## 5. Quelques exemples

Nous allons introduire par la suite quelques problèmes en traitement d'images pour lesquels la fonction d'énergie n'est *ni quadratique ni convexe*. C'est uniquement à ce type de problèmes *difficiles* que nous allons nous intéresser dans le chapitre suivant.

Les problèmes présentés ici feront l'objet de quelques démonstrations sur le prototype optoélectronique (Chap. VI), c'est pourquoi il s'agit plutôt de cas d'école. Nous n'avons pas abordé ici des traitements de type spatio-temporels qui pourraient bénéficier de l'approche bayésienne et de la régularisation markovienne ; cela sera l'objet du Chap. IV - pour le cas particulier de la détection du mouvement- et donnera également lieu à une démonstration sur le prototype optoélectronique (Chap. V et VI).

### 5.1 Débruitage d'images binaires et modèle d'Ising.

#### 5.1.1 Définition de la fonction de coût ou énergie.

Nous avons déjà présenté le problème de la restauration d'images à niveaux de gris. Intéressons nous maintenant au cas de la restauration d'images *binaires*. Supposons par ailleurs que le champ observé est lui-même binaire. Alors, la modélisation la plus simple du processus d'observation fait intervenir non pas un bruit gaussien, mais un *bruit de canal* de taux d'erreur  $\varepsilon$ , défini selon :

$$\Pr[O(s)=1 / E(s)=0] = \Pr[O(s)=0 / E(s)=1] = \frac{\varepsilon}{2}, \text{ et}$$

$$\Pr[O(s)=0 / E(s)=0] = \Pr[O(s)=1 / E(s)=1] = \frac{1-\varepsilon}{2}.$$

En notant  $\beta_o = \ln\left(\frac{1-\varepsilon}{\varepsilon}\right)$ , on peut exprimer le terme de vraisemblance marginal par :

$$\Pr[O(s)=o(s) / E(s)=e(s)] = \frac{\varepsilon}{2} \exp[\beta_o \cdot \delta(o(s)-e(s))],$$

avec  $\delta(0)=1$  et 0 sinon. La *vraisemblance* de l'observation  $O=o$  est alors :

$$\Pr(O = o / E = e) = \left(\frac{\epsilon}{2}\right)^{n \times m} \exp\left(\beta_o \sum_{s \in S} \delta(o(s) - e(s))\right)$$

L'énergie de rappel à l'observation est enfin (à une constante additive près, ce qui ne change pas la solution finale résultat de la recherche du *minimum* énergétique) :

$$U_{obs}(e, o) = -\beta_o \sum_{s \in S} \delta(o(s) - e(s)), \text{ soit finalement :}$$

$$U_{obs}(e, o) = -\frac{\beta_o}{2} \sum_{s \in S} (2o(s) - 1) \cdot (2e(s) - 1).$$

Par ailleurs, si l'on choisit un potentiel de régularisation spatial à deux niveaux  $\pm\beta_s$  :

$$\forall (r, s) \in S \times S, V_C(e(s), e(r)) = -\beta_s \cdot (2e(s) - 1) \cdot (2e(r) - 1),$$

alors l'énergie de régularisation spatiale s'écrit :  $U_{reg}(e) = -\frac{\beta_s}{2} \sum_{s \in S} \sum_{r \in V(s)} (2e(s) - 1) \cdot (2e(r) - 1)$

L'énergie *totale* à minimiser pour la restauration d'images binaires bruitées est finalement :

$$U(e, o) = -\beta_s \sum_{s \in S} \sum_{r \in V(s)} (2e(s) - 1) \cdot (2e(r) - 1) - \beta_o \sum_{s \in S} (2o(s) - 1) \cdot (2e(s) - 1) \quad (\text{Eq.I.e})$$

On reconnaît dans le premier terme de la formule le *Hamiltonien* d'un système de spins en interaction mutuelle ou *modèle d'Ising bidimensionnel* [Landau67]. Dans ce cas précis les interactions sont toutes *positives* (le coefficient de couplage vaut  $J_{rs} = \beta_s$ ) ce qui correspond au cas des matériaux *ferromagnétiques*. La configuration d'énergie minimale pour le modèle *ferromagnétique* ou *anti-ferromagnétique* est connue d'avance : il s'agit d'un champ d'étiquettes uniforme (+1 ou 0) pour le cas ferromagnétique et d'un champ en damier (direct et inverse) dans le cas anti-ferromagnétique. Ceci est possible car il n'y a pas de phénomène de "frustration" (autrement dit, les configurations d'énergie minimale *locale* sont compatibles avec la configuration d'énergie minimale *globale*). Le phénomène de frustration apparaît par exemple quand toutes les coefficients de couplage n'ont pas le même signe (*Verre de Spin*), ou quand on couple le système ferromagnétique avec un champ magnétique externe *non uniforme*. L'image à restaurer  $O=o$  joue ici le rôle de ce champ magnétique externe non uniforme (deuxième terme de la formule). Expliciter alors la configuration du minimum énergétique est un tout autre problème - très difficile.



### 5.1.2 Choix des paramètres $\beta_S$ et $\beta_o$ .

Si l'on définit la "granularité" (d'ordre 1) d'une image comme étant le rapport du nombre de paires de pixels voisins ayant la même étiquette au nombre total de pixels de l'image, alors on constate que pour des images générées selon la loi d'*a priori* du modèle, la granularité décroît avec  $\beta_S$  en suivant une courbe qui rappelle celle de la magnétisation d'un matériau ferromagnétique en fonction de la température. En particulier il existe une valeur critique qui fait chuter brutalement la granularité de l'image en faisant apparaître de domaines uniformes. Ce n'est pas par hasard : la température critique correspondant à une *transition de phase* pour l'Ising ferromagnétique bidimensionnel (défini par l'Hamiltonien du premier terme de l'éq.I.e) vaut à peu près  $T_{cm}=v.\beta_S$  en utilisant l'approximation du *champ moyen* (ce qui implique en principe que le cardinal du voisinage soit  $v \gg 1$ ) ; dans le cas d'un voisinage 4, cette température peut être calculé exactement [Diu] : elle vaut  $T_c=0,567.T_{cm} \approx 2,27.\beta_S$ . Or, la loi d'*a priori* du modèle de Markov Pr(E) correspond à la distribution de Gibbs du modèle d'Ising pour  $\beta_S'=\beta_S/T$ . Il n'est donc pas étonnant que la granularité chute quand  $\beta_S'=\beta_S/T_c=1/2,27=0,44$ . Prendre  $\beta_S$  en *dessus* de cette valeur signifie que les images à modéliser présentent principalement des régions homogènes ; inversement, choisir un paramètre  $\beta_S$  en *dessous* de cette valeur signifie que les images à traiter présentent naturellement une granularité élevée. On prendra dans nos exemples (voir fig.I.4)  $\beta_S$  supérieur à 0,44.

A un facteur multiplicatif près (ce qui n'a pas de conséquence sur le *résultat* de l'optimisation), la fonction de coût ne dépend en réalité que d'un seul paramètre, le *rapport*  $\lambda=\beta_S/\beta_o$  :

$$U(e, o) = - \sum_{s \in S} (2o(s) - 1)(2e(s) - 1) - \lambda \sum_{s \in S} \sum_{r \in V(s)} (2e(s) - 1)(2e(r) - 1).$$

Le facteur  $\lambda$  traduit la *confiance* (ou plutôt la *méfiance*) que l'on peut faire au champ de données issu de l'observation : si le bruit de canal est nul -l'observation correspond exactement à l'image originelle-, alors le facteur  $\lambda$  est lui aussi *nul* ( $\beta_o \rightarrow +\infty$ ) et il n'y aura pas d'effet de régularisation : le minimum de l'énergie correspond tout simplement à l'image observée. Si par contre le bruit atteint 50% (le champ d'observation ne contient alors aucune information pertinente,  $\beta_o \rightarrow 0$ ), le terme de contrainte est négligeable, traduisant le fait que seule la connaissance *a priori* que l'on a sur les images est à prendre en considération (le résultat de l'optimisation correspond à une image entièrement uniforme).

**Remarque** : Si l'on choisit d'utiliser l'estimateur MPM, alors le résultat de l'optimisation en présence d'un bruit important correspond à une image aléatoire de granularité correspondant à celle de l'image que l'on aurait voulu reconstruire (à condition que le paramètre  $\beta_S$  ait été bien estimé au préalable). C'est pourquoi à

notre avis l'estimateur du MPM semble se comporter mieux en présence d'un bruit important.

Enfin, le facteur  $\lambda$  peut être estimé de façon plus ou moins heuristique (cf.§3.3). Nous avons testé deux façons de le faire :

1. En imposant des contraintes sur le résultat de l'optimisation (c'est une version simplifiée de la technique d'Azencott [Azenco88]) L'idée consiste à rendre improbables certaines configurations locales ; ainsi pour un voisinage d'ordre 2, on voudra éviter qu'un pixel "s'aligne" avec la donnée de l'observation locale, si plus de 4 pixels du voisinage lui "indiquent" le contraire. Ceci se traduit exactement par l'inéquation (à condition d'utiliser l'estimateur du MPM)<sup>17</sup> :

$$( +5.\beta_S - 3.\beta_S ) > \beta_o, \text{ c.à.d. : } \lambda > 1/2.$$

2. Plus simplement on a tenté d'ajuster "à l'œil" le paramètre  $\lambda$  pour un type spécifique d'images donnée. Ceci nous a conduit, pour les images binaires très simples qui serviront à tester le démonstrateur optoélectronique (i.e. lettres de l'alphabet latin bruitées), à des valeurs de  $\lambda$  comprises entre 0.4 et 0.6 pour un voisinage 8 et entre 0,9 et 1,2 pour un voisinage 4 (voir fig.I.4).

---

<sup>17</sup> Nous allons utiliser par la suite l'estimateur du MAP. La condition tient aussi dans ce cas, mais avec une signification légèrement différente : elle implique que pendant toute la durée du recuit (i.e. à toute température, cf Chap.II), la configuration aléatoire tirée respecte localement ces contraintes ; cependant, puisque la configuration finale minimise l'énergie globale, la contrainte locale sera certainement violée beaucoup plus souvent dans celle-ci.

Images binaires bruitées (40x40)



Bruit de canal :

5% 15% 25% 35% 45%

---

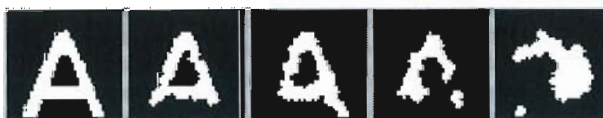
Filtrage passe-bas



Optimisation déterministe (ICM)



Optimisation Stochastique (SA)



Restauration ( $v=8$ )

---

Filtrage passe-bas



Optimisation déterministe (ICM)



Optimisation Stochastique (SA)



Restauration ( $v=4$ )

Fig.I.4 : Exemple de restauration d'images binaires bruitées. (Pour détails sur les méthodes d'optimisation, voir commentaires au Chap.II).

### 5.1.3 Restauration d'images avec prise en compte des bords.

Dans le cas de la restauration d'images à niveaux de gris continus (cas beaucoup plus intéressant en traitement d'images), la fonction d'énergie est quadratique par rapport aux variables du champ. Le minimum de la fonction d'énergie peut donc être obtenu par une méthode de type *descente de gradient*. On remarquera cependant que le modèle de débruitage proposé ne tient pas compte des discontinuités propres de l'image (le bruit est tenu pour responsable de tous les sauts brutaux entre niveau de gris). Ce n'est pas là simplement une remarque, c'est également un important défaut du modèle. Il est cependant facile de modifier le modèle pour tenir compte des bords à l'intérieur de l'image et en faire un modèle de restauration beaucoup plus réaliste. En effet, une technique maintenant classique ([Geman84], [Blake87]) consiste à introduire des informations *a priori* sur les discontinuités de l'image par l'adjonction d'un deuxième champ de Markov (binaire cette fois-ci) appelé *processus de ligne* et noté  $L$ .

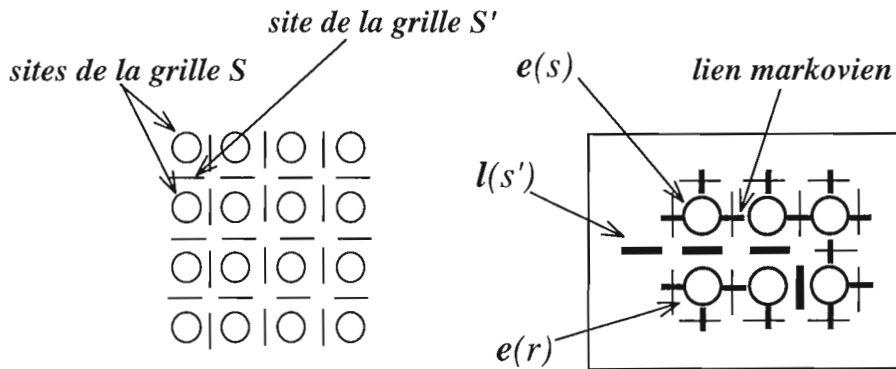


Fig.I.6 : Champ de Markov couplé pour la restauration avec prise en compte des discontinuités.

Les sites  $s'$  de la grille  $S'$ , support du champ  $L$ , s'intercalent entre les sites  $s$  du champ  $E$ . Le bord  $l(s') \in \{0,1\}$  agit comme un *interrupteur* coupant le "lien markovien" entre deux sites voisins du champ  $E$  (voir fig.I.6). Ce *champ de discontinuités ou bords*, forme avec le champ  $E$  de niveaux de gris un champ de Markov couplé  $(E,L)$  qui modélise l'ensemble des connaissances *a priori* sur l'image (niveaux de gris et bords). La fonction d'énergie à minimiser prend la forme suivante :

$$U(e, o, l) = \sum_{s \in S} (e(s) - o(s))^2 + \lambda \left\{ \sum_{s \in S} \sum_{r \in V(s)} (e(s) - e(r))^2 \cdot (1 - l_{sr}) + \Psi_B(l) \right\},$$

(dans l'expression ci-dessus,  $l_{sr}$  représente le bord du site  $s'$  entre le site  $s$  et  $r$ ). De la même façon que le champ  $E$  permet de prendre en compte les propriétés spatiales "agglutinantes" des niveaux de gris, le processus de ligne  $L$  peut prendre en compte les caractéristiques markoviennes spatiales des bords (par exemple, la probabilité d'avoir un bord entièrement entouré de bords est plutôt faible, la probabilité d'avoir un bord isolé

aussi, etc.). Ces propriétés ne sont pas modélisables par des potentiels simples comme pour le champ de niveaux de gris, et en général on se réfère à des tables ajustées de façon heuristique (fonction  $\Psi_B(l)$  dans l'expression ci-dessus). La recherche du minimum de la fonction d'énergie devient alors un problème *difficile*, dont la résolution n'est possible qu'en ayant recours à des algorithmes stochastiques sous-optimaux (cf. Chap.II). Prévost [Prévost95] étudie les principes et la faisabilité d'une rétine artificielle VLSI pouvant traiter le problème de la restauration d'images à niveaux de gris continus avec prise en compte des discontinuités (voir travaux précurseurs de Poggio et Koch [Poggio85]).

## 5.2 Codage en demi-teintes.

Le *codage en demi-teintes* ou *procédé de tramage* d'une image à niveaux de gris - *halftoning* en anglais - consiste à créer à partir de celle-ci une image binaire qui soit perçue par l'œil comme l'équivalent de l'image originale (fig.I.7). L'opération s'avère très utile dans le domaine de l'imprimerie (journaux, affiches et imprimantes de bureau), car il est plus facile de mettre au point des mécanismes d'impression en *tout ou rien* (imprimantes jet d'encre, à aiguilles, etc), non concernés en outre par des problèmes de réglage délicats du rendu des niveaux de gris. A notre connaissance, le procédé de codage en demi-teintes qui semble donner les meilleurs résultats est un problème d'optimisation difficile (voir plus bas), c'est pourquoi dans la pratique on utilise des variantes déterministes sous-optimales.



Fig.I.7 : A gauche, une image 476x476 pixels à 256 niveaux de gris (227Ko). A droite, image binarisée par la méthode de diffusion d'erreur (seulement 28 Ko car on n'a pas modifié la résolution). Même pour une résolution à l'impression qui reste assez faible (200ppi), la résolution tonale de l'image est acceptable.

Une argumentation du même genre peut être soutenue quant au problème de l'affichage dynamique : il est plus facile de réaliser des éléments optiques de commutation en tout ou rien que des pixels capables de reproduire avec précision une échelle d'intensité donnée (les matériaux électro-optiques sont par essence non-linéaires). De plus, par rapport à l'impression sur papier, la qualité de l'image sur écran se voit détériorée par la dispersion des caractéristiques entre pixels d'une même matrice. Il est

vraisemblable que des écrans binaires larges et/ou de haute résolution (ce qui améliore la qualité du rendu en demi-teintes) soient plus faciles à mettre au point que leurs homologues à niveaux de gris ; l'affichage d'images binarisées pourrait donc être une solution intéressante pour les applications ne nécessitant pas une restitution *exacte* de l'information binaire de l'image, mais seulement des qualités perceptibles à l'œil (tout écran d'affichage : affiches publicitaires, écrans d'ordinateur, et aussi télévision!, etc.). En résumé, le codage en demi-teintes en temps réel par recuit simulé pourrait (1) améliorer la *qualité* du rendu en demi-teintes des images larges dans le domaine de l'imprimerie (en particulier les affiches publicitaires) et (2) permettre l'utilisation d'écrans *binaires* de grande résolution et/ou taille pour l'affichage d'images à niveaux continus (le codage en demi-teintes en temps réel pourrait être à l'ordre du jour si le support multimédia à base *d'encre électronique* se généralise -système *gyricon* de Xerox [Gyricon] et l'*E-ink* de [MediaL99]).

Un autre avantage du codage en demi-teintes d'images est la possibilité de rendre celle-ci compatible avec un traitement numérique de type BIA (pour *Binary Image Algebra* [Huang89]) réalisable en technologie numérique sur des rétines artificielles VLSI (par exemple grâce à des opérateurs booléens invariants par translation [Zavido81], [Bernar92]). En toute généralité, le codage en demi-teintes est un cas particulier de *B-code* de l'image, représentation binaire pertinente de l'image relativement au système de conversion analogique-numérique donné (il est clair qu'un simple seuillage est un B-codage beaucoup plus pauvre). Ainsi, si l'on dispose à la fois d'un B-codeur efficace (rapide et de bonne qualité) et de dispositifs d'affichage binaires de grande résolution, la transmission (et le traitement) des images pourrait se faire sans avoir recours à un codage en niveaux de gris ; encore faudrait-il savoir si la bande passante nécessaire pour la transmission et le stockage d'images B-codées est plus grande ou plus petite que leur équivalent en niveaux de gris (bien évidemment l'image B-codée occupe moins d'espace mémoire si la résolution d'origine est conservée -voir *fig.I.7-* ; cependant en augmentant la résolution on améliore beaucoup la qualité visuelle du rendu tonal, mais la taille mémoire augmente...).

Venons enfin à la modélisation mathématique du problème. Le problème du codage en demi-teintes (ou B-codage *neuronal* [Bernar92]) peut être modélisé grâce à la fonction d'énergie à minimiser [Prémon99]:

$$U(e, o) = \sum_{s \in S} \{ [K^{-1} * o](s) - [K * e](s) \}^2 ,$$

où  $o$  est l'image à niveaux de gris,  $e$  est le champ d'étiquettes binaire à estimer (l'image B-codée),  $K^{-1}$  et  $K$  sont les noyaux de convolution permettant de réaliser respectivement le *rehaussement des bords* de l'image originale et le *flou* sur l'image binaire résultat (simulation de la réponse impulsionnelle de l'œil). L'énergie à minimiser est tout simplement la différence quadratique entre l'image binaire et l'image originale *telle*

qu'elles sont perçues par l'œil, le rehaussement des bords permettant d'éviter une perte de résolution spatiale. Il s'agit bien entendu d'un problème d'optimisation *difficile*.

## 6. Conclusion

Dans ce premier chapitre nous avons présenté une approche probabiliste générale permettant de **régulariser** les problèmes **mal posés** rencontrés en vision bas-niveau. Ceci est possible grâce à l'utilisation **d'estimateurs bayesiens**, dont par exemple l'estimateur du **maximum a posteriori** (MAP). Nous avons eu l'occasion de voir comment l'utilisation conjointe d'un **modèle de Markov** (MRF) chargé de représenter le champ de primitives, et d'une modèle probabiliste approprié du **processus d'observation**, permettait d'explicitier les termes de la formule de Bayes. Nous avons montré ensuite que l'estimation au sens du MAP est équivalente à la recherche de l'argument **minimisant une fonction de coût ou énergie** spécifique au problème, fonction que l'on a pu exprimer comme étant la somme de deux termes exerçant des influences "complémentaires" sur le résultat de l'optimisation (terme de contrainte aux données et terme de lissage spatial). Enfin, nous avons expliqué que la minimisation de la fonction d'énergie est, sauf dans quelques cas très particuliers, un problème d'optimisation **difficile** au sens algorithmique ; qui plus est, les seuls algorithmes (déterministes) que l'on puisse concevoir sont, comme on le verra au chapitre suivant, irrémédiablement sous-optimaux (le minimum global de l'énergie ne peut être trouvé que si l'on part d'une configuration déjà très proche de la solution).

L'objet du chapitre suivant sera donc de présenter la technique dite du *recuit simulé*, méthode d'optimisation *stochastique* permettant en principe de venir à bout de n'importe quel problème d'optimisation difficile - à condition toutefois d'avoir suffisamment de temps disponible.