



NICE du 20 au 24 MAI 1985

VERS UN SEUILLAGE AUTOMATIQUE DE CLASSES D'IMAGES

TOWARD AN AUTOMATIC THRESHOLDING FOR CLUSTERS OF IMAGES

S. LELANDAIS (*)

B. BRILLAULT (**)

(*) 28, rue Vincent Scotto 44300 NANTES

(**) EDF Etudes et Recherches
6, quai Watier 78400 CHATOU

RESUME

SUMMARY

Dans de nombreuses applications la première étape d'un traitement d'image est le seuillage de l'image, une image binaire pouvant être ensuite caractérisée par de nombreux paramètres (nombre de particules, centres de gravité, diamètres, contours...) relativement aisément et dans des temps raisonnables.

Il peut être assez facile sur certains types d'image de trouver une valeur de seuil satisfaisante. Ce n'est pas toujours le cas. Certaines classes de document contiennent peu de parties homogènes (éclairage, conditions de prises de vues,...). L'outil proposé ici prétend présenter suffisamment de souplesse pour s'adapter à de nombreux cas.

Nous avons adapté un programme semi-automatique de seuillage d'images granulométriques complexes à partir d'un algorithme développé précédemment dans notre laboratoire et centré sur deux idées : séparation de l'image en un nombre variable de quadrants, seuillage par la méthode des distances interquantiles.

L'efficacité de l'algorithme initial a été grandement améliorée par l'adjonction d'une contrainte de cohérence entre les différents seuils des quadrants de l'image, par la souplesse de l'analyse effectuée sur l'histogramme et par l'ajustement automatique d'un paramètre particulier.

Le niveau de la décomposition en quadrants, le nombre de quantils, le niveau de contrainte de cohérence des seuils sont choisis par l'opérateur. Nous avons constaté que ces paramètres dépendaient principalement de la classe d'images à traiter. Pour automatiser ce programme dans le cadre d'une application donnée nous procédons en deux étapes : apprentissage interactif des paramètres optimaux pour la classe d'images à traiter, ces paramètres étant fixés, intégration du programme dans la chaîne de traitement.

L'objet de ce papier est la présentation théorique de cet algorithme implanté sur un HP 1000 couplé à un système de traitement d'images I2S, le modèle 70, en Fortran 4, et la description des résultats obtenus.

In many applications, the first step of a picture processing is thresholding because of the easy characterization of a binary image (with parameters as number of particles, diameters, edges...)

Certain types of image are easy to threshold, others aren't. Some classes of images have few homogeneous areas (light, context of acquisition).

The tool we present here intend to be flexible enough to be used in many problems.

We have developed a half automatical thresholding of complex granulometric images from an algorithm previously developed in our laboratory and based on two main ideas : to divide the image in some quadrants, thresholding using interquantile distance.

To improve the initial algorithm efficiency we add a property of coherence between the different quadrant thresholds, we consider the possibility to get none, one or two thresholds for an area (we choose one afterwards) and we used an automatical adjustment of a particular parameter. The level of picture decomposition, the quantil number, the coherence level of the thresholds are chosen by the operator. We have noted that these parameters depend mainly on the picture classes to be processed.

To get an automatical tool fitting a class of images we proceed in two steps : interactive learning of the parameters of the class, integration of the program in a set of processings.

The subject of this paper is the theoretical presentation of the algorithm and of the results we get, running the program, written in FORTRAN 4, on a HP 1000 with a "modele 70" as picture processing computer.



VERS UN SEUILLAGE AUTOMATIQUE DE CLASSES D'IMAGES

TOWARD AN AUTOMATIC THRESHOLDING FOR CLUSTERS OF IMAGES

I - INTRODUCTION

Dans de nombreuses applications en traitement d'images, on est conduit à utiliser des images binaires qui présentent de multiples avantages (1) :

- . Description compacte de l'image (1 bit par pixel).
- . Aucune signification du niveau de gris dans certains cas (reconnaissance de l'écriture, dessin,...).
- . Possibilité d'utiliser des opérateurs Booléens pour faire de nombreux calculs (vitesse), de nombreux traitements simples et efficaces.
- . Immunité au bruit dans la transmission.
- . Acquisition et visualisation peuvent être faites sur du matériel bas de gamme (peu coûteux).

Ces images binaires sont obtenues soit directement à partir d'un capteur approprié, soit à partir d'une image comportant N niveaux gris. Dans ce deuxième cas, il faut trouver une valeur de seuil conservant, dans l'image résultat, tous les détails pertinents pour l'application.

De nombreuses méthodes de seuillage existent (1,2,5), par exemple :

- . Décision Bayésienne (théorique ou expérimentale).
- . seuillage adaptatif.
- . analyse de la dérivée de l'histogramme des niveaux après avoir appliqué un algorithme de lissage sur celui-ci.
- . méthode paramétrique globale.

Certaines sont assez complexes à mettre en oeuvre ou nécessitent un temps de calcul important ou encore ne donnent des résultats satisfaisants que dans des cas précis. Le choix parmi ces méthodes est donc difficile et résulte le plus souvent d'un compromis entre simplicité, rapidité, qualité. Elles ont souvent des applications extrêmement restrictives. Par exemple BRU et LORETTE (3) se fondent sur une analyse très précise de l'histogramme des niveaux pour isoler les différents éléments d'un circuit intégré. FAVIER et coll. (4) partent de la connaissance de la valeur moyenne de l'image (M) et des écarts admissibles des niveaux (N) pour calculer un seuil $S=M+N$. LATROUSSE et BOUCHER (5) ont cherché à obtenir une procédure de seuillage automatique. Celle-ci suppose un histogramme des niveaux bimodal et optimise le choix du seuil en utilisant un calcul de variances inter-classe. ROSENFELD (6) propose une méthode d'amélioration de l'histogramme qui pourrait constituer un prétraitement de cette courbe et faciliter la détermination d'une valeur de seuil. Mais cet algorithme qui aurait pu permettre une approche plus générale du problème, donne des résultats assez peu encourageants.

Au laboratoire image de l'EDF, nous sommes confrontés à des images de genres très différents émanant de services chargés de surveillance, mesure, essais... Ces documents sont analysés et nous proposons, si possible, des moyens de traitement pouvant être implantés sur les sites d'acquisition. Ces moyens répondent à deux conditions :

- . automatisation maximum.
- . rapidité.

Une des étapes d'un traitement d'image est très souvent une opération de seuillage. Cette opération réalisée manuellement est fastidieuse et pas toujours efficace, nous avons cherché un programme assez souple pour pouvoir être adapté rapidement à des cas variés. Nous allons, dans les paragraphes suivants, détailler l'algorithme utilisé, les méthodes choisies pour améliorer les résultats et rendre le programme reconfigurable, et enfin présenter les résultats finaux.

II - ALGORITHME UTILISE

Nous avons repris une méthode qui avait été développée dans notre laboratoire par Georges MEGHABGHAB (7). Son but était de faire de la granulométrie automatique sur des images relativement complexes. La première étape du traitement consistait en une binarisation de l'image. La méthode utilisée est basée sur deux idées :

- . séparation de l'image en sous-images.
- . utilisation de la distance inter-quantile.

Les problèmes dus à la complexité des images ont poussé MEGHABGHAB à s'éloigner de la méthode de départ pour obtenir des résultats satisfaisants. Nous avons donc réécrit l'algorithme en reprenant l'esprit des deux principes ci-dessus.

1°) Séparation de l'image en sous-images

C'est une idée déjà exploitée (8) et qui permet souvent d'obtenir de bons résultats.

La répartition spatiale des niveaux de gris dans une image est rarement homogène. Ceci est dû à la variété des sujets présents dans le document, aux caractéristiques du capteur réalisant la prise de vue et aux conditions d'éclairage de la scène. Trouver une seule valeur de seuil pour une image entière est illusoire dans un certain nombre de cas.

On divise donc l'image en 2^n sous-images, ou quadrants et on calcule une valeur de seuil pour chacun de ces quadrants. Le seuil trouvé prend donc en compte les conditions locales de la photographie (taille des objets, zones floues, variation de la luminosité, ...).

2°) Méthode de la distance interquantile (9)

Si une image comporte deux régions homogènes bien distinctes : la forme et le fond, l'histogramme des niveaux de gris est composé de deux pics correspondant à chacune des régions. La valeur optimum de seuillage est située dans la vallée séparant ces pics.

En analysant l'histogramme cumulé $H(n)$ on cherche à isoler les zones fortement peuplées de l'histogramme $H(n)$, correspondant aux pics dans le cas simple évoqué ci-dessus (figure 1). Pour cela on divise $H(n)$ en un nombre donné "p" de quantiles. On obtient C_p classes contenant chacune une population donnée Q_i des points de l'image.

$$Q_i = \frac{\sum_{n=0}^{255} H_N(n)}{P}$$



VERS UN SEUILLAGE AUTOMATIQUE DE CLASSES D'IMAGES
TOWARD AN AUTOMATIC THRESHOLDING FOR CLUSTERS OF IMAGES

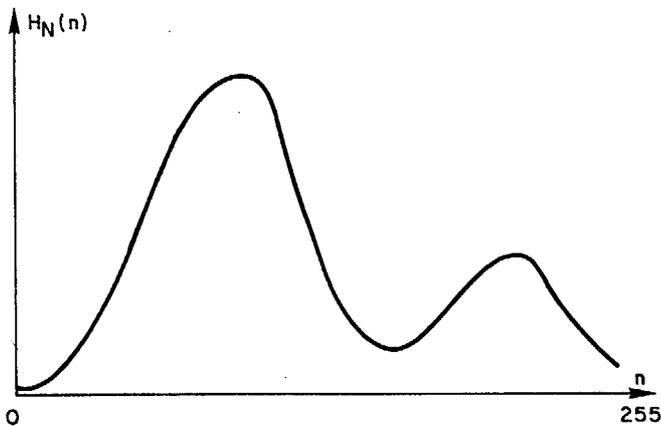


Figure 1 : Histogramme par niveaux

Aux P quantiles correspondent P-1 valeurs de niveaux de gris, représentant les frontières entre classes. Le quantile Q_1 s'étendra du niveau 0 au niveau α_1 , puis Q_2 de α_1 à α_2 et ainsi de suite (figure 2).

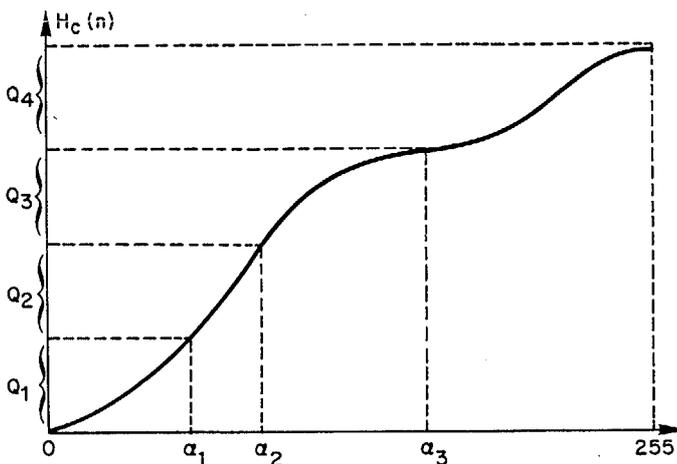


Figure 2 : Histogramme cumulé divisé en 4 quantiles

On constate sur cette figure 2 que lorsque la pente de $H(n)$ est raide on se trouve dans des classes fortement peuplées de $H_N(n)$.

Le but de la méthode est d'agréger les quantiles voisins dont la distance est inférieure à un seuil δ_S que l'on appelle indice de dissimilarité.

On dira que deux quantiles consécutifs peuvent être groupés ensemble ou similaires au sens de δ , si :

$$\delta(Q_i, Q_{i+1}) = \alpha_{i+1} - \alpha_i = \Delta\alpha_i < \delta_S$$

On sélectionne ainsi les portions de plus grande pente de l'histogramme cumulé, c'est à dire les pics de l'histogramme.

Une fois que l'on a agrégé les classes voisines, on cherche une valeur de seuil de l'image, soit le niveau de gris correspondant à la mi-distance entre deux "super" classes (classes constituées de l'agrégation de plusieurs quantiles).

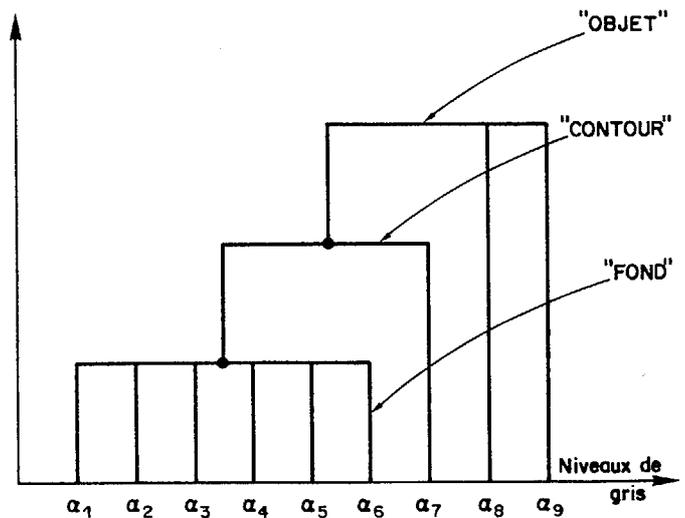


Figure 3

Constitution d'un nombre réduit de classes à partir de 10 quantiles et recherche du seuil.

Considérons la figure 3, les quantiles ont été agrégés de la manière suivante :

- Une "super classe" allant de α_1 à α_6 correspondant à une région de l'image, le fond par exemple.
- Un quantil restant isolé allant de α_6 à α_7 représentant le contour de l'objet.
- Puis, une "super classe" allant de α_7 à α_9 et qui représente l'objet.

Dans un tel cas, certes idéal, le seuil de l'image serait située entre la classe "fond" et la classe "contour de l'objet" :

$$s = \frac{\alpha_6 + \alpha_7}{2}$$

III - AMELIORATION DE LA METHODE

Une fois réécrit l'algorithme précédent nous avons travaillé dans trois directions :

1°) Amélioration des résultats par l'utilisation du contexte

Les premiers résultats obtenus étaient dans de rares cas satisfaisants. Dans certains quadrants les seuils trouvés conduisaient à un résultat en contradiction avec le reste de l'image, ceci pouvant être dû à la présence d'un autre objet, ou à un défaut de l'image ou bien à l'absence complète d'objet dans cette zone.

Une image présente souvent des variations spatiales du niveau gris moyen. Malgré tout, ces variations restent relativement faibles d'un quadrant à l'autre.

Nous avons considéré chaque quadrant comme une image à part entière, donc isolée de son contexte ; nous avons réintroduit cette notion en limitant les écarts possibles entre les seuils. Pour cela on calcule la moyenne de tous les seuils trouvés et on fixe une valeur maximum d'écart à cette moyenne.

2°) Utilisation de deux seuils - Cas de l'absence de seuil

Pour résoudre des cas plus complexes (présence locale



VERS UN SEUILLAGE AUTOMATIQUE DE CLASSES D'IMAGES

TOWARD AN AUTOMATIC THRESHOLDING FOR CLUSTERS OF IMAGES

d'un objet "intrus", tâche...) nous avons introduit la possibilité de trouver deux seuils sur un seul quadrant. Ces deux seuils correspondent à la présence de deux "super-classes" équivalentes.

Le seuil retenu finalement sera le plus proche de la valeur moyenne évoquée précédemment. Celle-ci est alors recalculée en tenant compte de ce choix et la contrainte "d'écart à la moyenne" est alors appliquée.

Il est également possible de ne pas trouver de seuil du tout (absence d'objet) dans le cas d'une superclasse unique s'étendant de 0 à 255. Ce quadrant aura pour seuil la valeur moyenne décrite ci-dessus.

3°) Adaptativité de l'indice de dissimilarité

Une valeur d'indice de dissimilarité, permettant d'agréger les classes entre elles, est fixée a priori puis adaptée de manière automatique dans la suite de l'algorithme :

- . Si les différents quantiles ont été correctement agrégés c'est à dire si on a obtenu de une à trois "super-classes" (0, 1 ou 2 seuils), on essaie un indice de dissimilarité supérieur pour tenter d'agréger les classes supplémentaires voisines (élargissement des pics de l'histogramme), et affiner ainsi le résultat.
- . Si l'agrégation ne peut pas se faire correctement (toutes les distances inter-quantiles supérieures ou inférieures à l'indice proposé), on classe par ordre croissant toutes les distances et on prend comme nouvel indice l'élément situé au tiers de cette liste. On itère jusqu'à obtention d'une à trois "super-classes" (cf. cas précédent).

4°) Discussion

Toutes ces modifications, de mise en oeuvre aisée, peu coûteuses en temps de calcul et en place mémoire, nous ont permis d'obtenir des résultats très satisfaisants (cf. paragraphe 7). Elles ne supposent aucune connaissance a priori de l'image à traiter ou de son histogramme. Elles permettent d'adapter la méthode à un grand nombre de cas.

IV) VERS UN PROGRAMME AUTOMATIQUE

Ce programme est actuellement conçu comme un outil de laboratoire. Il nous permet de rechercher rapidement des valeurs de seuillage pour des images variées et éventuellement complexes.

L'opérateur doit fournir trois paramètres pour permettre au programme de s'exécuter :

- . Nombre de quadrants à traiter dans l'image.
- . Nombre de quantiles à extraire dans $H(n)$.
- . Ecart maximum autorisé pour le seuil d'un quadrant par rapport à la moyenne des seuils.

En testant ce programme sur différentes images, nous avons constaté que ces trois paramètres ne changeaient généralement pas pour des documents similaires et qu'il était souvent possible de trouver un optimum pour chacune des valeurs dans un cas donné. Par exemple dans le cas des images granulométriques (paragraphe 7), une décomposition en 4 quadrants avec séparation de l'histogramme cumulé en vingt quantils et un écart maximum à la moyenne de 25 donnent de très bons résultats.

On envisage donc, après une phase d'apprentissage des paramètres permettant de trouver ces optima, de rendre le programme complètement autonome face à une application précise.

Parmi les images traitées, les images radiographiques de soudure se sont avérées plus complexes à seuiller. Nous avons dû procéder à un prétraitement de l'image, en utilisant des transformations issues de la Morphologie Mathématique (10) avant de seuiller l'image. Nous avons obtenu des résultats très prometteurs.

V - CONCLUSION

Le programme sous sa forme actuelle est présenté sur l'organigramme de la figure 4. Il est écrit en FORTRAN 4 et implanté sur un HP 1000 couplé à un système de traitement d'images I2S, le modèle 70. Il n'est pas portable sous cette forme, comportant des appels à des fonctions du modèle 70. Un seuillage est effectué en une quinzaine de seconde (temps qui pourrait être nettement amélioré sur un matériel adapté).

Nous avons vanté sa souplesse et son efficacité. Ce n'est cependant pas une solution universelle et si elle permet de résoudre certains cas difficiles, la recherche des paramètres optimaux n'est pas toujours aisée.

Les quelques idées très simples qui nous ont permis d'adapter la méthode initiale sont issues d'une approche descriptive des images à traiter. Le succès de cette démarche justifie l'interprétation des résultats obtenus présentée dans le paragraphe suivant.

VII - RESULTATS

Le programme a été testé sur des images de types variés.

1. Images granulométriques

L'image de la figure 5 représente des grains dans un matériau.

L'image de la figure 6 représente des gouttelettes d'aéroréfrigérants.

Toutes les images traitées étaient d'un type voisin : tâches de taille petite à moyenne sur fond uniforme réparties sur toute l'image. Elles ont toutes été seuillées en divisant l'image en 4 quadrants, l'histogramme en 20 quantils et en imposant un écart maximum à la moyenne de 25.

2. Les images suivantes sont beaucoup plus difficiles (figure 7) : il s'agit de radiographies de soudure. Les exemples présentés ne sont pas exactement de même type. Il s'agit de 2 parties différentes d'une soudure numérisée à deux grossissements différents, une par la caméra et l'autre par un microdensitomètre (les textures de l'image sont très différentes).



VERS UN SEUILLAGE AUTOMATIQUE DE CLASSES D'IMAGES

TOWARD AN AUTOMATIC THRESHOLDING FOR CLUSTERS OF IMAGES

La soudure, fréquemment inhomogène, comporte éventuellement des défauts dont la taille, l'intensité et la définition sont très variables.

Ce sont ces défauts que nous désirons isoler. Isoler 1,2 ou 3 objets, souvent de petite taille, n'est possible que si l'image est seuillée localement et si ce seuil peut être choisi avec une certaine liberté. Effectivement des seuillages à peu près satisfaisants ont été obtenus par des degrés de décomposition de l'image élevés et pour un écart maximum par rapport à la moyenne égal à 50 : dans le premier exemple le défaut apparaît bien ; dans le deuxième exemple le défaut central est correctement seuillé, par contre, les défauts situés plus haut ne sont pas séparés du fond localement plus sombre. Un pré-traitement serait nécessaire.

La même soudure a été numérisée au micro-densitomètre à un grossissement plus faible. La soudure devient l'objet de l'image alors qu'elle en était le fond, elle est privilégiée par rapport aux petits défauts lors du seuillage, par conséquent inefficace. Après pré-traitement ("gommage" de la soudure) l'image a été de nouveau seuillée. La plupart des défauts apparaissent.

3. Nous avons enfin seuillé des images de scènes. Différents paramètres permettent un bon seuillage, le degré de détail ressortant étant différent pour chaque triplet. Un niveau de décomposition de l'image supérieur permet d'atteindre des objets plus petits cependant cela exige un écart maximum à la moyenne plus faible pour préserver la cohérence des gros objets et suppose donc une certaine homogénéité d'éclairage de l'image globale.

BIBLIOGRAPHIE

- (1) H.MAITRE Elements of picture processing Polycopie de cours ENST
- (2) WESZKA A survey of threshold selection techniques. CGIP7 1978
- (3) T.BRU, G.LORETTE Robot de soudure automatique des circuits intégrés et hybrides. Premier colloque Image GRETSI-CESTA Biarritz, 21-25 Mai 1984, pp 857-861
- (4) ch.FAVIER, G.THOMAS, ch.BREBANT, R.MOGAVERO Système d'analyse automatique de défauts en imagerie X. Premier colloque Image GRETSI-CESTA Biarritz, 21-25 Mai 1984, pp 881-884
- (5) S.LATROUS, J.BOUCHER Comptage du zooplancton par analyse d'image. Automatisation de l'extraction des paramètres morphologiques du zooplancton. Premier colloque Image GRETSI-CESTA Biarritz 21-25 Mai 1984, pp 1023-1029
- (6) N.AHUYA, A.ROSENFELD A Note on the use of second order Gray-Level statistics for threshold selection. IEE Transactions on system, man and cybernetics vol. SMC-8, n° 12, décembre 1978.
- (7) G.MEGHABGHAB Granulométrie par traitement d'images thèse de 3^e cycle. Université de PARIS VI 1er Janvier 1983.
- (8) V. DUTRON Traitement numérique de radiographies industrielles. 5^{ème} conférence internationale sur les méthodes de contrôle non destructif BORDEAUX 30 Mai-3 Juin 1983.
- (9) SCOLLAN, HUAN, WEIDUER Image Enhancement using the median and the interquantile distance. ICASSP 1983 BOSTON.
- (10) J.SERRA Image Analysis and mathematical morphology Academic Press 1982.

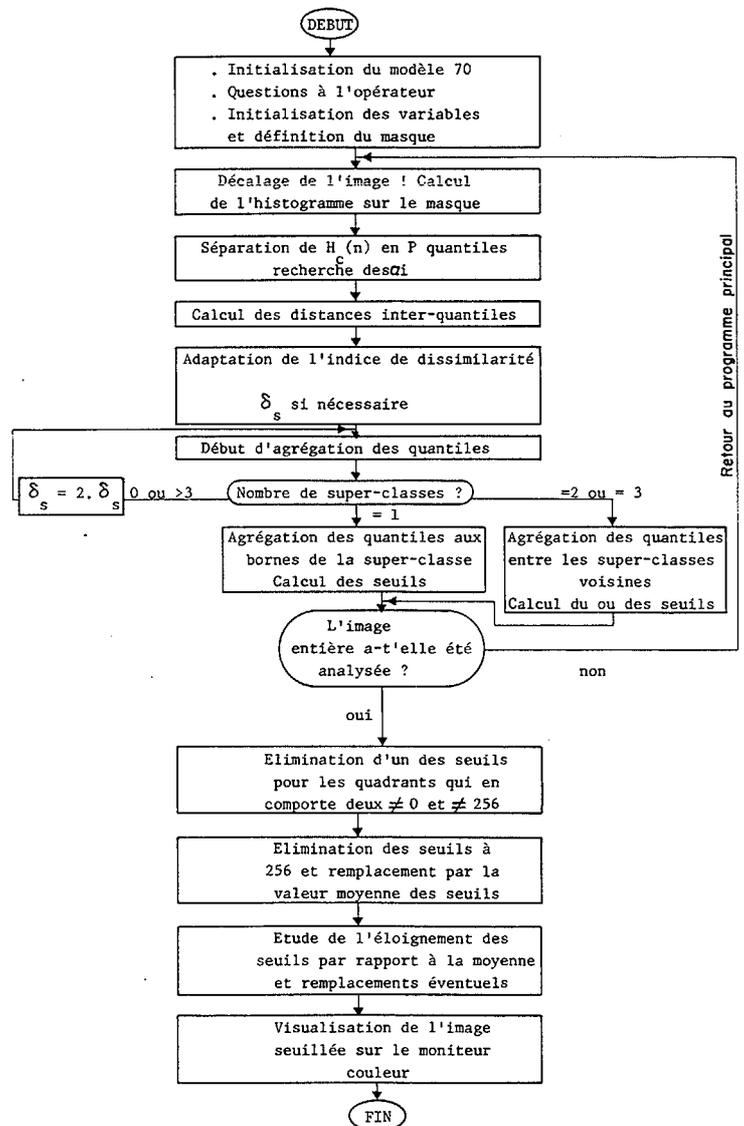


Figure 4 : Organigramme du programme complet

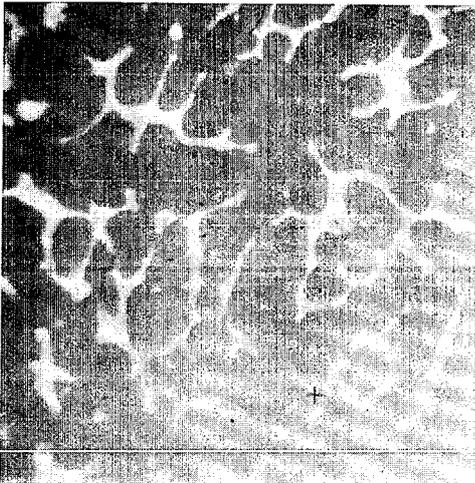


Figure 5

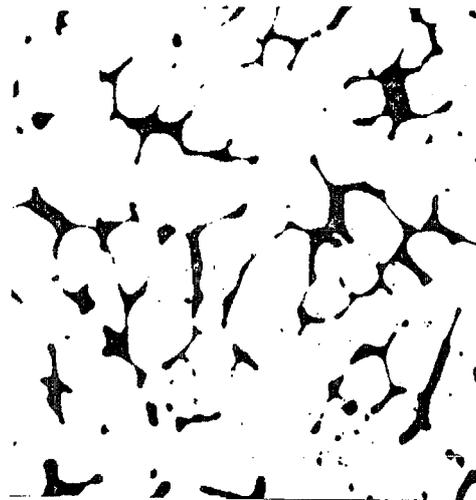


Figure 5

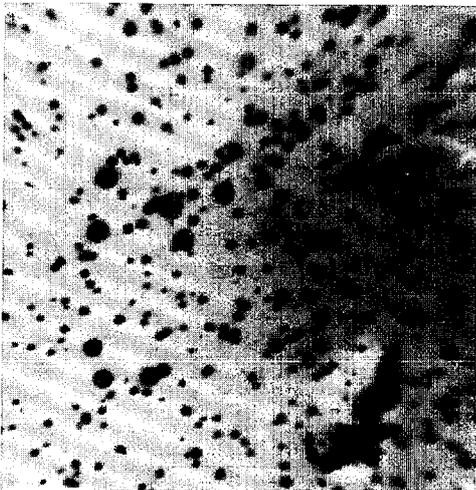


Figure 6



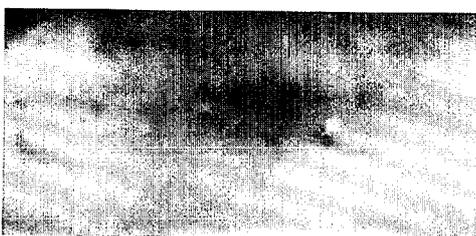
Figure 6



Figure 7



Figure 7



après
pré-traitement

