

# Application

---

## De la création d'images composites à la coopération d'algorithmes en contexte multicapteurs

---

### *From Image Fusion to Algorithms Cooperation*

par **Philippe CORNIC, Alain MICHEL**

ONERA — 29, avenue de la Dr. Leclerc  
F-92320 Châtillon

#### Résumé

L'intérêt de l'utilisation de données multi-sources pour l'interprétation et la classification d'images satellitales n'est plus à démontrer. La fusion d'informations complémentaires et/ou redondantes peut s'opérer à différents niveaux. La fusion pixel (bas niveau) se fonde exclusivement sur la radiométrie et conduit à des images composites destinées à synthétiser sur un même support des informations issues de chacun des modes. Les résultats obtenus à partir d'images Spot et Ers sont assez décevants. Une fusion se plaçant à un niveau d'abstraction plus élevé manipule des décisions prises par un ou plusieurs classificateurs pour élaborer une décision finale. Cette démarche est illustrée par une classification à partir d'un couple d'images Spot et Ramses (Sar aéroporté). Enfin, faisant l'objet de la dernière partie, une utilisation coopérative d'informations exogènes floues associées à un modèle générique simple assure la détection automatique d'aéroports.

**Mots clés :** Fusion de données, Classification Multi-capteur, Détection d'objets complexes.

#### Abstract

*This paper deals with a few aspects of radar-optronic image fusion. Low level fusion uses radiometry to create synthetic images. A higher level fusion involves decisions taken by one or several sensors (or algorithms) so as to elaborate a final decision on the ground of evidences at hand. This approach is illustrated by a classification problem with a couple of Spot and Ramses (airborne Sar) images. The last part is dedicated to synergic use of Ers and Spot images coupled with ancillary informations and a model, to automatically detect airports.*

**Key words :** Image fusion, Classification, Algorithms cooperation.

## 1. Introduction

Cet article présente une synthèse sur le plan applicatif de trois des aspects que peut revêtir la fusion d'images optiques et millimétriques. Les résultats présentés ont été obtenus dans le cadre de travaux d'aide à la photo-interprétation.

La fusion pixel (bas niveau), dont les résultats sont assez décevants, est évoquée dans la première partie. Elle se fonde exclusivement sur la radiométrie et conduit à la création d'images composites. La deuxième partie présente et compare les performances de deux stratégies de fusion appliquées à un problème de classification. Dans la première, fusion centralisée, un centre de décision unique élabore une décision au vue de l'ensemble de l'information disponible. Dans la seconde, fusion décentralisée, ce sont des décisions prises indépendamment sur chacune des images qui sont manipulées. La troisième partie est consacrée à l'utilisation coopérative d'informations exogènes (modèle numérique de

terrain, cartographie sommaire) pour permettre la détection automatique d'aéroports à partir d'images Spot panchromatique et radar (Ers).

## 2. Images composites pour la photo- interprétation

L'essor que connaît l'imagerie satellitale grâce au lancement de nouveaux capteurs SAR spatiaux (Synthetic Aperture Radar) et à l'avènement de la haute résolution visible est sur le point de générer un flux de données qui dépasse les capacités d'analyse des photo-interprètes. La motivation principale de la fusion pixel destinée à la photo-interprétation est de concentrer sur un support unique, sous forme d'une image composite, des informations spécifiques à chacune des images initiales. Les renseignements

que l'on souhaite mettre en valeur peuvent être aussi bien d'ordre géomorphologique (relief) que cartographique (réseaux, superstructures...). La seconde motivation relève d'un facteur humain. La diffusion d'images SAR spatiales est relativement récente, leur interprétation n'est pas aisée et, engendre quelques réticences de la part des photo-interprètes. Des images composites faisant clairement ressortir les points forts du radar tout en restant facilement interprétables seraient susceptibles d'emporter plus facilement leur adhésion.

L'idée de fusionner des images visibles avec des images SAR n'est pas nouvelle. R. Welch [1] a généré une image composite en utilisant une transformée IHS inverse. Il disposait d'une image SIR-B avec laquelle il a modulé le canal intensité et des bandes 3 et 4 de Landsat respectivement utilisées pour la teinte et la saturation. R. Harris [2] a étendu cette méthodologie à l'intégration d'images radar avec d'autres données de télédétection. Y. Yésou [3] a étudié pour ce type d'approche le poids qu'il convient d'attribuer au substitut du canal intensité. Il a préconisé l'utilisation d'une combinaison linéaire optronique / radar pour éviter que le radar ne prenne une place trop importante lors de la transformation inverse, et ne vienne de ce fait altérer l'information Spot (perte du parcellaire par exemple). L'Analyse en Composantes Principales fait partie des méthodes qu'il a expérimentées pour une fusion Seasat et Spot XS. Il a ainsi obtenu un réseau hydrographique rehaussé et la mise en évidence au sein du parcellaire agricole de zones humides initialement non visibles dans Spot. On remarquera toutefois que les images utilisées sont de basse résolution.

La fusion pixel ne va pas sans poser de problèmes. Les contraintes concernant le recalage des images entre elles sont très fortes. L'obligation d'utiliser des informations qui proviennent de mêmes régions impose une tolérance inférieure au pixel. Lorsque les images présentent des géométries différentes, comme c'est le cas des images Sar et des images optroniques, des traitements préalables s'imposent. Outre ces problèmes s'ajoutent ceux posés par le multi-dates et des résolutions différentes.

Les différentes évaluations que nous avons faites de ces méthodes à partir d'images Spot panchromatique de résolution 10m et Ers 1 ramené à la résolution de 20 m se sont révélées, pour la photo-interprétation, assez décevantes. L'utilisation de la Transformée en Ondelette, à l'instar de ce qui se pratique avec la transformée IHS s'est avérée tout aussi insatisfaisante. Les images composites générées sont d'un aspect peu engageant, et supporte difficilement la comparaison avec l'image Spot originale. Nous n'avons pas pour autant renoncé à leur utilisation, mais dans un tout autre contexte : celui de la classification.

### 3. Fusion centralisée, fusion décentralisée et classification

Nous nous plaçons ici dans le cadre d'un problème de **classification supervisée**. Sur la base d'un couple d'images, Spot panchro-

matique et Ramses (radar aéroporté ONERA) recalées, du site de Dampierre (Illustrations 1 et 2), nous nous proposons d'établir une carte de l'occupation du sol. Il s'agit d'affecter des régions issues de segmentations des images à des classes répertoriées a priori (champ, eau, forêt, bâtiment). Le premier paragraphe concerne la fusion centralisée où une décision est prise sur une image com-



Illustration 1. – Spot Dampierre 10 m. © CNES 1990 – Distribution SPOT IMAGE.



Illustration 2. – Ramses Dampierre 10 m.

posite. Le deuxième paragraphe est consacré à la fusion décentralisée où l'on manipule des décisions prises indépendamment sur chaque image. Le troisième paragraphe compare les résultats obtenus avec ceux d'autres classifications.

### 3.1. CONSTITUTION D'UNE RÉFÉRENCE ET MULTI-DATES

Les méthodes de classification supervisées nécessitent une référence afin d'initialiser correctement le processus de classification et de mesurer précisément la fiabilité des résultats obtenus. Le multi-dates pose pour la constitution d'une référence de sérieux problèmes. Sauf cas rarissimes, les images issues de plates-formes différentes ne sont pas enregistrées simultanément : au mieux à quelques heures d'intervalle au pire à quelques années. S'il est vrai qu'à une résolution décimétrique les entités surfaciques demeurent souvent inchangées, l'occupation au sol, en particulier pour le parcellaire agricole ou les zones inondables, est fonction de la saison. On peut constater que les bancs de sable à fleur d'eau dans l'image radar sont immergés dans l'image Spot (Illustration 1 et 2).

La vérité terrain au moment des prises de vues est très rarement disponible. Il est cependant généralement admis que pour des grandes classes d'occupation du sol, telles celles définies par la BD Carto de l'IGN, moins de 1% de la surface totale de l'image change d'affectation en 1 an. C'est pourquoi, nous avons choisi de constituer un modèle simple en 4 classes (eau, forêt, champ, bâtiment) en actualisant une carte IGN au 1/25 000 à l'aide de toutes les informations dont nous disposons (images et photos aériennes). Ce choix a pour conséquence une plus grande diversité intra-classe dont il faudra tenir compte lors de l'apprentissage. La Référence est présentée dans l'Illustration 3.

### 3.2. FUSION CENTRALISÉE

Dans ce type d'approche, une décision unique est élaborée au vue de l'ensemble de l'information disponible (Figure 1).

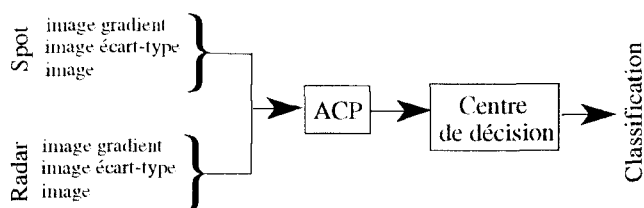


Figure 1. – Schéma de fusion centralisée

Les meilleurs résultats selon ce schéma ont été obtenus en utilisant une méthode de classification région par mesure de texture sur une image composite bicanal. L'avantage de l'image composite par rapport aux deux canaux originaux sera explicité au § 3.4.1.



Illustration 3. – Référence adoptée.

#### 3.2.1. Obtention de l'image composite

L'image composite, utilisée pour la **classification de régions** par mesure de texture, a été obtenue en conservant les deux premier axes d'une Analyse en Composante Principale (ACP) effectuée à partir de 6 images : les images Spot et radar et les images gradient classique et écart-type (calculés sur un pavé 5\*5) qui en sont dérivées. Selon la terminologie consacrée, les individus sont les pixels et les 6 variables associées sont respectivement : les radiométries des images Spot et Radar pour le pixel considéré, et les valeurs de ce même pixel dans chacune des images gradient et écart-type. Les gradients et écart-types ont été introduits pour prendre en compte une mesure de l'hétérogénéité locale des images.

L'image bicanal est segmentée par un algorithme de croissance de régions, dont le seuil est adaptatif et dépend de l'intensité et du gradient calculés localement au point d'accès. L'apprentissage est effectué sur un échantillonnage systématique en pavés 16\*16, où l'on retient un pavé sur 4.

#### 3.2.2. Classification région par mesure de texture

La classification se fait par mesure de texture (Mt). La méthode retenue pour le calcul de la distance permettant l'affectation des régions issues de la segmentation à la classe la plus proche, s'appuie sur des matrices de cooccurrences sur 16 niveaux. Nous avons adopté une mesure de type inertie, tenant compte de l'organisation spatiale des distributions au sein des matrices

de cooccurrences. Elle s'est avérée donner de meilleurs résultats que la distance du  $\chi^2$  et est définie de la manière suivante :

$Fr$  désigne la matrice des cooccurrences sur 16 niveaux de la région  $r$ , calculée sur toute la région avec une distance  $dx = dy = 1$ .

$Vr$  est le vecteur de taille  $M$ , ordonné dans le sens décroissant, des éléments non nuls de  $Fr$ .

$Fc$  désigne la matrice des cooccurrences sur 16 niveaux de la classe  $c$ , calculée sur l'ensemble des échantillons d'apprentissage de cette classe avec une distance  $dx = dy = 1$ .

$Vc$  est le vecteur de taille  $N$ , ordonné dans le sens décroissant des éléments non nuls de  $Fc$ .

Soit  $xr(k), yr(k)$  tels que :  $Vr(k) = Fr(xr(k), yr(k))$   
( $k = 1..M$ )

Soit  $xc(q), yc(q)$  tels que :  $Vc(q) = Fc(xc(q), yc(q))$  ( $q = 1..N$ )

La distance d'une région  $r$  à une classe  $c$  est donnée par :

$$D(r, c) = \sum_{k=1}^{\inf\{N, M\}} [(xr(k) - xc(k))^2 + (yr(k) - yc(k))^2] * |Vr(k) - Vc(k)|$$

La fiabilité globale de la classification (Mt ACP) obtenue est de 75%. Le tableau ci-après donne la matrice de confusion.

**Tableau 1. — Matrice de confusion classification Mt ACP.**

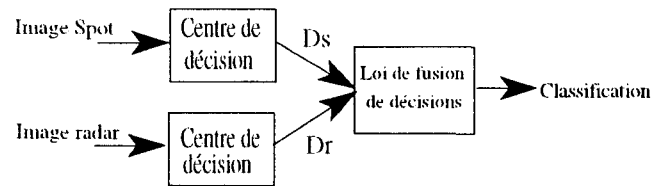
Classification	Champ	Eau	Forêt	Bâtiment
Réf. Champ	82%	1%	10%	5%
Réf. Eau	36%	60%	3%	20%
Réf. Forêt	18%	1%	79%	17%
Réf. Bâtiment	45%	0%	15%	38%

La classe bâtiment est de loin la moins bien classée. Il y a à cela deux raisons : d'une part les bâtiments sont en faible proportion dans l'image et donc peu représentés au niveau de l'apprentissage et d'autre part la référence est incomplète, quelques bâtiments semblent avoir été oubliés (absents de la carte).

A partir des même images nous avons aussi adopté une démarche alternative, où la fusion s'opère à un niveau d'abstraction plus élevé.

### 3.3. FUSION DÉCENTRALISÉE

Les images originales sont segmentées à l'aide du même algorithme que précédemment. La fusion porte sur les décisions relatives aux régions, prises indépendamment sur chacune des images Spot et Radar (Figure 2).



**Figure 2. — Schéma de fusion décentralisée.**

C'est à partir de classifications monomodes par maximum a posteriori, avec une loi de fusion bayésienne qu'ont été obtenus les meilleurs résultats.

#### 3.3.1. Classification monomode par Maximum a Posteriori

Étant donnée une mesure  $m$  de  $p$  variables caractérisant un individu, ici une région issue de la segmentation, il s'agit de décider de son appartenance à l'une des 4 classes répertoriées a priori, représentatives de la nature du sol, ici champs ( $H_0$ ), eau ( $H_1$ ), forêt ( $H_2$ ) et bâtiments ( $H_3$ ). On décide de l'hypothèse dont la probabilité a posteriori est la plus grande. Le théorème de Bayes permet de calculer aisément la probabilité a posteriori d'une hypothèse, connaissant les probabilités a priori de chacune des hypothèses, données par leurs fréquences relatives d'apparition au sein de l'échantillon d'apprentissage, et les probabilités conditionnelles  $p(m/H_i)$  estimées à partir des histogrammes normalisés calculés sur une population pour laquelle on sait l'hypothèse vérifiée.

$$\forall i = 0 \dots 3 \quad p(H_i/m) = \frac{p(m/H_i) p(H_i)}{\sum_{j=1}^3 p(m/H_j) p(H_j)}$$

Pour chacune des populations d'apprentissage Spot et radar, les histogrammes font apparaître des distributions unimodales. Les matrices de corrélation des 10 variables (paramètres d'Haralick) calculées sur les distributions d'apprentissage montrent que l'hypothèse d'indépendance n'est pas légitime. Le cas multivarié est très délicat à traiter sauf si l'on fait l'hypothèse gaussienne. La densité de probabilité conjointe  $p(m/H)$  est alors entièrement déterminée par les moments d'ordre 1 (moyenne) et 2 (matrice de covariance).

Sous cette hypothèse, la meilleure combinaison de variables s'est révélée être :  $m =$  (moyenne, écart-type, homogénéité).

L'homogénéité, variable introduite par Haralick [14] à partir des matrices de cooccurrences est ici calculée de la manière suivante :

$$h = \sum_{i=1}^{16} \sum_{j=1}^{16} \frac{1}{1 + (i - j)^2} p(i, j, -45)$$

où  $p(i, j, -45)$  désigne la fréquence d'apparition d'un niveau  $j$  suivant une direction de  $-45^\circ$  et à une distance  $dx = dy = 1$  d'un niveau  $i$ , pour la région à classer.

Nous donnons ci-après les matrices de confusion issues des classifications bayésiennes.

**Tableau 2. — Matrice de confusion radar.**

Classification	Champ	Eau	Forêt	Bâtiment
Réf. Champ	64%	7%	24%	2%
Réf. Eau	11%	84%	2%	0%
Réf. Forêt	36%	2%	57%	1%
Réf. Bâtiment	54%	17%	9%	18%

Fiabilité globale de la classification Map de l'image radar : 64%.

**Tableau 3. — Matrice de confusion SPOT.**

Classification	Champ	Eau	Forêt	Bâtiment
Réf. Champ	64%	15%	18%	1%
Réf. Eau	8%	77%	13%	0%
Réf. Forêt	9%	4%	85%	0%
Réf. Bâtiment	50%	0%	13%	35%

Fiabilité globale de la classification Map de l'image Spot : 71%.

### 3.3.2. Fusion Bayésienne : Maximum a priori

Soit  $D_s$  la décision concernant l'affectation d'une région dans l'image Spot.

Soit  $D_r$  la décision concernant l'affectation de cette même région dans l'image radar.

Dans le cas où les prises de décision sont indépendantes, ce qui peut être raisonnablement envisagé ici :

$$p(H_i/D_s, D_r) = \frac{P(D_s/H_i) P(D_r/H_i) P(H_i)}{P(D_s, D_r)}$$

Les  $P(D_s/H_i)$  et  $P(D_r/H_i)$ , qui représentent les vraisemblances de l'hypothèse  $H_i$  dans les deux classifications sont approximées par les matrices de confusion issues des classifications mono-mode.  $p(D_s, D_r)$  demeure constant quelque soit  $i$ .

La loi de fusion que nous avons retenue est celle du **maximum a posteriori** :

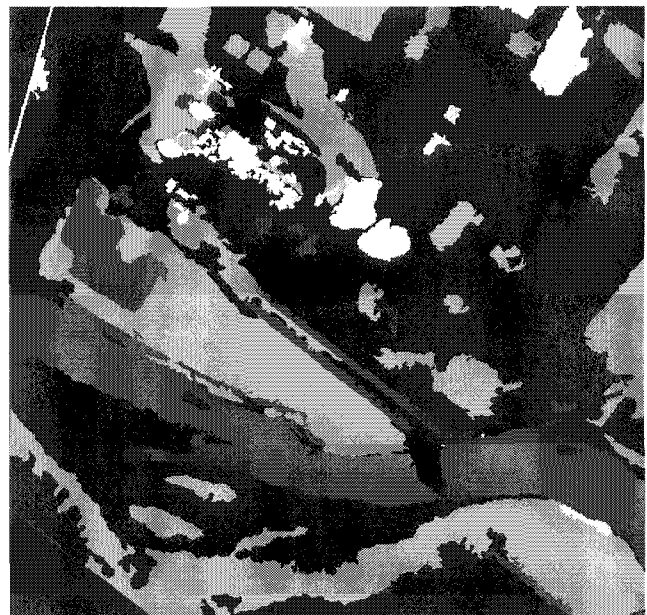
$$\text{Décision : } \max_i (p(D_s/H_i) p(D_r/H_i) P(H_i))$$

La classification bimode, notée C Map-F Map (pour Classification Map – Fusion Map), obtenue donne la matrice de confusion suivante :

**Tableau 4. — Matrice de confusion bimode (C Map-F Map)**

Classification	Champ	Eau	Forêt	Bâtiment
Réf. Champ	76%	4%	16%	2%
Réf. Eau	14%	82%	2%	0%
Réf. Forêt	13%	3%	82%	0%
Réf. Bâtiment	37%	0%	10%	51%

La fiabilité globale en bimode est de 77%. L'illustration 4 permet d'apprécier le résultat de cette classification.



**Illustration 4. – Classification Map-Fusion Map.**

La fusion de décision améliore ici de façon sensible la fiabilité globale de la classification. L'apport essentiel concerne les classes champs et bâtiments pour lesquelles l'amélioration est flagrante.

L'approche décentralisée permet ici un gain de 2% de fiabilité globale par rapport la fusion centralisée. Cependant aux vues d'autres expérimentations que nous avons conduites, il est difficile de dégager des conclusions définitives quant à ce dernier point.

### 3.4. COMPARAISON AVEC D'AUTRES MÉTHODES DE CLASSIFICATION

Nous n'aborderons pas les classifications à celle des pixels dont les performances sont nettement inférieures à celle des classifications

région. D'autre part l'aspect « pointilliste » des classifications obtenues est réhhibitoire.

### 3.4.1. Comparaison de classifications région

La figure 3 donne les performances globales de trois méthodes de classification région : mesure de texture (Mt) (cf. § 3.2.2),  $k$  plus proches voisins (Kppv) et maximum a posteriori (Map) (cf. § 3.3.1), respectivement sans fusion (classification monomode), selon un schéma centralisé, et avec fusion décentralisée.

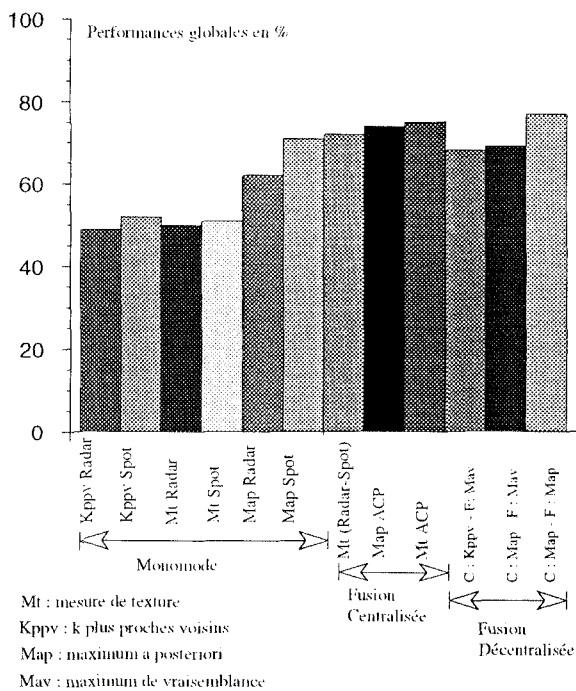


Figure 3. – Performances globales de différentes classifications.

#### Monomode :

Parmi les classifications monomodes on remarquera que c'est le Map qui donne les meilleurs résultats.

La méthode des Kppv a été mise en oeuvre avec les paramètres suivants : affectation d'un individu à une classe si au moins 5 des 10 plus proches voisins appartiennent à cette classe (distance de mahalanobis). La sélection des variables pertinentes dépend du capteur : moyenne, écart-type pour le radar, et moyenne, écart-type, inertie pour Spot, cette dernière étant définie comme suit :

$$I = \sum_{i=1}^{16} \sum_{j=1}^{16} (i - j)^2 p(i, j, -45)$$

#### Centralisée :

Une classification par mesure de texture utilisant les deux premiers canaux de l'ACP (Mt ACP), présentée au § 3.2.2, permet

l'obtention de 3% de fiabilité en plus par rapport à la même méthode utilisant à la place directement les canaux Radar et Spot (Mt Radar-Spot). On peut imputer cette amélioration à l'utilisation de canaux décorrelés et à l'intégration dans les composantes principales d'une mesure d'hétérogénéité.

La classification par Map de l'ACP (MAP ACP) est faite avec les variables suivantes : (moyenne, écart-type, inertie, kurtosis, skewness).

#### Décentralisée :

Nous avons aussi évalué pour notre application la loi de fusion du maximum de vraisemblance, appliquée aux classifications Kppv (C Kppv - F Max) et aux classifications Map (C Map - F Max). Les performances obtenues sont inférieures à celles de la fusion selon le Map des classifications Map (C Map - F Map).

### 3.5. CONCLUSION

La fusion optronique / radar permet, avec un couple d'image Spot panchromatique, Sar Ramses (Aéroporté), d'améliorer sensiblement les performances des classifications monomodes, avec ici un léger avantage pour le schéma décentralisé.

L'intérêt de ce schéma réside surtout dans sa souplesse. Les segmentations des images Spot et radar conduisent à des régions (individus) différentes. La relation du § 3.3.2 n'est applicable que lorsque les décisions  $D_s$  et  $D_r$  concernent le même individu, ce qui est très facilement réalisable en considérant comme nouveaux individus ceux résultant de l'intersection des différentes régions Spot et radar. Les prises de décisions peuvent donc dans chacun des modes concerner des régions différentes.

Cependant, nous avons pu vérifier que l'apport d'un Sar **Spatial** (Ers) dans un processus de classification est très limité dès lors que l'on dispose d'informations multispectrales (Landsat TM ou Spot XS par exemple) qui permettent d'obtenir de bien meilleurs taux de bonne classification.

## 4. Fusion haut niveau pour la détection automatique de sites

La détection automatique de sites (aéroports) sur des images optronique (satellitaires ou aériennes) a été récemment étudiée à l'ONERA [7], [8]. La détection automatique d'objets cartographiques complexes associant imagerie SAR et optronique n'apparaît, à notre connaissance, dans la littérature qu'à travers le projet PASTIS de l'INRIA [9], [13]. V. Clément y a développé une extension du système d'interprétation de scènes MESSIE appliqué à la détection.



L'objectif de notre étude était de montrer l'intérêt de l'utilisation d'informations exogènes pour accélérer et améliorer la fiabilité de la détection de sites sur un ensemble d'images préalablement recalées. La méthode recherche des candidats par adéquation à un modèle générique, en appliquant sur chacune des images disponibles des filtres opérant sur des caractéristiques complémentaires des sites recherchés (radiométrie, forme, contexte). Un formalisme flou s'est avéré particulièrement adapté à la prise en compte simultanée d'informations hétérogènes telles que la surface équivalente radar (Ser) ou la topographie. Nous nous focaliserons ici sur un type de sites : les aéroports. La détection débute par la recherche de la piste principale, suivie d'une phase de confirmation où l'on cherche, à proximité de la piste, les taxiways et les superstructures associées.

## 4.1. DONNÉES DISPONIBLES ET LEUR RÔLE

Les différentes données utilisées ont préalablement été recalées sur un référentiel géographique commun (WGS 72). Nous disposons :

- d'une image Sar ERS 1 géocodée (Lambert 3), à la résolution de 30 m, consacrée à la détection de la piste principale (Illustration 5).
- d'une image Spot panchromatique (10 m) utilisée en phase de validation pour la détection des taxiways (Illustration 6).
- d'informations cartographiques sommaires donnant une répartition en 13 postes de l'occupation du sol, largement suffisante pour notre objectif. Les zones aquatiques et forestières y sont généralement bien délimitées et les agglomérations importantes y sont également répertoriées. Elles seront utilisées pour l'élaboration d'un masque. En excluant la construction d'un aéroport en zone urbaine, forestière ou aquatique, on restreint spatialement la zone de recherche, accélérant d'autant les traitements.
- d'informations altimétriques. Des contraintes très strictes régissent l'inclinaison du terrain lors de la construction des aéroports. Cela permettra d'éliminer les régions trop pentues.

Le choix d'ERS pour la détection de la piste principale se justifie par le fait qu'il y a dans Spot une forte variation de la radiométrie de la piste rapportée à l'histogramme total (particulièrement sensible sur l'image d'Istres) et qu'elle est tributaire du matériau utilisé pour la construction.

## 4.2. LE MODÈLE GÉNÉRIQUE D'AÉROPORT

Afin d'avoir un modèle d'aéroport indépendant des images, tout en étant le plus exhaustif possible, nous nous sommes penchés sur la littérature spécialisée.

L'ouvrage de référence dans le domaine civil est « Airport engineering » [10]. Les dimensions des différentes structures y sont

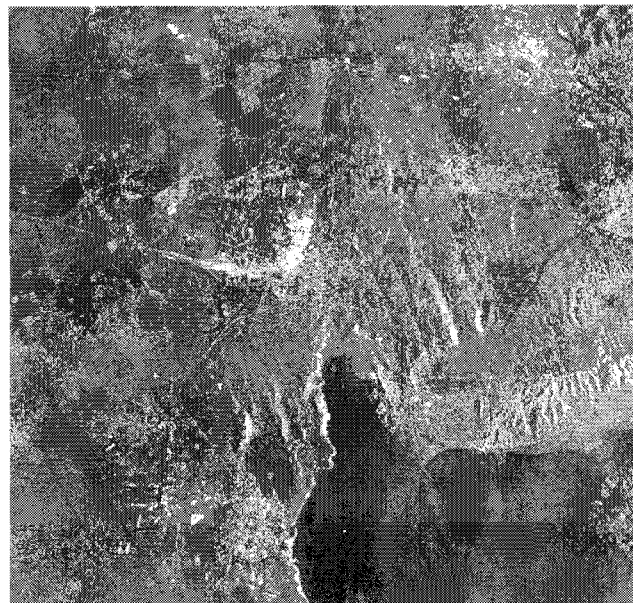


Illustration 5. – Ers Istres 30 m.

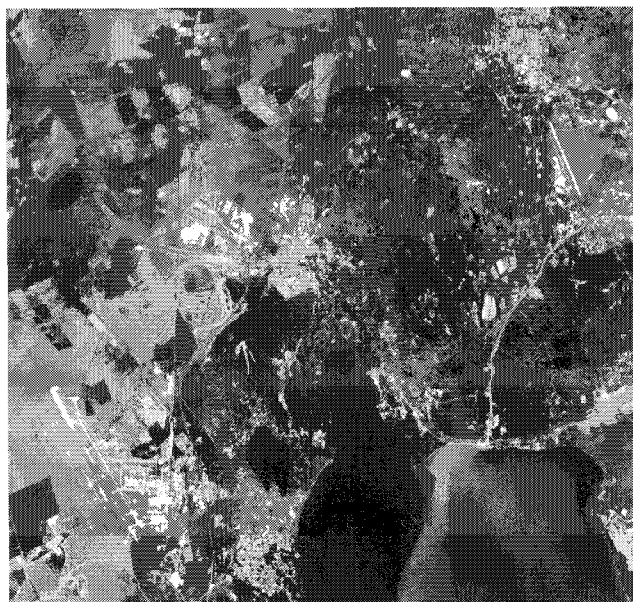


Illustration 6. – Spot Istres 30 m. © CNES 1992 – Distribution Spot IMAGE.

indiquées en fonction de l'importance de l'aérodrome. Les différents paramètres que nous allons décrire en sont directement issus. Nous avons opté pour une représentation calquée sur le modèle des variables linguistiques.

Rappelons simplement qu'une variable linguistique est représentée par un triplet  $(V, X, T)$  où  $V$  désigne la variable prenant ses valeurs dans l'ensemble de référence  $X$ .  $T$  est un ensemble de sous-ensembles flous de  $X$  qui sont utilisés pour caractériser  $V$

[11]. L'avantage de ce formalisme, où interviennent des descriptions imprécises, est d'éviter les bornes artificiellement rigides et de permettre une plus grande souplesse de caractérisation.

### 4.3. CARACTÉRISTIQUES DES INFRASTRUCTURES

Nous avons défini la variable **pen**te :

$V = \text{pen$ te;  $X = [0 \dots 100]$ ;

$T = \{\text{très faible, faible, moyenne, forte}\}$ . Les fonctions d'appartenance des sous-ensembles flous de  $T$  sont définies comme suit :

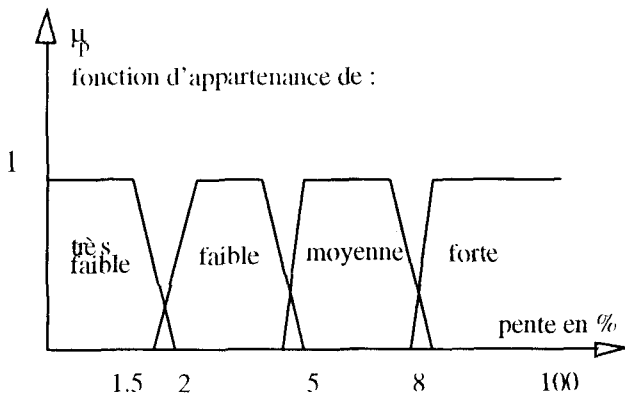


Figure 4. – Fonctions d'appartenance des s.e.f. de pente.

La pente maximale autorisée par l'US NAVY et l'US AIR FORCE est de 1%. Les spécifications des aéroports civils recommandent une pente inférieure à 1.5%, c'est-à-dire très faible selon notre terminologie.

La longueur de la piste principale est donnée suivant l'importance du trafic : de 1200 à 4000 m. Celle des taxiways est de l'ordre de 150 m.

Nous avons défini la variable :

**Longueur** = (longueur,  $R^+$ , {taxiway, piste}) avec les fonctions d'appartenance :

On définit de même la variable :

**Largeur** = (largeur,  $R^+$ , {piste, taxiway})

La référence [10] donne une largeur comprise entre 22 et 45 m pour la piste principale et de 10 à 23 m pour les taxiways.

Les matériaux utilisés pour la construction des pistes sont le bitume ou le béton. L'implantation des bâtiments ne peut se faire qu'à une distance de sécurité de la piste principale. Faute de spécifications concernant la dimension des superstructures, nous avons arbitrairement fixé la taille minimale des hangars à  $100 \times 100$  m et la taille maximale à  $200 \times 500$  m.

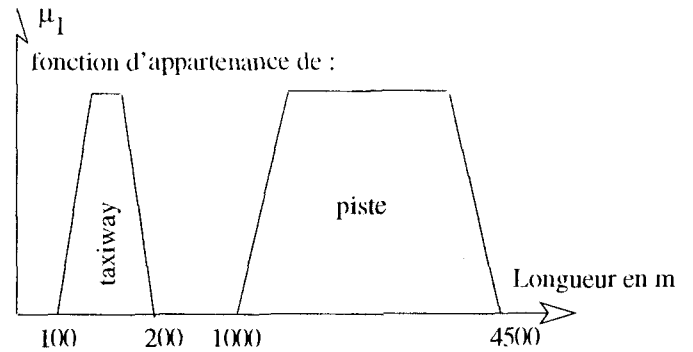


Figure 5. – Fonctions d'appartenance des s.e.f. de Longueur.

### 4.4. ALGORITHME DE DÉTECTION DE LA PISTE PRINCIPALE

La détection de la piste principale s'effectue en 5 étapes successives dans l'image ERS :

- 1 – filtrage multi-critères : détection des morceaux de pistes,
- 2 – étiquetage pixel : recherche des morceaux connexes et suppression des régions de faible surface,
- 3 – vectorisation : détermination de la direction principale pour chaque région
- 4 – reconnexion de vecteurs : détermination des pentes et longueurs des vecteurs, essai de reconnexion des morceaux de pistes disjoints,
- 5 – validation : fondée sur l'intensité, l'hétérogénéité et la longueur des segments candidats.

Le candidat piste est ensuite confirmé ou réfuté par la recherche dans Spot des infrastructures secondaires (taxiways) ou des superstructures d'accompagnement (hangars, tour de contrôle,...). La figure 6 illustre l'enchaînement des différentes phases de l'algorithme.

Les paragraphes suivants sont consacrés à la description de chacune de ces étapes.

### 4.5. FILTRAGE FLOU

C'est la partie la plus sensible de l'algorithme de détection et qui conditionne toute la suite. Nous avons été amenés à définir un filtre prenant en compte **simultanément** des informations de radiométrie, d'homogénéité et de relief. Il s'est avéré qu'une réduction sensible du temps de calcul ne pouvait être obtenue qu'au prix d'une minimisation des traitements « bas niveau ». Il importait donc de limiter au maximum le balayage de l'image et le calcul de caractéristiques attachées aux pixels.

Une piste d'aéroport apparaît sur une image ERS-1 comme **une région très sombre, très homogène et très allongée**. Le filtre flou spatial se propose d'extraire en une seule passe ce type de



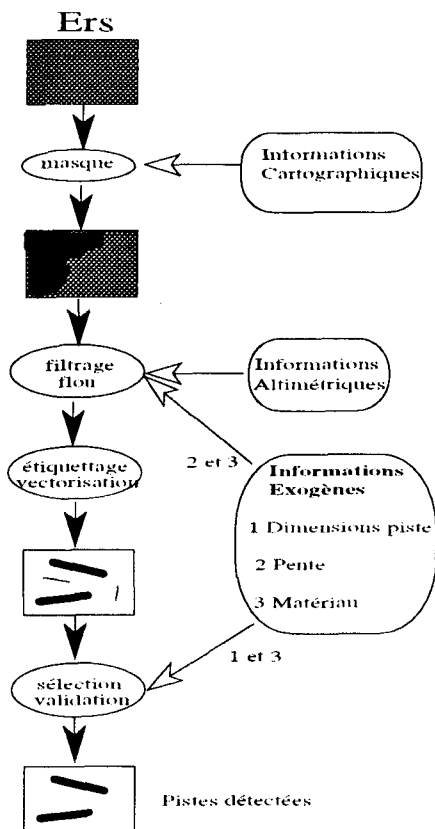


Figure 6. – Principe de la détection des pistes.

segment avec comme condition supplémentaire la satisfaction des contraintes inhérentes à la pente et à l'occupation du sol.

Le filtre extrait en une passe les segments linéaires élémentaires de 3 pixels, susceptibles de constituer une piste. Pour chaque pixel de l'image, un indice donnant sa « possibilité d'être un élément de piste » est calculé à partir de l'image ERS-1 et du modèle numérique de terrain. Les caractéristiques du filtrage sont les suivantes :

**élément structurant** : pavé  $1 \times 3$  (élément allongé)

**intensité** : très faible (Ser bitume très faible)

**hétérogénéité** : très faible (même matériau)

**pente** : très faible (spécifications techniques)

**domaine d'exploration** : 4 orientations principales (horizontale, verticale, oblique avant, oblique arrière) sont systématiquement explorées (Figure 7).

Le choix de la longueur de l'élément structurant résulte d'un compromis entre la largeur des pistes cherchées et le nombre de directions à explorer. En se limitant à 3 pixels, il y a effectivement 12 directions à explorer, dont il s'est avéré suffisant de ne considérer que les 4 principales.

La principale difficulté réside dans la définition de la modalité « très faible » des variables **intensité** et **hétérogénéité**; la modalité

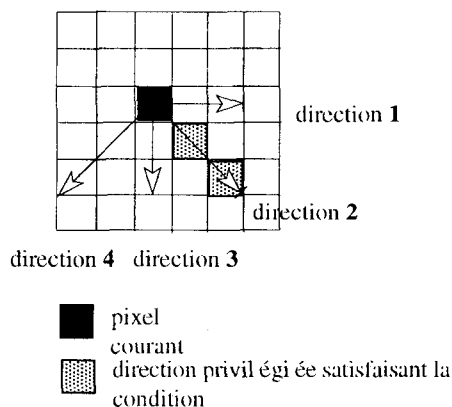


Figure 7. – Domaine d'exploration.

de la variable **pente** précisée dans les spécifications techniques ne pose pas de problème. Nous avons défini les modalités de ces variables en fonction d'un pourcentage de leurs distributions cumulées, et non dans l'absolu faute d'avoir pu trouver un document de référence précisant les Ser en bande C des matériaux « béton » et « bitume » ainsi que leur variabilité en fonction de paramètres tels que l'humidité, ou l'angle d'incidence du faisceau d'éclairage.

Le noyau des modalités (fonctions d'appartenance) pour les variables linguistiques **intensité** et **hétérogénéité** comprend par ordre croissant respectivement 5% (très faible), 15% (faible), 40% (moyenne), 15% (forte) et 1% (très forte) de la distribution de l'image non masquée.

$D$  désigne un pavé directionnel de taille  $1 \times 3$ , et  $i(x, y)$  la valeur du pixel de coordonnées  $x, y$ .

On définit :

$$\text{moy}(D) = \frac{1}{4} \sum_D i(x, y), \text{ moyenne dans la direction } D.$$

$$\text{het}(D) = \frac{1}{4} \sum_D (i(x, y) - \text{moy}(D)), \text{ hétérogénéité dans la direction } D.$$

$$\text{pente}(D) = \frac{|MNT(x, y) - MNT(x_d, y_d)|}{\sqrt{(x-x_d)^2 + (y-y_d)^2}}$$

où  $(x_d, y_d)$  désigne le pixel le plus éloigné dans la direction  $D$  et  $MNT(x, y)$  l'élévation du pixel  $(x, y)$ .

La direction  $D$  étiquetée est celle correspondant à la valeur maximum des 4 directions :

$$\max_{D=1,4} [\min(\mu_A(\text{moy}(D)), \mu_A(\text{het}(D)), \mu_P(\text{pente}(D)))]$$

Une décision est prise à l'aide de l'heuristique floue suivante :

Si ( $\text{moy}(D)$  très faible ET  $\text{het}(D)$  très faible ET  $\text{pente}(D)$  très faible) Alors  $D$  est un morceau de piste,

qui n'est appliquée que si le degré de réalisation des prémisses, mesuré par :

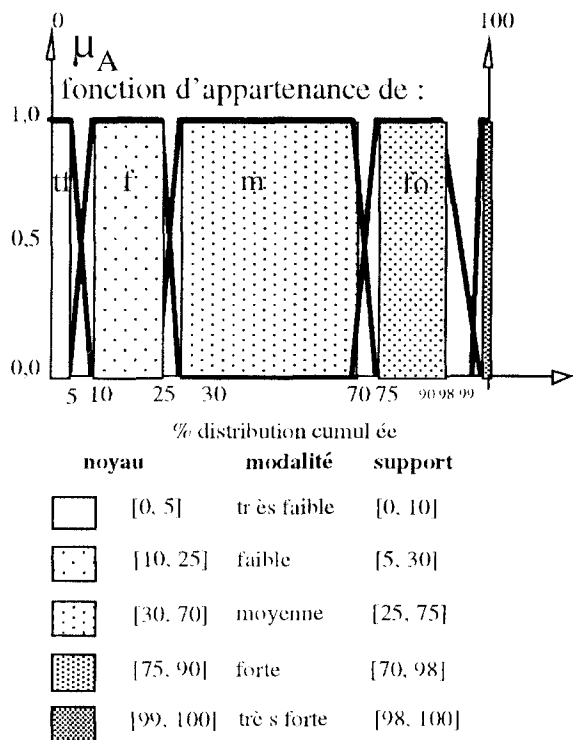


Figure 8. – Définition des modalités.

$\Psi(\mu_{Atf}moy(D), \mu_{Atf}het(D), \mu_{Plf}pente(D))$ , où  $\Psi$  désigne un opérateur d'agrégation, est supérieur à un seuil arbitraire que l'on a fixé à 0.5.

Nous avons retenu l'opérateur d'agrégation MIN, entre autres pour sa simplicité. Il en existe une multitude d'autres [12] qui ne modifient pas ici les résultats de manière sensible. Un comportement plus ou moins sélectif peut être obtenu en jouant sur la définition des modalités.

#### 4.6. VECTORISATION PAR TRANSFORMÉE DE HOUGH

Les régions 8-connexes obtenues par étiquetage doivent être transformées en segments de droites. La polygonalisation par l'algorithme de la corde ne s'est pas avérée adaptée. En effet, ce traitement est trop local et dépend fortement du point d'accès et de la tolérance sur la distance à la corde. Les vecteurs obtenus étaient beaucoup trop nombreux, trop courts et trop imprécis, quelle que soit la distance à la corde tolérée. La reconnexion des segments devenait ensuite trop difficile à gérer compte tenu de leur faible longueur, et de l'imprécision sur leur pente. Nous avons adopté une approche plus globale en limitant le nombre de droites résultats à une par label. La transformée de Hough accumule dans un espace judicieusement choisi l'ensemble des droites passant par les points précédemment étiquetés.

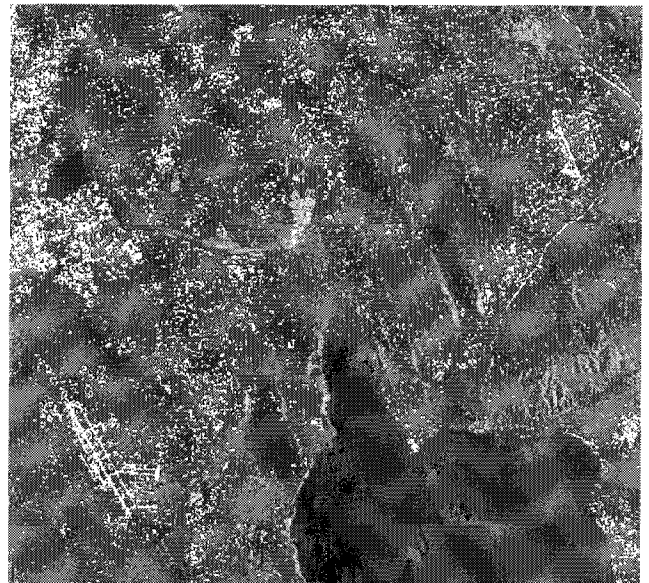


Illustration 7. – Résultat du filtre flou sur ERS.

#### 4.7. RECONNEXION

La continuité de la détection n'est pas toujours assurée par le filtrage. Il peut s'avérer nécessaire de raccorder des segments candidats provenant d'une seule piste, fragmentés lors des phases précédentes. Le raccordement se fait sous les trois contraintes de :

- proximité : les deux segments candidats doivent être spatialement proches. La distance  $\delta$  (Figure 9) séparant les deux extrémités les plus proches des deux segments ne doit pas excéder 10% de la longueur du segment le plus court.
- directions identiques : les deux segments candidats censés constituer une piste doivent pouvoir être raccordés en conservant une direction commune. Suite à l'imprécision cumulée du filtrage et de l'orientation calculée par la transformée de Hough, une tolérance angulaire de  $10^\circ$  nous a semblé être un bon compromis.

$$|\theta_1 - \theta_2| < 10^\circ$$

- afin de ne pas raccorder deux segments parallèles et voisins, le segment résultat doit avoir une longueur supérieure à la plus grande des deux longueurs des segments candidats.

#### 4.8. VALIDATION

Il s'agit ensuite de sélectionner parmi l'ensemble des segments candidats (éventuellement raccordés) ceux satisfaisants les conditions de définition d'une piste, à savoir un élément linéaire dont les dimensions sont comprises entre 1400 m et 4000 m, possédant

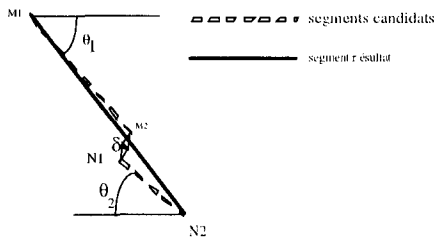


Figure 9. – Raccordement de 2 segments.

une SER et une hétérogénéité très faibles. Cette décision sera binaire (piste ou non piste) et compte tenu des différentes approximations effectuées (transformée de Hough et raccordement), elle portera à nouveau sur le support des variables **longueur**, **intensité** et **hétérogénéité**. Pour un segment  $S$ , cette condition se traduit par :

$$1400 < L(S) < 4000 \text{ m ET } I(S) < \delta(I) \text{ ET } H(S) < \delta(H)$$

Les valeurs de  $\delta(I)$  et  $\delta(H)$  correspondent aux valeurs maximum de la modalité très faible des variables **intensité** et **hétérogénéité**.

#### 4.9. DÉTECTION DES TAXIWAYS

Afin de d'augmenter le crédit de la détection, nous avons cherché à détecter les structures annexes, telles que les taxiways ou les bâtiments, qui sont des infrastructures obligatoires pour des aéroports. La Figure 10 décrit les étapes qui confirment ou infirment la détection.

La restriction spatiale permet de limiter la zone de recherche aux abords de la piste pour les taxiways, et dans un voisinage défini par des distances minimum et maximum autour de la piste pour les bâtiments.

En limite de résolution du capteur radar, les taxiways ne se distinguent pas toujours clairement sur l'image Ers-1. Par contre, le capteur Spot de meilleure résolution spatiale permet de les détecter.

Nous avons constaté sur plusieurs images SPOT que les taxiways se distinguent de leur environnement d'intensité moyenne par une luminosité forte si le revêtement est du béton, et faible si le revêtement est du bitume. Des opérateurs de morphologie mathématique permettent de mettre en évidence des éléments clairs sur fond sombre et vice versa. Il s'agit respectivement du top-hat (image originale moins dilatation de l'érodée) et du bottom-hat (image originale moins érodée de la dilatation), que l'on utilise ici avec des éléments structurants de taille  $3 \times 3$ .

Seules 3 orientations pistes / taxiways sont couramment utilisées ( $30^\circ$ ,  $45^\circ$  et  $90^\circ$ ). Dans notre cas, les taxiways étaient principalement à angle droit avec la piste principale, nous avons donc réduit la recherche à ce seul cas, sachant que son extension aux deux

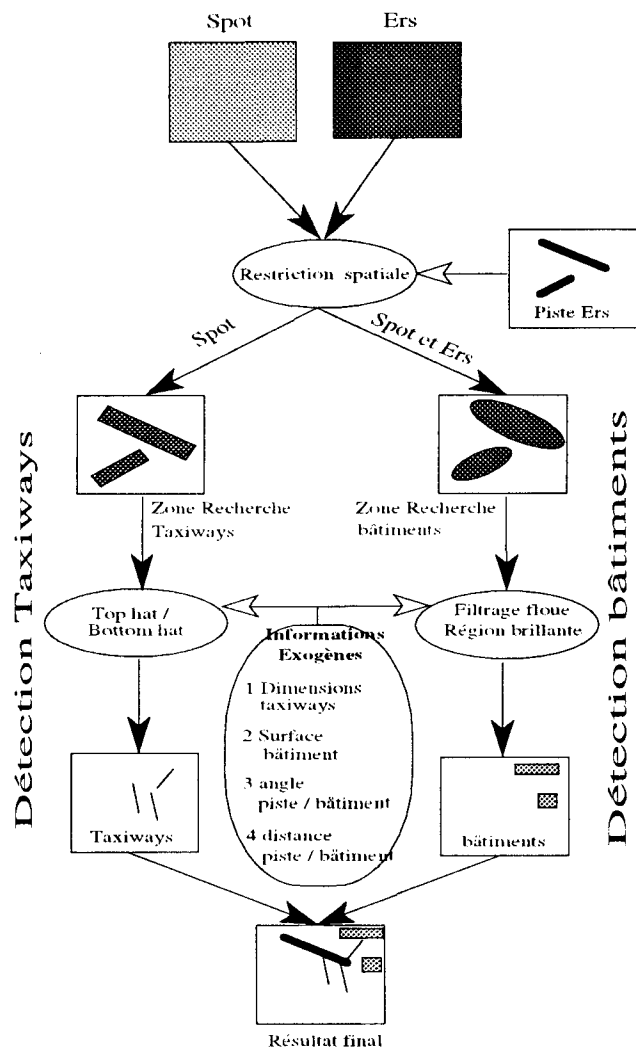


Figure 10. – Détection des structures annexes.

autres orientations ne pose aucun problème (si ce n'est de temps calcul...). Nous avons retenu :

- une orientation  $\theta$  de  $90^\circ$
- une longueur usuelle de taxiway  $D$  de 150 m

L'algorithme appliqué au résultat du top hat (resp bottom hat), pour lequel seuls les pixels extraits sont non nuls, est le suivant :

- 1 – Pour chaque pixel  $P$  de la piste, détermination de l'extrémité potentielle  $T$  d'un taxiway, à la distance  $D$  de  $P$  dans la direction  $\theta$ .
- 2 – Pour chaque segment  $PT$ , si le nombre  $N$  de pixel de valeur supérieure à 1 est plus grand que le rapport, où  $R$  désigne le pas de quantification de l'image, on décide de la présence d'un taxiway. La connexité est ici implicite du fait de la finesse des taxiways.

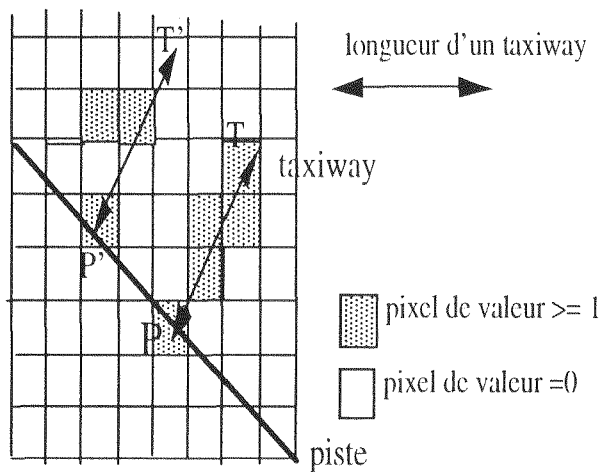


Figure 11. – Recherche des taxiways.

Dans la figure 11, le segment  $PT$  est un taxiway, il recouvre des pixels non nuls du top-hat. Sur les images dont nous disposons, les taxiways ne sont détectés qu'avec une fiabilité très moyenne, en particulier à cause d'un taux de fausses alarmes élevé.

#### 4.10. DÉTECTION DES SUPERSTRUCTURES

En guise d'ultime confirmation nous avons cherché les superstructures (bâtiments, hangar...) qui accompagnent généralement les aéroports. Aussi bien sur l'image ERS que sur l'image Spot, elles apparaissent comme des formes approximativement rectangulaires très claires par rapport au reste de l'image. Nous avons utilisé le même principe que pour la détection des pistes avec la règle d'inférence :

$SI$  ( $\text{moy}_{ERS}(M)$  très forte  $OU$   $\text{moy}_{Spot}(M)$  très forte)

ALORS  $M$  est un morceau de bâtiment où :

$\text{moy}_{ERS}(M)$  désigne la moyenne sur un pavé  $2 \times 2$  de la radiométrie dans l'image radar.

$\text{moy}_{Spot}(M)$  est la moyenne sur un pavé  $2 \times 2$  de la radiométrie dans l'image Spot.

La conclusion est inférée lorsque le degré de réalisation des prémisses est supérieur à un seuil fixé a priori, soit :

$$\text{MAX}(\mu_{\text{Atrès fort}}(\text{moy}_{ERS}(M)), \mu_{\text{Atrès fort}}(\text{moy}_{Spot}(M))) > 0.5$$

Cette opération n'est bien sûr effectuée qu'à proximité des pistes détectés et en respect des distances de sécurité. Le rayon de recherche a arbitrairement été fixé à 1000 m.

Nous n'avons pas tenu compte dans notre algorithmie de la régularité de leurs contours. Il serait donc envisageable de com-

plexifier cette méthode en intégrant de nouvelles restrictions sur la base d'informations exogènes.

## 5. Conclusion

Cet article illustre trois aspects de la fusion d'images. Au niveau pixel tout d'abord, avec la création d'images composites, à un niveau plus symbolique avec une fusion de décision pour un problème de classification, et enfin à un troisième niveau où des informations de natures très différentes sont utilisées en synergie. C'est cette dernière approche qui nous semble la plus prometteuse. Elle nous a permis de détecter des aéroports sans fausses alarmes sur cinq sites différents. Le résultat de la détection sur Istres est présenté ci-après (Illustration 8).



Illustration 8. – Résultat de la détection (Istres).

Bien que l'échantillon sur lequel a été effectuée cette étude ne soit pas représentatif de l'ensemble des aéroports implantés dans le monde, le modèle adopté est suffisamment simple et générique pour pouvoir être efficace. Ce type de démarche, faisant coopérer des informations de nature différente, a montré son efficacité en termes de temps de calcul : le traitement d'une image  $1000 \times 1000$  nécessite environ de 2 mn sur une station SPARC 10. Nous sommes actuellement en train d'étendre cette méthodologie à un système de détection multicapteurs s'adaptant aux données disponibles afin d'en tirer le meilleur parti, sans qu'aucune source d'information ne constitue un impératif pour le déroulement de la détection, qui ne devrait plus exclusivement concerner les aéroports.

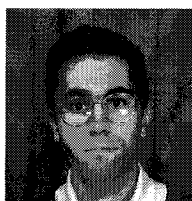
## BIBLIOGRAPHIE

- [1] R. Welch and M. Ethlers, « Cartographic feature extraction with integrated SIR-B and Landsat TM images », in *International Journal of Remote Sensing*, vol 9, n°5, 1988.
- [2] J. Harris and R. Murray, « IHS transform for the integration of radar imagery with other remotely sensed data », in *Photogrammetric Engineering and Remote Sensing*, vol 56, December 1990.
- [3] Y. Yéso and al., « Intégration des données satellitaires Seasat et Spot, contribution à la cartographie géologique structurale en milieu tempéré », in *Physical Measurements and Signatures in Remote Sensing*, Courchevel January 1991.
- [4] A. Michel, P. Cornic, « Fusion d'images radar et optique en télédétection », *14ème Colloque GRETSI*, Septembre 1993.
- [5] S. Mallat, « A theory for multiresolution signal decomposition : the wavelet representation », in *IEEE transactions on pattern analysis and machine intelligence*, vol 11, n°7, July 1989.
- [6] I. Debauchies, « Orthogonal bases of compactly supported wavelets », in *Communications on pure and applied mathematics*, Vol XLI, pp. 909-996, 1988.
- [7] L. Benquet, « Extraction automatique des aéroports dans les images Spot », *14ème Colloque GRETSI*, Septembre 1993.
- [8] L. Benquet, T. Quiguer, « Airport Recognition using contextual information », in *EUSIPCO 94*, à paraître.
- [9] S. Houzelle, G. Giraudon, « Automatic Feature Extraction Using Data Fusion in Remote Sensing », in *Proceeding of SPIE*, Boston December 1991.
- [10] N. Ashford, P.H. Wright, « Airport engineering », WileyIntersciens.
- [11] E. Bouchon - Menier, « La logique floue », in *Que sais-je?* PUF 1993.
- [12] I. Bloch, « Information combinaison operators for data fusion : a comparative review with classification », *Rapport ENST 1994*.
- [13] V. Clement , G. Giraudon, S. Houzelle, « A knowledge-based Interpretation system for Fusion of Sar and Spot Images », in *IEEE 1992*.
- [14] R. Haralick and alt, « Textural features for image classification », in *IEEE Transaction on systems, Man, and Cybernetics*, Vol 3 no6 November 1973.

Manuscrit reçu le 30 Mai 1994.

## LES AUTEURS

### Philippe CORNIC



Philippe Cornic est ingénieur ENSIEG (traitement du signal, 1989). Au sein de la Direction des Etudes de Synthèse de l'ONERA depuis 1992, il s'occupe plus particulièrement de fusion en imagerie satellitale.

### Alain MICHEL



Alain Michel est titulaire d'une thèse de doctorat (EHESS 1988). Son principal domaine d'intérêt est le développement d'outils de PIAO (Photo-Interpretation Assisté par Ordinateur).