

HUITIEME COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET SES APPLICATIONS



NICE du 1^{er} au 5 JUIN 1981

METHODE ADAPTATIVE D'EXTRACTION DE CONTOURS A SEUIL CONSTANT

P. CHAMBON - G. GIRAUDON

SINTRA-ALCATEL , 26 RUE MALAKOFF 92600 ASNIERES

RESUME

Notre étude se situe dans le domaine du traitement numérique d'image et plus particulièrement dans celui de l'extraction de contours dans les images.

Une des méthodes de détection de contours les plus utilisées est le rehaussement d'image suivi d'un seuillage. C'est une méthode locale qui se fait par la mesure des discontinuités de niveaux de gris sur un voisinage centré, suivi d'une comparaison par rapport à un seuil. Les problèmes de mise en oeuvre d'une telle méthode sont d'une part la synthèse du filtre, et d'autre part le choix de la valeur du seuil.

Nous présentons dans cet article, deux nouvelles méthodes complémentaires, une de synthèse générale d'opérateurs différentiels linéaires et non récursifs adaptés aux types de contours, et l'autre permettant le calcul du seuil optimal en fonction du taux de fausses alarmes désiré.

Après avoir explicité la méthode de manière théorique, celle-ci est illustrée et comparée aux méthodes existantes, sur des images de scènes naturelles.

L'avantage et l'originalité de cette méthode par rapport à celles existantes est de permettre la sélection a priori du type de signal à détecter ou à rejeter comme faisant partie d'un contour. Elle opère donc de manière sélective sur les contours contenus dans une image, la détection se faisant à seuil constant quelque soit le rapport S/B sur l'image, à taux de fausses alarmes fixé.

SUMMARY

This paper is concerned with digital image processing and more specially with edge extraction in images.

One of the most useful methods is enhancement / thresholding edge detection. It is a local method which measures the gray levels discontinuities in centered neighbourhood, following by thresholding. The problems of this method are firstly the filter design, and secondly the threshold value choice.

In this paper, a new general design method of differential, linear and non recursive operators, matching to different typical edges is presented. And afterward, a new method, providing an optimal threshold value in fonction of desired false alarm rate, is designed.

Theoretical method and experimental comparaisons of this method are presented, by means of natural images.

The new method advantage is to allow to retain any kind of signal to be detected or to be rejected as edge.

It works, therefore, to selective manner over all edges of images, the threshold value being optimal for a desired false alarm rate.



INTRODUCTION.

L'extraction de contours est une méthode très utilisée dans le traitement d'image, pour des domaines aussi variés que la compression d'information, l'amélioration d'image, l'analyse de scène et la reconnaissance de forme. Son intérêt réside dans le fait que l'information sémantique contenue dans les contours d'un objet permet souvent de caractériser cet objet indépendamment de son environnement. Le système visuel humain semble, lui-même, utiliser largement la détection de contours pour sa perception scénique [1].

Mais malheureusement, il n'existe pas de définition formelle des contours applicable de manière universelle. La difficulté de généralisation tient au fait que le choix d'une méthode de détection dépendra grandement du type d'images à traiter, de la richesse informationnelle que l'on désire retirer et bien sûr des moyens que l'on peut mettre en oeuvre pour sa réalisation. Les contours que l'on peut observer sont de nature très différentes dans leurs formes [2]. En suivant le profil de niveau de gris d'une ligne, on peut obtenir par exemple (fig. 1) :

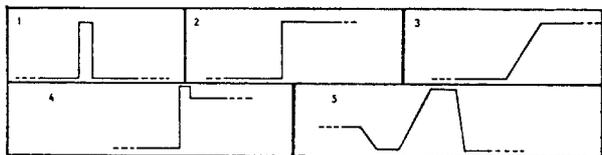


Fig. 1 : Différents profils de contours

Un contour est une frontière entre deux régions d'intensités photométriques différentes et uniformes, que l'on peut caractériser de manière générale par les paramètres suivants [3], (fig. 2) :

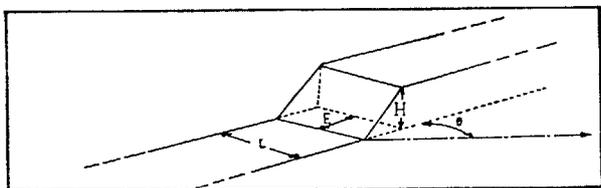


Fig. 2: Paramètres d'un contour idéal

- son épaisseur E, mesurant la largeur de transition.
- sa hauteur H, mesurant le contraste de luminosité entre le niveau haut et le niveau bas.
- sa longueur L
- son orientation θ , par rapport à l'axe horizontal de l'image.

La détection de contours peut alors se diviser en deux problèmes différents et complémentaire [4] :

1°) Un problème local : il s'agit de mettre en évidence des petits ensembles de points susceptibles d'appartenir à un contour grâce à des opérateurs locaux qui mesurent les paramètres E, H et éventuellement θ .

2°) Un problème global : il s'agit de relier entre eux ces ensembles de points (paramètre L) : c'est le suivi de contours.

Nous nous intéressons ici à la méthode locale de rehaussement d'image suivi d'un seuillage.

I. DETECTION DE CONTOURS PAR REHAUSSEMENT-SEUILLAGE

Le principe de cette méthode est indiqué sur la figure 3 :

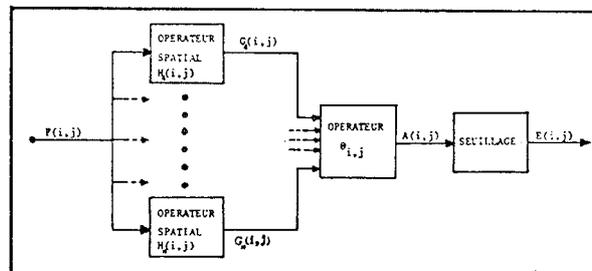


Fig. 3 : Système de détection de contours par rehaussement et seuillage d'image.

Une image numérisée $F(i,j)$ subit d'abord un traitement spatial par l'intermédiaire de N opérateurs linéaires $H_k(i,j)$ produisant un ensemble de N fonctions gradients de l'image :

$$G_k(i,j) = F(i,j) \otimes H_k(i,j)$$

où \otimes désigne la convolution bidimensionnelle.

Sur chaque voisinage centré de l'image F, les fonctions gradients sont combinées par l'intermédiaire d'un opérateur linéaire ou non linéaire $Q\{\cdot\}$, donnant une image $A(i,j)$ marquant plus nettement les discontinuités de luminosité ; c'est l'image rehaussée :

$$A(i,j) = Q\{G_k(i,j)\} \quad k=1, \dots, N$$

Il suffit alors de comparer chaque élément de l'image $A(i,j)$ par rapport à un seuil pour obtenir une image en tout ou rien qui est l'image des contours $E(i,j)$:

$$\begin{aligned} \text{Si } A(i,j) \gg t & \text{ existence d'un contour} \\ A(i,j) < t & \text{ pas de contour} \end{aligned}$$

En utilisant cette méthode, nous avons rencontré les deux problèmes suivants qu'il nous a paru important de résoudre.

- définition des caractéristiques d'un contour.
- Nous avons vu (fig. 1) qu'une image pouvait contenir

une très grande variété de types de contours. Or si tous les contours de l'image représentent une information intrinsèque de l'image, ils ne représentent pas forcément tous l'information utile que l'on désire obtenir. C'est à dire qu'il est intéressant de pouvoir a priori sélectionner quel type de contour est à accepter et de pouvoir rejeter les autres.

Nous voyons donc qu'un premier problème important est de définir une méthode permettant d'obtenir des opérateurs adaptés aux caractéristiques informationnelles désirées de l'image.

- Le deuxième problème réside dans la minimisation de la sensibilité des traitements aux bruits sur l'image. Le bruit (au sens large) peut avoir plusieurs sources dont les plus importantes sont, le bruit lié au senseur et le bruit aléatoire sur l'image (sel-et-poivre). Ce problème est caractérisé par la difficulté à obtenir la bonne valeur du seuil t , car ce dernier variera en fonction des conditions de prise de vue pour une même image.

Aussi, nous proposons dans cet article une méthode de générale permettant de synthétiser des opérateurs bidimensionnels, linéaires, non récursifs, adaptés, ainsi qu'une méthode de seuillage donnant un taux de fausses alarmes constant quelque soit le rapport S/B, ce qui se traduit par une détection à seuil constant.

II. SYNTHESE DES OPERATEURS ADAPTES

Les opérateurs de convolution spatiale sont entièrement définis par un ensemble de N matrices $M \times M$ $H_k(i,j)$, $k=1, \dots, N$.

Nous pouvons distinguer deux types d'opérateurs :

- les opérateurs différentiels
- les opérateurs adaptés.

Nous expliciterons ici que les opérateurs adaptés et nous renverrons le lecteur à la référence [3] pour les opérateurs différentiels.

Les opérateurs linéaires adaptés, non récursifs sont définis par un ensemble de matrices représentant une approximation discrète des contours idéaux à détecter pour différentes directions. Les plus connus sont ceux construits à partir des opérateurs différentiels de Prewitt [5], de Sobel [6], et de Kirsch [7]. Les fonctions gradients sont combinées en général par l'opérateur non linéaire $\theta\{.\}$:

$$A(i,j) = \theta \left\{ G_k(i,j) \right\} = \text{Max}_k \left\{ |G_k(i,j)| \right\}$$

La méthode que nous proposons consiste à chercher des bases bidimensionnelles de polynômes orthogonaux (matrices $H_k(i,j)$) adaptés à certaines caractéristiques

de contours. Pour plus de compréhension, nous allons d'abord développer le principe de la méthode dans le cas unidimensionnel.

1° - Opérateurs adaptés unidimensionnels.

Le problème peut se poser ainsi :

Déterminer une base H_f à N points ($N=2M+1$), adaptée à une fonction $f(h)$ et qui rejette par exemple, les fonctions de type constante ($g_1(h)=c$) et les fonctions "rampes" ($g_2(h)=h$).

Le problème s'écrit :

Trouver H_f tel que

$$\begin{aligned} & \langle H_f, f \rangle \text{ soit maximal} \\ & \text{sous les contraintes} \\ (1) \quad & \begin{cases} (2) \quad \langle H_f, g_1 \rangle = 0 \\ (3) \quad \langle H_f, g_2 \rangle = 0 \\ (4) \quad \langle H_f, H_f \rangle = A, \quad A > 0 \end{cases} \end{aligned}$$

Sans perte de généralité, prenons $g_1(h)=1$ et posons $H_f(h)=x_h$, $h=-M, \dots, M$.

Le problème posé par l'équation (1) revient à maximiser une forme quadratique $L(x_{-M}, \dots, x_M)$ définie positive qui s'écrit :

$$L(x_{-M}, \dots, x_M) = \sum_{-M}^M (f(h)x_h + a_1 x_h + a_2 h x_h + a_3 x_h^2)$$

La solution est alors donnée par :

$$(5) \quad x_h = -(a_1 + a_2 h + f(h))/2a_3 \quad h \in [-M, M]$$

La contrainte (2) nous donne :

$$(6) \quad a_1 = - \sum_{-M}^M f(h) / N$$

De (3), nous tirons :

$$(7) \quad a_2 = - \sum_{-M}^M h f(h) / 2 \sum_{-M}^M h^2$$

Enfin de (4), nous avons :

$$(8) \quad 4a_3^2 = \left(\sum_{-M}^M f(h)^2 - 2a_2 \sum_{-M}^M h^2 - a_1^2 \right) / A$$

Exemple :

Considérons la fonction $f(h)$ comme décrite sur la figure 4 :

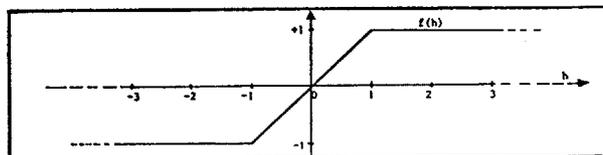


Fig. 4 : fonction f(h)

Prenons $N=5$ et $A=10$.



Des équations (5),(6),(7) et (8), nous tirons :

$$a_1 = 0 \quad ; \quad a_2 = -3/5 \quad ; \quad a_3 = 1$$

La solution $H_f(h)$ s'écrit alors :

$$H_f = (-1, 2, 0, -2, 1)$$

La réponse au signal $f(h)$ se traduit alors comme la convolution de la base H_f par le signal $f(h)$. On vérifie bien que cette réponse est nulle pour les signaux $g_1(h)$ et $g_2(h)$, et optimale pour $f(h)$ sous les contraintes données.

$$B_{00} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad B_{01} = \begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ -2 & -1 & 0 & 1 & 2 \\ -2 & -1 & 0 & 1 & 2 \\ -2 & -1 & 0 & 1 & 2 \\ -2 & -1 & 0 & 1 & 2 \end{bmatrix}$$

$$B_{10} = \begin{bmatrix} -2 & -2 & -2 & -2 & -2 \\ -1 & -1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 \end{bmatrix} \quad B_{11} = \begin{bmatrix} 4 & 2 & 0 & -2 & -4 \\ 2 & 1 & 0 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ -2 & -1 & 0 & 1 & 2 \\ -4 & -2 & 0 & 2 & 4 \end{bmatrix}$$

Remarque

- Il est bien évident que les bases adaptées trouvées par cette méthode de synthèse ne dépendent que de la forme du signal de référence et non pas des valeurs spécifiques prise en chaque point, i.e. le signal $(\dots, -1, -1, 0, 1, 1, \dots)$ aura, à un facteur de normalisation près, la même base adaptée que le signal $(\dots, 0, 0, 10, 20, 20, \dots)$
- La deuxième remarque est que l'accroissement de la complexité de la forme du signal de référence et/ou l'accroissement de la résolution spatiale obligera à prendre une base adaptée à plus de coefficients ($N \gg 7$).

2°- Base adaptée bidimensionnelle.

Nous allons maintenant décrire le principe de la méthode dans le cas bidimensionnel, et pour plus de clarté, nous allons le décrire à l'aide d'un exemple simple.

Soit un signal bidimensionnel 5×5 décrit par les matrices suivantes :

$$F_1 = \begin{bmatrix} -1 & -1 & 0 & 1 & 1 \\ -1 & -1 & 0 & 1 & 1 \\ -1 & -1 & 0 & 1 & 1 \\ -1 & -1 & 0 & 1 & 1 \\ -1 & -1 & 0 & 1 & 1 \end{bmatrix} \quad F_2 = \begin{bmatrix} -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$F_3 = \begin{bmatrix} -1 & -1 & -1 & -1 & 0 \\ -1 & -1 & -1 & 0 & 1 \\ -1 & -1 & 0 & 1 & 1 \\ -1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{bmatrix} \quad F_4 = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 1 & 1 \\ -1 & -1 & 0 & 1 & 1 \\ -1 & -1 & -1 & 0 & 1 \\ -1 & -1 & -1 & -1 & 0 \end{bmatrix}$$

On désire trouver une base 5×5 qui maximise la réponse à ce signal et qui minimise la réponse à des signaux continus (constantes et rampes) décrits par les matrices :

La base $B = (B_{02}, B_{20}, B_{03}, B_{30})$ est alors solution de :

$$B_{02} = H_1 + a_1 B_{01} + b_1 B_{10} + c_1 B_{11}$$

$$B_{20} = H_2 + a_2 B_{01} + b_2 B_{10} + c_2 B_{11}$$

$$B_{03} = H_3 + a_3 B_{01} + b_3 B_{10} + c_3 B_{11}$$

$$B_{30} = H_4 + a_4 B_{01} + b_4 B_{10} + c_4 B_{11}$$

avec les contraintes d'adaptativité qui s'écrivent :

$$\langle B_{0i}, H_i \rangle + a_i \cdot \|B_{01}\|^2 = 0, \quad i=1,2,3,4$$

$$\langle B_{10}, H_i \rangle + b_i \cdot \|B_{10}\|^2 = 0$$

$$\langle B_{11}, H_i \rangle + c_i \cdot \|B_{11}\|^2 = 0$$

Sachant que la base est orthogonale, nous avons :

$$\langle B_{01}, H_1 \rangle = 30, \quad \langle B_{10}, H_1 \rangle = 0, \quad \langle B_{11}, H_1 \rangle = 0$$

$$\langle B_{01}, H_2 \rangle = 0, \quad \langle B_{10}, H_2 \rangle = 30, \quad \langle B_{11}, H_2 \rangle = 0$$

$$\langle B_{01}, H_3 \rangle = 20, \quad \langle B_{10}, H_3 \rangle = 20, \quad \langle B_{11}, H_3 \rangle = 0$$

$$\langle B_{01}, H_4 \rangle = -20, \quad \langle B_{10}, H_4 \rangle = 20, \quad \langle B_{11}, H_4 \rangle = 0$$

Or, comme $\|B_{10}\|^2 = \|B_{01}\|^2 = 50$ et $\|B_{11}\|^2 = 100$, on obtient :

$$a_1 = -3/5, \quad b_1 = 0, \quad c_1 = 0$$

$$a_2 = 0, \quad b_2 = -3/5, \quad c_2 = 0$$

$$a_3 = -2/5, \quad b_3 = -2/5, \quad c_3 = 0$$

$$a_4 = -2/5, \quad b_4 = 2/5, \quad c_4 = 0$$

Ce qui donne la base bidimensionnelle adaptée suivante

$$B_{02} = 5H_1 - 3B_{01}, \quad B_{30} = 5H_3 - 2(B_{01} + B_{10})$$

$$B_{20} = 5H_2 - 3B_{10}, \quad B_{03} = 5H_4 - 2(B_{10} - B_{01})$$

Nous avons ainsi une base orthogonale bidimensionnelle $(B_{02}, B_{20}, B_{03}, B_{30})$ adaptée au signal de référence et rejetant les signaux continus (constantes-rampes). Quelques photos, données en annexe, illustrent les diverses possibilités de cette méthode.

Il est bien entendu que les remarques énoncées pour le cas unidimensionnel sont encore valables pour le cas bidimensionnel. En particulier sur la nécessité d'augmenter la taille NxN des matrices de la base pour un signal de complexité accrue et/ou une résolution spatiale plus fine.

III. METHODE DE DETECTION T.F.A.C.

Nous avons vu que le deuxième problème important des méthodes de détection de contours par rehaussement d'image et seuillage est le choix de la valeur à affecter au seuil, car cette valeur dépend du rapport S/B sur l'image ou bien encore du changement de luminosité entre deux prises de vue de la même scène.

Pour résoudre ce problème, nous proposons une méthode permettant d'avoir un taux de fausses alarmes constant sur la détection de contour, ce qui se traduit par une détection à seuil constant, dont le principe est le suivant :

Soit une image numérique PxP, F(i,j) telle que

$$F(i,j) = a_{i,j} \quad , \quad i,j \in [1,P]^2$$

où $a_{i,j}$ est la brillance du pixel (i,j).

Soit une base adaptée NxN, d'opérateurs $B_{hk}(i,j)$, pour $i,j \in [1,N]^2$.

La convolution de l'image F par la base adaptée nous donne sur chaque fenêtre de convolution NxN :

$$b_{hk} = \sum_{i,j} B_{hk}(i,j) \cdot a_{i,j}$$

où $a_{i,j}$ est la valeur du pixel de l'image F sur la fenêtre NxN.

On peut alors définir les quantités suivantes :

- pour chaque opérateur de la base

$$d_{hk} = (b_{hk} / \|B_{hk}\|)^2$$

- la norme carrée de l'image centrée de moyenne nulle sur chaque fenêtre NxN

$$S_1 = \sum_{i,j} (a_{i,j} - Z)^2$$

où Z est la brillance moyenne

On peut alors effectuer la convolution de l'image centrée par la base adaptée. Soit donc à calculer sur chaque fenêtre NxN :

$$(9) \quad b_{hk} = \sum_{i,j} (a_{i,j} - Z) \cdot B_{hk}(i,j)$$

Nous pouvons définir par exemple la base adaptée comme B_{02} , B_{20} , B_{30} , B_{03} les opérateurs adaptés au signal de référence.

On définit alors une mesure du bruit de l'image centrée dans la fenêtre de traitement NxN par :

$$S_3 = (S_1 - (d_{02} + d_{20})) / L$$

$$S_4 = (S_1 - (d_{03} + d_{30})) / L$$

où L représente le degré de liberté du système

On calcule alors les valeurs :

$$C_3 = (C_1 / S_3)^{1/2} \quad \text{et} \quad C_4 = (C_2 / S_4)^{1/2}$$

où $C_1 = \text{Max}(d_{02}, d_{20})$ et

$C_2 = \text{Max}(d_{03}, d_{30})$

qui représentent les valeurs maximales du gradient en un point de l'image centrée, par rapport à tout ce qui n'est pas adapté à la base choisie.

Enfin,

$$C = \text{Max}(C_3, C_4)$$

donne la valeur du gradient normalisé sur une fenêtre NxN de l'image originale.

L'image des contours est donnée par une détection par rapport à un seuil. Soit t ce seuil, alors

$$\begin{aligned} C(i,j) \geq t &\Rightarrow E(i,j) = 1 & i,j \in [1,P]^2 \\ C(i,j) < t &\Rightarrow E(i,j) = 0 \end{aligned}$$

Cet algorithme permet d'avoir un seuil de détection constant quelque soit le rapport S/B sur l'image à taux de fausses alarmes constant. A titre d'exemple, nous donnons dans le tableau suivant une estimation du seuil en fonction du taux de fausses alarmes désiré.

Taux de fausses alarmes	Seuil
10 %	3
5 %	4
1 %	8
0.5 %	11

Le résultat de cette méthode est illustré, en Annexe, par des photos.

IV. CONCLUSION

Pour répondre aux problèmes liés par la détection de contours par rehaussement d'image et seuillage, nous avons présenté dans cet article une nouvelle méthode de synthèse générale d'opérateurs différentiels linéaires et non récursifs adaptés aux types de contours, ainsi qu'une méthode permettant d'obtenir un seuil constant quelque soit le rapport S/B de l'image, à taux de fausses alarmes donné. L'avantage et l'originalité de cette méthode par



rapport à celles existantes est qu'elle donne un moyen de sélectionner a priori quel type de signal est à détecter comme un contour et de rejeter le reste.

Elle opère donc de manière sélective sur l'information contenue dans les contours d'une image et permet d'obtenir seulement l'information désirée, le problème du seuil ne se posant plus, à taux de fausses alarmes désiré.

Nous avons testé ces méthodes sur de nombreuses images de grande diversité, et les résultats obtenus sont encourageants. Toute fois, un effort d'analyse du signal (nature des contours, etc...) doit être effectué si l'on veut effectivement tirer partie de toute la puissance de la méthode d'extraction de contours adaptative.

Remerciements.

Les auteurs tiennent ici à remercier tout particulièrement Monsieur SILVERA, pour l'aide fructueuse apportée à ces travaux.

Références bibliographiques:

- /1/ Rosenfeld A. et Kak A.C., Digital Picture Processing. New York. Academic Press 1976.
- /2/ Griffith A.K., 'Edge Detection in Simple Scene Using a priori Information', I.E.E.E. Trans. C. Vol-22, pp371-381, 1973
- /3/ Abdou I.E. et Pratt W.K., 'Quantitative Design and Evaluation of Enhancement/Thresholding Edge Detectors', Proc. I.E.E.E., Vol-67, n°5, pp 753-763, mai 1979
- /4/ Basseville M., 'Détection de contours: méthodes et études comparatives', Annal. Télécom., 34, n° 11-12, pp 559-579, 1979.
- /5/ Prewitt J.M.S., 'Object enhancement and extraction' Picture Proc. and Psychopictories, B.S. Lipkin, A. Rosenfeld, Ed. Academic Press, New York, 1970.
- /6/ Duda R.O. et Hart P.E., Pattern Classification and Scene Analysis, New York, Wiley, 1973.
- /7/ Kirsch R., 'Computer determination of the constituent structure of biological images', Comput. Biomed. Res., Vol. 4, n° 3, pp 315-328, 1971.

ANNEXE
+++++

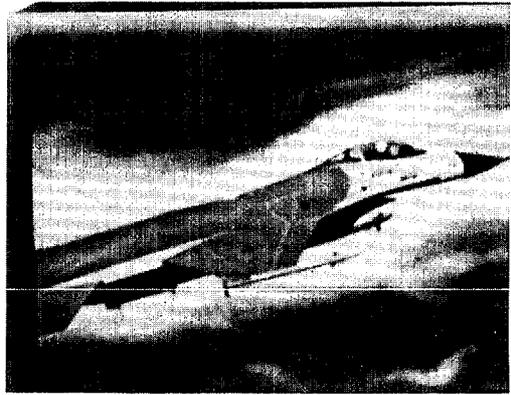


Photo n° 1

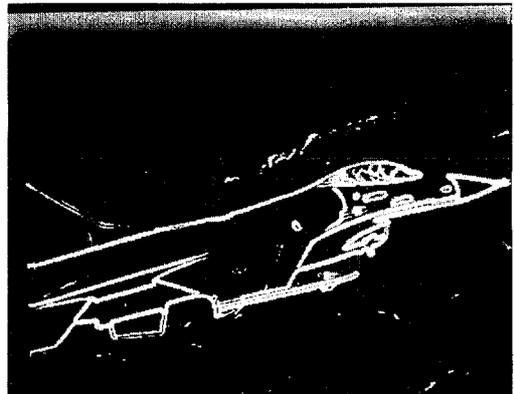


Photo n° 2

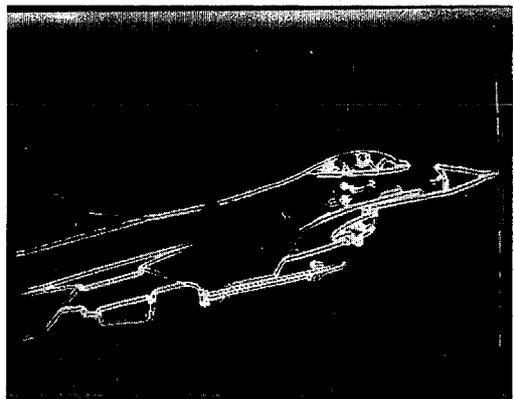


Photo n° 3

METHODE ADAPTATIVE D'EXTRACTION DE CONTOURS A SEUIL CONSTANT



Photo n° 4

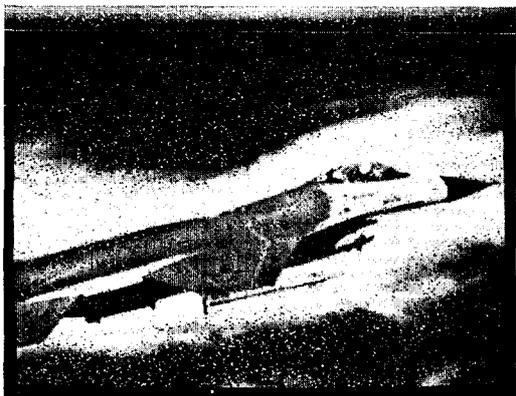


Photo n° 5



Photo n° 6

Illustration de la détection TFAC.

Photo n° 5 : image originale bruitée.

Photo n° 6 : détection non TFAC.

Photo n° 7 : détection TFAC.



Photo n° 7

