

DETECTION DE CONTOURS ET LISSAGE D'IMAGE
 PAR DES ALGORITHMES DETERMINISTES DE
 RELAXATION.
 MISE EN OEUVRE SUR LA CM2.

Josiane Zerubia et Florimond Ployette

INRIA-Sophia - 2004 route des lucioles - 06560 - Valbonne.
 FRANCE
 et
 GDR TdSI no 134

Résumé

L'utilisation des champs de Markov en vision bas-niveau a fait l'objet de nombreuses recherches ces dernières années. Dans cet article, nous nous intéressons à l'implantation parallèle de deux algorithmes déterministes de détection de contours et de lissage d'image: le GNC (Graduated Non Convexity) et le recuit par champs moyens (MFA).

Nous donnons d'abord une brève description des deux méthodes de relaxation, puis nous présentons l'architecture de la CM2 ainsi que certains aspects de l'implantation des algorithmes. Enfin, nous comparons les deux algorithmes.

1 Introduction

L'utilisation des champs de Markov en vision artificielle fournit un modèle mathématique simple et général que l'on peut appliquer à plusieurs types de problèmes tels que la détection de contours, la restauration d'image, la reconstruction de surface, la stéréoscopie, la synthèse et la classification de texture.

Dans toutes ces applications, le problème qui se pose est de minimiser une fonction de coût qui est non-convexe à cause du champ processus de ligne. Un inconvénient majeur est le coût de calcul relatif à certains algorithmes obtenus à partir d'une telle modélisation. Les temps de calcul sont très fortement liés à la forme de l'énergie utilisée. C'est pourquoi nous avons choisi de paralléliser deux méthodes déterministes de relaxation: le GNC ("Graduated Non Convexity") [1] et le recuit par champs moyens [2]. Ces algorithmes itératifs ont été mis en oeuvre sur la machine à connexions CM2 de l'INRIA.

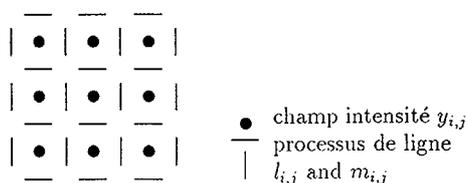


Figure 1: Modélisation de l'image

2 Les deux méthodes déterministes de relaxation

2.1 Le modèle mathématique

La modélisation de l'image par Champs de Markov permet de prendre en compte les interactions au niveau du pixel. L'image est représentée par deux champs: un champ intensité et un champ booléen appelé processus de ligne. Ce champ, introduit pour la première fois par Geman & Geman [3], représente explicitement la présence ou l'absence de discontinuité. Dans le modèle considéré, nous avons deux processus de ligne (horizontal et vertical).

Abstract

Recently, many researchers have investigated the use of Markov Random Fields (MRF) for early vision. In this paper, we focus on the parallel implementation of two deterministic algorithms for edge detection and image smoothing: the Graduated Non Convexity and the Mean Field Annealing. First, we briefly describe the two relaxation methods. Then, we review the architecture of the CM2 and we give some details about the implementation. Finally we compare both algorithms.

Nous supposons que la donnée d'entrée est une image bruitée décrite par:

$$d_{i,j} = y_{i,j} + n_{i,j} \quad (1)$$

où $n_{i,j}$ est un bruit blanc gaussien et $y_{i,j}$ la donnée originale.

L'énergie peut alors s'exprimer par:

$$E = \mathcal{D} + \mathcal{S} + \mathcal{P} \quad (2)$$

avec $\mathcal{D} = \sum_{i,j} (u_{i,j} - d_{i,j})^2$ qui assure une bonne adéquation entre la solution trouvée et les données.

$$\mathcal{S} = \sum_{i,j} \lambda^2 [(u_{i,j} - u_{i-1,j})^2 (1 - l_{i,j}) + (u_{i,j} - u_{i,j+1})^2 (1 - m_{i,j})]$$

qui représente la contrainte de lissage classique ($\lambda^2 * gradient^2$) en tenant compte du processus de ligne.

$\mathcal{P} = \sum_{i,j} \alpha (l_{i,j} + m_{i,j})$, où $\alpha (\alpha > 0)$ est le coût à payer pour l'introduction d'une ligne.

Le modèle décrit par l'équation (1) est celui d'une membrane à contraintes de continuité lâches [1], [2].

2.2 "Graduated Non Convexity"

Pour minimiser l'énergie décrite par l'équation (1) qui est une fonction non-convexe, Blake et Zisserman [1] ont proposé un algorithme déterministe de relaxation, appelé GNC ("Graduated Non Convexity").

Les techniques classiques de minimisation consistent soit en l'application d'une méthode déterministe (auquel cas le système peut être bloqué dans un minimum local quelconque), soit en l'utilisation d'une technique stochastique de recuit simulé (qui assure la convergence asymptotique vers le minimum global mais qui demande a priori un coût de calcul très élevé).



La méthode du GNC assure de trouver le minimum global dans des cas particuliers (par exemple si les discontinuités sont assez éloignées les unes des autres) et donne de bonnes solutions dans les autres cas (cf. Chapitre 7 de [1] pour plus de détails). Nous en rappelons brièvement le principe ci-dessous.

L'idée de base du GNC est la suivante: la première étape est la construction d'une approximation convexe E^* de l'énergie E ; la minimisation de E^* donne donc un minimum global. Puis une séquence d'énergies $E^{(p)}$ est construite telle que $E^1 = E^*$ et $E^0 = E$. Le GNC optimise toute la séquence d'énergies $E^{(p)}$ (par exemple, $p=1, 1/2, 1/4, 1/8...$ pour une image) en utilisant comme conditions initiales:

- les données, pour la minimisation de E^* ,
- la solution trouvée lors de la minimisation de la fonction énergie précédente dans les autres cas.

(cf.[9] pour plus le détail des équations)

2.3 Recuit par champs moyens

Une autre méthode déterministe de relaxation basée sur un recuit par champs moyens ("Mean Field Annealing") et appliquée à notre problème par Geiger et Girosi [2], Zerubia et Chellappa [8].

La technique d'approximation par champs moyens est souvent utilisée en physique statistique. En effet, l'obtention de façon analytique de la valeur moyenne d'un champ comme fonction explicite des données et des paramètres n'est en général pas possible. La physique statistique utilise alors comme outil l'approximation par champs moyens afin d'obtenir une solution approchée donnée de façon implicite par un système d'équations non linéaires.

Tous les calculs reposent sur la fonction de partition Z qui contient toute l'information du système physique considéré. Le calcul de la fonction de partition est équivalent à l'évaluation d'une intégrale multidimensionnelle qui ne peut pas être obtenue explicitement du fait de l'interaction entre les champs. L'approximation par champs moyens consiste à substituer l'interaction entre les champs de différents sites par l'interaction du champ de chaque site avec les valeurs moyennes des champs des sites voisins (voisinage défini suivant l'ordre du champ de Markov considéré). Alors la fonction de partition peut se factoriser en un produit de fonctions de partition relatives à un seul site et être calculées de façon implicite par un système d'équations non-linéaires obtenues grâce à la dérivée du logarithme de la fonction de partition approximée (cf.[9] pour la détail des équations).

3 La Machine à Connexions

Dans ce paragraphe, nous donnons une description succincte de cette machine dont on trouvera une documentation plus complète en [4], [6]. La machine à connexions est une machine SIMD ("Single Instruction Multiple Data") ayant de 8K à 64K processeurs. Chaque processeur est un processeur 1 bit disposant de 32K octets de mémoire locale et une horloge à 8 MHz. Les macro-instructions provenant du calculateur frontal sont traitées par un micro-contrôleur qui diffuse des nano-instructions à tous les processeurs.

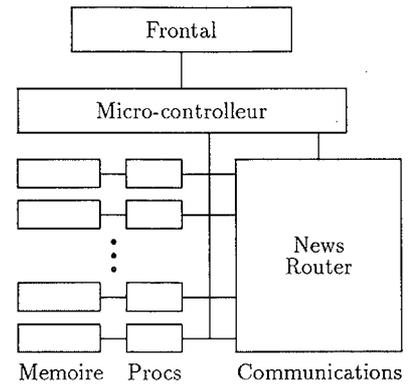


Figure 2: L'architecture de la CM-2

L'organisation physique de l'architecture est la suivante:

- Une puce regroupe 16 processeurs.
- Une section regroupe 2 puces de 16 processeurs, la mémoire locale de ces processeurs et une autre puce contenant l'unité de calcul flottant.
- Les puces processeurs (contenant chacune 16 processeurs) sont interconnectées en un hypercube à 12 dimensions (il y a de 1K à 4K puces processeurs). Chaque processeur a une adresse codée sur 16 bits et deux processeurs sont directement reliés si leur adresse diffère de 1 bit.

Pour chaque application, l'utilisateur peut dynamiquement définir une géométrie particulière pour l'ensemble des processeurs qu'il utilise. Cette géométrie décrit l'organisation logique de ces processeurs: le nombre de dimensions de la structure et le nombre de processeurs par dimension. Il peut, en outre, s'affranchir de la limite imposée par le nombre de processeurs physiques grâce au mécanisme des processeurs virtuels: auquel cas un processeur traite plusieurs éléments de la donnée.

La machine à connexions est donc une machine à parallélisme de données. Une même opération est effectuée en parallèle sur une grande masse de données. Ce type de parallélisme est bien adapté au traitement d'image bas niveau puisque les données (l'image) sont de taille importante et que les traitements sont locaux et identiques pour chaque pixel [5][7].

4 Mise en oeuvre des algorithmes de relaxation sur CM2

La géométrie adoptée en général pour le traitement d'images est une grille à deux dimensions dont la taille est égale à la puissance de deux immédiatement supérieure au nombre de pixels. On obtient donc un nombre de processeurs virtuels supérieur ou égal au nombre de pixels de l'image. Les opérations parallèles que l'on applique sur cette grille sont les extensions naturelles des opérations arithmétiques et logiques courantes des langages de programmation. Ces opérations sont effectuées en parallèle sur chacun des processeurs. Outre ces opérations parallèles, nous utilisons le réseau de communication locales NEWS (pour "North, East, West, South") qui est très rapide et particulièrement bien adapté à la propriété de localité des champs de Markov.

D'un point de vue mise en oeuvre, ces deux algorithmes sont assez semblables puisqu'il s'agit de deux algorithmes de relaxation itératifs. A l'issue de chaque itération, une nouvelle image lissée est calculée et est utilisée à l'itération suivante.

La valeur de la fonction énergie est calculée à chaque itération et sert dans le test d'arrêt de la boucle (à la $k_{i\grave{e}me}$ itération, arrêt si $\Delta E = E_k - E_{k-1} < \epsilon$). Le calcul de cette valeur nécessite une opération globale de réduction sur toute l'image (une somme globale).

Dans l'algorithme GNC, le calcul du gradient se fait en deux passes successives. On considère l'image comme un échiquier, le calcul du gradient est d'abord effectué sur les cases blanches de cet échiquier, puis, à la seconde passe, sur les cases noires. Ces deux passes sont séquentielles car, à la seconde passe on utilise le résultat du gradient calculé à la première passe. En effet, chaque processeur communique avec ses quatre voisins cardinaux. La minimisation de l'énergie se fait par une méthode de type SOR ("Successive Over-Relaxation").

Dans l'algorithme MFA on utilise une méthode de descente optimale par gradient conjugué qui nécessite un nombre plus élevé de calcul de l'énergie.

5 Performances des algorithmes

Nous avons testé ces algorithmes sur des images très variées fournies par l'ENST dans le cadre du Groupe de Recherche 134 Traitement de Signal et Images ainsi que sur des images provenant de USC et de l'INRIA [8][9].

5.1 Choix des paramètres

Pour la méthode du GNC, les deux paramètres à choisir sont le facteur de lissage λ qui est également un facteur d'échelle

(cf.[1]) et le seuil $h = \sqrt{\frac{2\alpha}{\lambda}}$ qui fixe la limite de détection des contours. Le test d'arrêt de l'algorithme est basé sur un seuillage de la variation d'énergie ΔE calculée après chaque itération.

Pour la méthode du recuit par champs moyens (MFA), le choix de la loi de variation de température est très important. En effet, lorsque les images sont peu contrastées ou fortement bruitées, une variation trop rapide de la température donne des images lissées de mauvaise qualité ainsi qu'une détection de contours moyenne alors que le test de convergence de l'algorithme (basé sur la variation d'énergie ΔE) est correct. Après avoir testé différentes lois de variation de température proposées dans la littérature, nous avons choisi une loi empirique qui donne de bons résultats sur toutes les images testées:

- $\beta = 0.0002$ à l'initialisation.
- $\beta = \beta * 4$ tant que $\beta \leq 1$.

Les paramètres à fixer sont le facteur de lissage λ et le coût α à payer pour créer une discontinuité (ce qui revient à choisir h).

5.2 Comparaison des temps de calcul

D'un point de vue temps de calcul, les deux algorithmes de relaxation du GNC et MFA sont sensiblement équivalents, avec toutefois un petit avantage pour le GNC qui comporte moins d'instructions par itération. La figure 3 montre que le temps d'exécution par itération est proportionnel au nombre de pixels par processeurs. Toutefois, quand le nombre de pixels par processeurs augmente, le rapport entre le temps

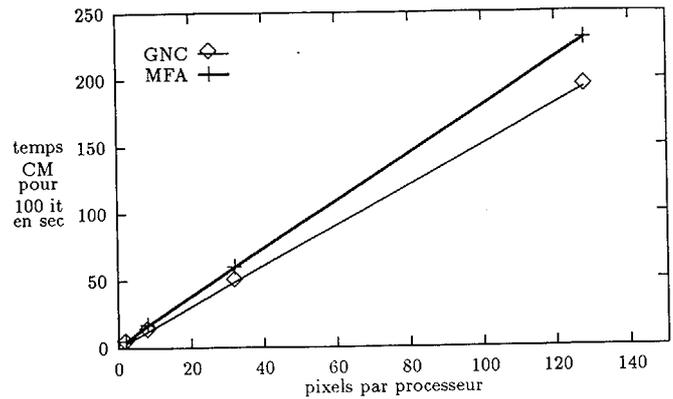


Figure 3: Temps d'exécution des algorithmes GNC and MFA

	VPR	Tps.CM	Tps.Total	Nb.Iter	Tps.CM par It
GNC	8	10.17s	11.22s	244	0.0416s
MFA	8	15.46s	18.22s	98	0.16s

Table 1 : Image infra-rouge 256x256.

d'exécution et le nombre de pixels par processeur décroît légèrement. Ce phénomène est dû au fait que, plus le VPR est élevé, plus fort est le pourcentage de communications locales.

A titre d'exemple la table 1 indique les résultats obtenus pour une image infra-rouge 256x256. Où le VPR est le nombre de pixels par processeurs et le temps CM le temps d'exécution des instructions CM.

5.3 Comparaison de la qualité des résultats

En ce qui concerne la qualité des résultats, on constate que le GNC et le recuit par champs moyens donnent des résultats analogues tant en détection de contours qu'en lissage d'image.

Dans le cas d'images non bruitées, les algorithmes déterministes de relaxation donnent des contours fermés mais présentent des inconvénients:

- pas d'affinage de contours,
- quelques problèmes de détection dans le cas de texture ou de parcellaire.

Dans le cas d'images bruitées ou de mauvaise qualité, ces algorithmes donnent des résultats acceptables en détection de contours même lorsqu'il y a un faible rapport signal sur bruit. De plus, il n'y a pas de coût supplémentaire à payer pour obtenir un image lissée.



6 Conclusion

Nous avons présenté les résultats obtenus en segmentation et lissage d'image à partir d'algorithmes basés sur une modélisation Markovienne de l'image. Les principaux avantages d'une telle approche sont:

- la prise en compte des interactions locales au niveau du pixel,
- l'utilisation des processus de ligne pour la détection des contours,
- l'incorporation de contraintes dans l'énergie,
- l'obtention d'une solution stochastique ou déterministe qui peut être implantée à l'aide d'un algorithme parallèle itératif.

Les résultats obtenus sont intéressants particulièrement dans le cas des images naturelles bruitées ou de mauvaise qualité. Les temps de calculs sont encore importants malgré l'implantation parallèle. Nous nous attachons à optimiser le nombre d'itérations des algorithmes sans altérer la qualité des résultats.

7 Bibliographie

1. A. Blake et A. Zisserman, "Visual reconstruction", *MIT Press, Cambridge - MA*, 1987.
2. D. Geiger et F. Girosi, "Parallel and deterministic algorithms for MRFs : surface reconstruction and integration", *Proc. ECCV90*, Antibes, Avr. 1990.
3. S.Geman et D.Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images", *IEEE Trans. Pattern Analysis, Machine Intel.*, Vol. PAMI-6, pp 721-741, Nov. 1984.
4. W.D. Hillis, "The Connection Machine ", *Cambridge, MA, MIT Press*, 1985.
5. J. Little, G. Belloch et T. Cass, " Algorithmic techniques for computer vision on a fine-grain parallel machine", *IEEE Trans. on PAMI*, Vol. 11, pp 244-257, Mar. 1989.
6. Thinking Machine Corporation, "Connection Machine, Model CM2 Technical Summary" *TMC, Cambridge, MA*, Mai 1989.
7. H. Voorhees, D. Fritzsche et L. Tucker, "Exploiting data parallelism in Vision on the Connection Machine system", *Proc. 10th ICPR*, Atlantic City, Juin. 1990.
8. J. Zerubia et R. Chellappa, "Mean field approximation using Compound Gauss-Markov Random Field for edge detection and image restoration", *Proc. ICASSP90*, Albuquerque, Avr. 1990.
9. J. Zerubia et F. Ployette, "Détection de contours et restauration d'image par des algorithmes déterministes de relaxation. Mise en oeuvre sur la machine à connexions CM2", *Rapport de recherche INRIA no.1291*, Oct.1990.
10. J. Zerubia et F. Ployette, "Détection de contours et lissage d'image par deux algorithmes déterministes de relaxation. Mise en oeuvre sur la machine à connexions CM2", à paraître dans la revue *Traitement du Signal*.



Figure 4: Image IRM 256x256

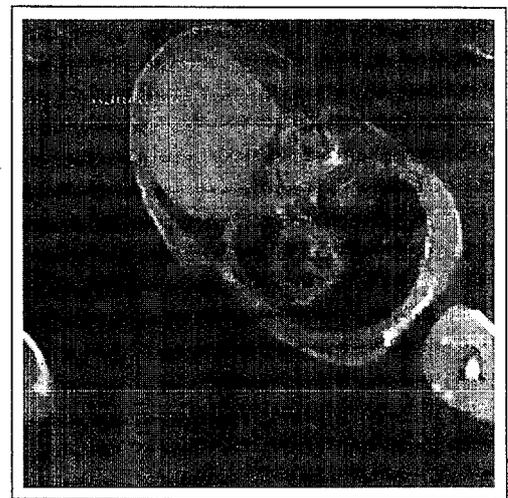


Figure 5: Image lissée par algorithme MFA



Figure 6: Image des contours par algorithme MFA