

SEGMENTATION DES SCENES EN POURSUITE TELEVISION
SCENE SEGMENTATION FOR T.V. TRACKING

Robert NEGRE - Roger SAMY

CETIA - CSEE - Av. des Frères lumière Z.I. Ste Claire - 83160 LA VALETTE DU VAR

RESUME

On présente dans une première partie les différents algorithmes de segmentation des scènes de poursuite.

Le caractère temps réel de la poursuite nous a conduit à effectuer une sélection parmi ces algorithmes.

On décrit, dans une deuxième partie, l'architecture d'un système de poursuite multimode.

SUMMARY

First, we present some image segmentation algorithms for target tracking.

We select well suited algorithms for designing a realtime video tracker.

And then, a multiprocessor architecture is described for the video tracker.



INTRODUCTION ET GENERALITES

La segmentation de scènes est une opération fondamentale dans la reconnaissance et l'interprétation d'images ; le but de cette opération est d'effectuer une partition de l'image en régions qui correspondent à des entités ou des objets caractérisant de la scène [FAU1].

La segmentation fait souvent l'hypothèse que l'objet recherché présente une homogénéité dans les attributs de niveau de gris, couleur ou texture. Cette hypothèse a pour conséquence deux approches complémentaires : le groupement (fusion) de petites régions qui présentent les caractéristiques identiques (croissance de régions) et la détection de frontières où un ou plusieurs attributs changent brusquement (détection de contours).

La segmentation des scènes en poursuite télévision est destinée à extraire une ou plusieurs cibles présentant non seulement les caractéristiques précédentes, mais également certaines propriétés spécifiques.

En effet, en poursuite on traite des séquences d'images pour lesquelles on connaît grossièrement les positions des cibles [Fig 1] ; la segmentation permet d'extraire la position des cibles afin de pouvoir asservir la caméra sur la position des objets poursuivis [HUAL].

Il arrive que des obstacles soient présents dans le champs de vision de caméra. La segmentation d'image peut alors fournir des fausses alarmes. On demandera donc à l'algorithme correspondant d'être robuste ou de signaler rapidement qu'on est en présence d'obstacles. Par ailleurs la conception d'un système de poursuite à partir d'une caméra de télévision suppose que le traitement puisse être effectué en temps réel (à la cadence de signal vidéo : 25 images/seconde).

Cette contrainte de temps a des conséquences sur l'algorithme de segmentation. En effet, l'algorithme doit être peu complexe pour être intégrable sur une structure cablée ou implémentée sur un microprocesseur de traitement de signal (microprocesseur en tranche).

On présente un certain nombre d'algorithmes qui satisfont à ces contraintes et on décrit une réalisation matérielle d'un écartomètre télévision.

I - SEGMENTATION DES SCENES DE POURSUITE

La première étape dans le traitement des images en poursuite télévision réside dans la réduction du volume d'information à traiter afin de satisfaire les exigences liées aux capacités réelles de traitement et de mémorisation du système.

Cette réduction est réalisée par l'extraction d'attributs. Les attributs utilisés en poursuite sont les histogrammes d'intensités et les contours (contraste). En effet, les histogrammes se prêtent bien au problème de traitement en temps réel, car on passe de $N \times N$ points (N étant la taille de la fenêtre à traiter) à 256 (codage sur 8 bits) valeurs à traiter et de plus le calcul d'un histogramme repose sur l'incrémentement d'un compteur.

Il existe en toute rigueur une deuxième solution qui consiste à effectuer de l'extraction de contours. Cette solution séduisante a priori est difficile à réaliser avec la technologie disponible. On peut cependant imaginer de le faire en développant des VLSI spécialisés. Les algorithmes présentés dans la suite mettent en oeuvre des histogrammes d'intensité. Ceci suppose que la cible et le fond ont des histogrammes différents. Cette hypothèse n'est pas vérifiée si, pour exemple, le fond est texturé ; on doit alors envisager d'autres méthodes. Une solution à ce problème réside dans le calcul de trois histogrammes qui correspondent à l'intensité, au gradient de l'intensité et à une composante de texture.

II - SEGMENTATION PAR ANALYSE STATISTIQUE DES INTENSITES

On présente un algorithme d'extraction de cible basé sur l'analyse discriminante linéaire des intensités [GIL]. En effet, les intensités de la cible sont différentes de celles du fond immédiat.

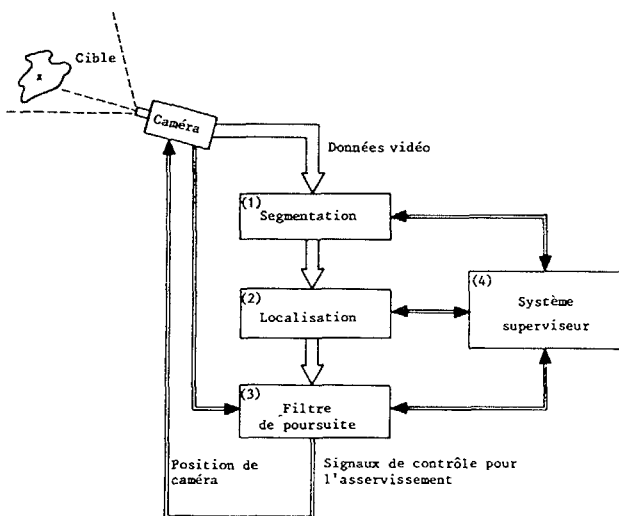


Figure - Système de poursuite vidéo

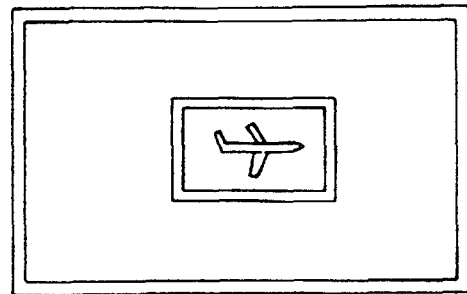


Figure - Fenêtre de poursuite

On effectue donc une partition de la fenêtre de poursuite en deux régions BR et TR ; on suppose que la région BR ne contient que les points du fond et que la région TR contient des points de la cible et du fond.

On dresse alors les histogrammes en intensité de ces deux régions BR et TR ; ces histogrammes sont normalisés pour qu'ils soient indépendants des dimensions des différentes régions.

On peut alors écrire les relations suivantes :

$$\forall x \in [0, 255] \left\{ \begin{array}{l} h^{BR}(x) = h^B(x) \quad (1) \\ h^{TR}(x) = \alpha h^B(x) + (1 - \alpha)h^T(x) \quad (2) \end{array} \right.$$

$$0 \leq \alpha \leq 1$$



SEGMENTATION DES SCENES EN POURSUITE TELEVISION
SCENE SEGMENTATION FOR T.V. TRACKING
Robert NEGRE - Roger SAMY

Il existe plusieurs méthodes pour estimer le coefficient de mélange :

L'équation (2) permet d'écrire :

$$\forall X \in [0, 255] \quad h^T(X) = \frac{h^{TR}(X) - \alpha h^B(X)}{1 - \alpha}$$

($\alpha \neq 1$)

$h^T(X)$ étant une fonction positive, on déduit une relation pour :

$$h^T(X) \geq 0 \implies \begin{cases} \forall X \in [0, 255] \\ h^B(X) \neq 0 \end{cases} \quad \alpha \leq \frac{h^{TR}(X)}{h^B(X)}$$

d'où :

$$X \in [0, 255] \quad \alpha = \text{Inf} \left[\frac{h^{TR}(X)}{h^B(X)} \right]$$

$h^B(X) \neq 0$

Comme il s'agit d'une séquence d'images, on peut faire un filtrage temporel de ces histogrammes $h^T(X)$ et $h^B(X)$.

Ainsi deux filtres temporels d'ordre deux peuvent être utilisés :

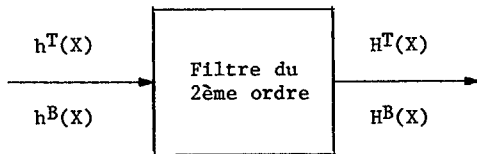


Figure - Filtrage temporel des histogrammes

$$H(i, i) = wH(i, i-1) = (1 - w) h_i(X)$$

$$H(i+1, i) = 2H(i, i) - H(i-1, i-1)$$

où i désigne le numéro de l'échantillon sur l'axe temporel et H désigne les valeurs filtrées des histogrammes.

Comme on effectue la poursuite de la cible le coefficient d'apprentissage w de la cible est différent de celui du fond.

III - SEGMENTATION D'IMAGES PAR CLASSIFICATION FLOUE

On présente ici un algorithme de segmentation d'image basé sur la théorie des ensembles flous [KAU1].

En effet, dans des situations réelles il arrive que l'algorithme de segmentation basé sur la classification bayésienne donne des résultats erronés.

Dans le but d'améliorer la segmentation, on a introduit [INI1] la classification floue.

Cette méthode s'apparente aux problèmes de linguistique et est plus précisément liée à la notion de nuance. En effet, on introduit quatre niveaux :

- 1^2 = très vraisemblable
- 1 = vraisemblable
- ul = invraisemblable
- ul^2 = très invraisemblable

La règle de décision est alors :

$$BP = l_B^2 + l_B + l_B ul_B + ul_B (\bar{l}_T + \bar{l}_B^2)$$

$$TP = (l_T + l_T^2) (ul_B + ul_B^2) + ul_B^2$$

(B : fond et T : cible)

$$\text{On prend : } \begin{cases} l_B(i) = H_i^B(X) = h_i^B(X) \\ l_T(i) = H_i^T(X) = h_i^T(X) \end{cases}$$

Si on suppose que les lois de l'algèbre booléenne peuvent s'appliquer aux ensembles flous, on obtient alors :

$$BP = l_B (1 + ul_B) + (ul_B) l_T^2$$

$$TP = l_T ul_B + (ul_B)^2$$

La règle de décision reste identique à celle de la classification bayésienne :

$$\begin{aligned} \text{si } BP > TP &\implies \text{FOND} \\ \text{si } TP > BP &\implies \text{CIBLE} \end{aligned}$$

L'évaluation de cet algorithme montre que la classification floue permet de résoudre certains problèmes de l'algorithme de segmentation précédent, mais la complexité du calcul ne justifie pas encore son emploi dans un système temps réel.

IV - ARCHITECTURE D'UN SYSTEME DE POURSUITE DE CIBLES EN TEMPS REEL

1 - Généralités

On présente l'architecture d'un système de poursuite à partir d'une caméra de télévision. Ce système de poursuite (écartomètre télévision) présente la particularité d'être multimode (multi-algorithmes).

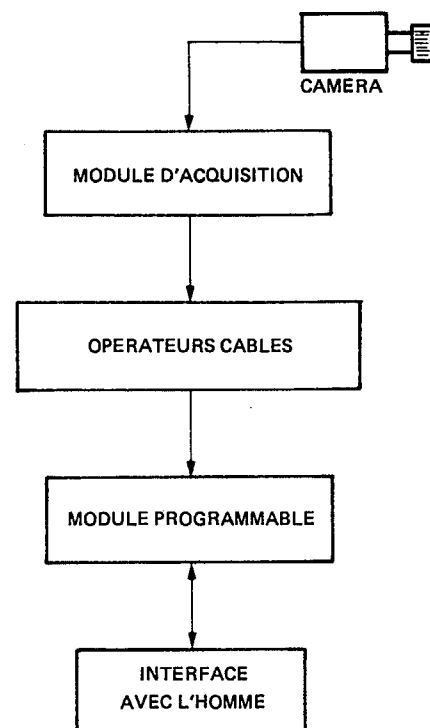


Figure - Architecture de l'écartométrie télévision multimode.



SEGMENTATION DES SCENES EN POURSUITE TELEVISION
SCENE SEGMENTATION FOR T.V. TRACKING
Robert NEGRE - Roger SAMY

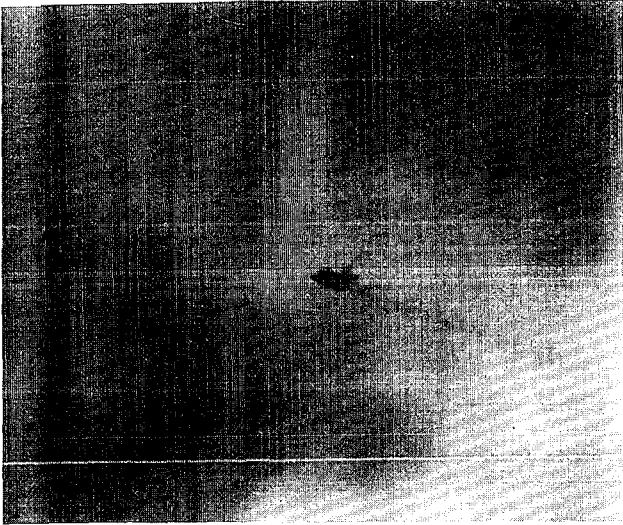


Photo 1 - Image originale n° 1

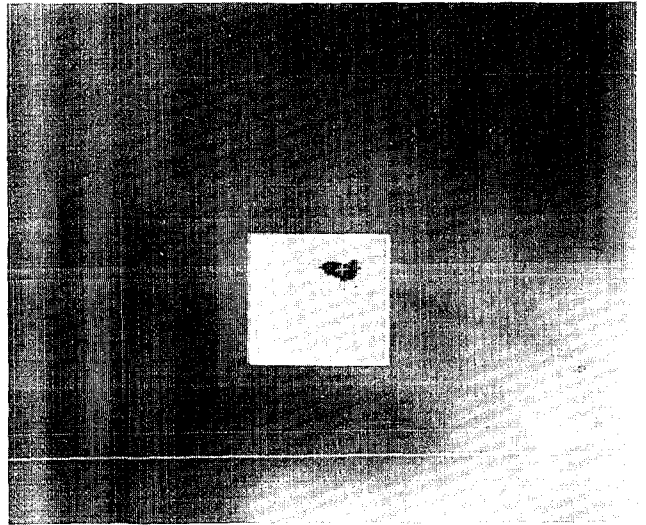


Photo 4 - Image segmentée n° 1



Photo 2 - Image originale n° 2



Photo 5 - Image segmentée n° 2



Photo 3 - Image originale n° 3

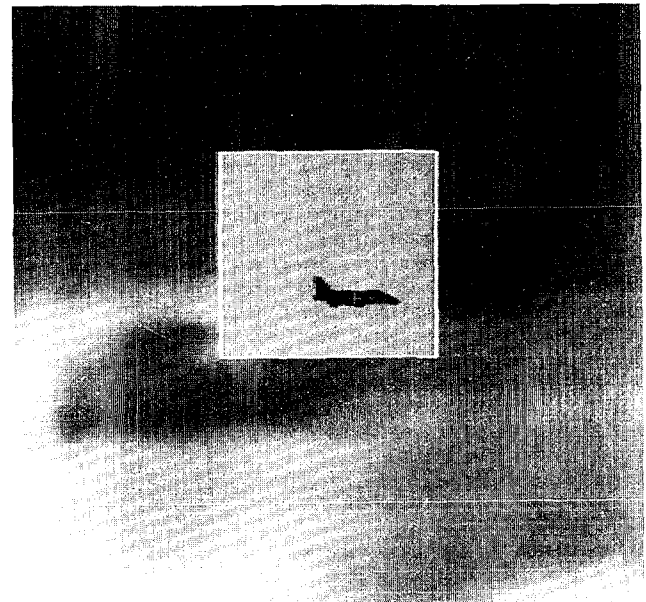


Photo 6 - Image segmentée n° 3



SEGMENTATION DES SCENES EN POURSUITE TELEVISION
SCENE SEGMENTATION FOR T.V. TRACKING
Robert NEGRE - Roger SAMY

Les images traitées par ce système sont issues d'une caméra qui est au standard vidéo CCIR à 625 lignes avec un balayage de 25 images par seconde (50 trames/sec) et une bande spectrale d'environ 5MHz (pour conserver le rapport S/B de 45 dB).

Ce signal vidéo est numérisé et codé sur 8 bits (octect) (256 niveaux de gris). On obtient un débit de 10 M octects ; une image étant représentée par 256 K octects.

Le débit résultant étant de 6,5 M octect/seconde est trop élevé vis à vis des capacités de traitement d'une machine classique programmable.

Certains algorithmes de traitement d'images nécessitent en effet plusieurs dizaines d'opérations par point ou "pixel", soit une cadence de plusieurs dizaines de millions d'opérations par seconde.

Il est donc nécessaire de réduire le débit de traitement et de choisir les architectures permettant de tenir une cadence élevée d'opérations. Nous nous proposons de présenter une architecture adaptée à ce problème.

2 - Opérateurs câblés

2.1 - Opérateur d'histogrammes

Cet opérateur reçoit la vidéo numérisée et on calcule des histogrammes sur les fenêtres dont les positions et dimensions sont fournies par un contrôleur externe.

Cet opérateur câblé permet, de plus, d'effectuer le filtrage temporel des histogrammes [SCH1].

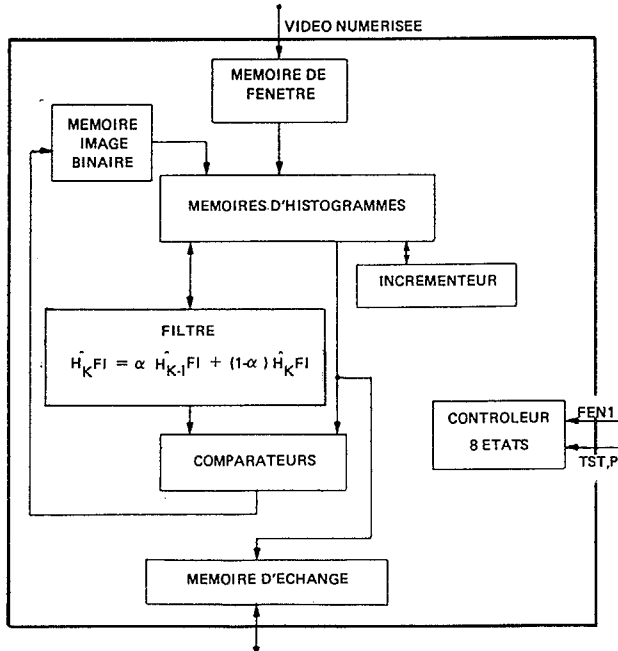


Figure - Architecture de l'opérateur câblé d'histogrammes

L'image binaire résultant de l'opérateur de segmentation est traitée pour "binariser" la cible. La localisation de cible peut être faite par un calcul de projections verticale et horizontale (i.e. barycentre) ou par une méthode de corrélation binaire ; la corrélation binaire étant plus robuste que les projections en présence d'obstacles [MUR1].

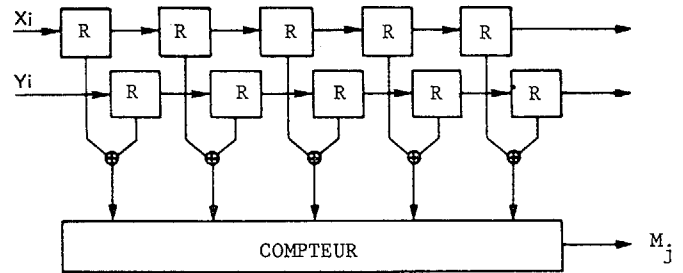


Figure - Circuit de corrélation binaire

(R : Registre à décalage de 1 bit, ⊕ : ou exclusif)

2.2 - Opérateur de gradient

Quand la taille apparente de la cible est faible les histogrammes n'ont plus de valeur statistique en raison du nombre peu élevé de pixels traités. On peut alors utiliser le critère de contraste pour détecter la cible.

En effet, on calcule le gradient de l'image par les masques de Prewitt. On effectue la convolution de de l'image par les deux masques 3 x 3 suivant :

$$M_1 = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \quad M_2 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

$$G(j,k) = |F(j,k) \otimes M_1| + |F(j,k) \otimes M_2|$$

Une recherche de maximums locaux est faite sur une fenêtre plus large. On obtient alors une représentation synthétique de la scène dans laquelle on discrimine cible, fond et obstacles.

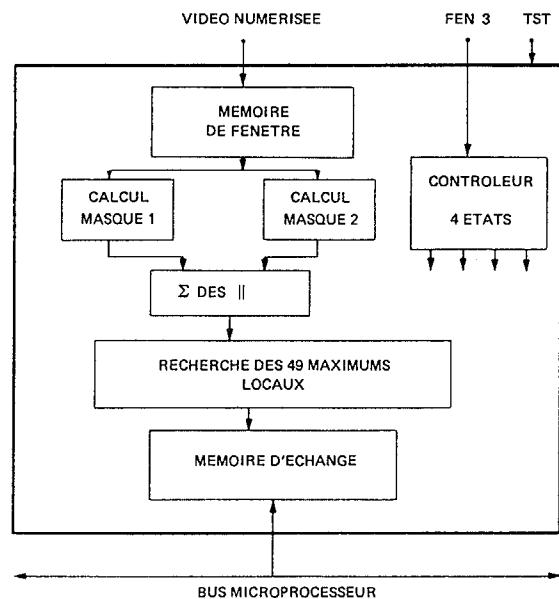


Figure - Architecture de l'opérateur câblé de gradient



SEGMENTATION DES SCENES EN POURSUITE TELEVISION
SCENE SEGMENTATION FOR T.V. TRACKING
Robert NEGRE - Roger SAMY

3 - Module programmable

La souplesse d'exploitation de l'écartomètre est obtenue grâce à un microprocesseur qui est chargé de recueillir les résultats des opérateurs cablés. Il caractérise ces résultats pour "donner une note" correspondant à un degré de confiance.

* Les résultats de l'opérateur de gradient sont traités pour extraire la cible du fond et la discriminer des obstacles.

* La fonction d'intercorrélacion est caractérisée par recherche de pic et calculs de moyennes et variances spatiales destinées à déterminer si la corrélation a un sens.

* Les résultats de l'opérateur d'histogramme sont traités de manière à déterminer les fausses détections :

- Test de la répartition de l'histogramme cible.
- Test des projections verticales et horizontales.
- Test de la continuité temporelle du nombre de points de la cible détectés.

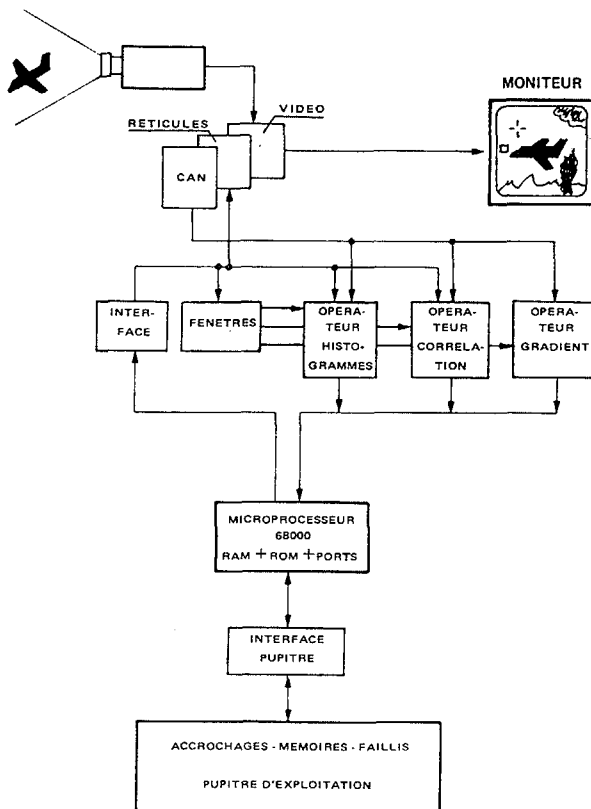
4 - Architecture proposée

La structure proposée est hiérarchisée modulaire et synchrone.

Hiérarchisation - de type vertical : Rapidité de traitement au sommet, souplesse et intelligence en bas.

Modulante - Parallélisme des opérateurs - Parallélisme des modules logiciels.

Synchronisme - L'ensemble de la logique numérique est synchrone et de type "pipe-line".



V - CONCLUSION

Cette architecture a été réalisée sous forme de maquette dans les laboratoires d'études avancées de CSEE et a permis des poursuites dans des milieux très complexes en présence de nombreux obstacles. D'autres opérateurs sont actuellement à l'étude pour accroître la puissance de l'écartomètre. Les études futures s'orientent vers une intégration des fonctions en circuits prédifusés CMOS de manière à diviser par deux le volume de l'appareil qui représente à l'heure actuelle environ 15 cartes au format 1/2 ATR.

REMERCIEMENT

Nous remercions Mr Claude A. BOZZO, Directeur scientifique de la CSEE, pour avoir dirigé cette étude et pour les conseils et remarques fructueuses qu'il nous prodigués lors de la rédaction de cet article.

BIBLIOGRAPHIE

- [FAU1] O.D. FAUGERAS and M. BERTHOD
"Improving consistency and reducing ambiguity in stochastic labelling : An optimisation approach"
IEEE - Trans. PAMI Vol. 2 - N° 1 1980
- [GIL1] A.L. GILBERT et al.
"A Real time video tracking system"
IEEE - Trans. PAMI Vol. 2 N° 1 - 1980
- [HUA1] T.S. HUANG
"Image sequence analysis"
Springer - Verlag - 1981
- [INI1] R.M. INIGO, M. HINKEY and C.RUEST
"The use of fuzzy set theory image in segmentation"
Symposium IFAC - Marseille - 1983
- [KAU1] A. KAUFMAN
"Le traitement mathématique de l'imprécis par la théorie des sous-ensembles flous"
- [MUR1] H. MURAKAMI - V. KUMAR
"Correlation of binarized images"
IEEE - Trans. AES - 19 - n° 2 - 1983
- [SCH1] T. SCHLUMPF
Conception d'un histogrammeur cablé pour écartomètre télévision - E.N.S.P - 1983

