

# Détection d’objets ponctuels sur fond de clutter

Vincent SAMSON<sup>1</sup>, Frédéric CHAMPAGNAT<sup>1</sup>, Jean-François GIOVANNELLI<sup>2</sup>

<sup>1</sup>Office National d’Études et de Recherches Aérospatiales  
29, avenue de la division Leclerc, 92322 Châtillon Cedex, France

<sup>2</sup>Laboratoire des Signaux et Systèmes  
Supélec, Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France  
samson@onera.fr, fchamp@onera.fr, giova@lss.supelec.fr

**Résumé** – Nous nous intéressons à la détection de cibles ponctuelles sur fond nuageux dans une séquence d’images en imagerie satellitaire. Le problème est abordé sous l’angle de l’estimation de deux composantes, objets ponctuels d’un côté et fond structuré de l’autre, qu’il s’agit de séparer. À partir d’un modèle d’observation *mono-image*, la solution est obtenue selon une approche régularisée en minimisant un critère composite convexe et robuste. Nous proposons également un traitement conjoint sur deux images successives de la séquence qui prend en compte le mouvement apparent de translation du fond, et qui permet de détecter à la fois des cibles se déplaçant rapidement par rapport au fond et des cibles lentes.

**Abstract** – This paper deals with point target detection in a cloud background for remote surveillance in satellite imaging. Our goal is to separate the two components of the data, cluttered background on the one hand and isolated point targets on the other hand. The single frame solution is defined in a regularized framework as the minimizer of a penalized robust criterion, which is chosen convex. An extension to the simultaneous processing of several frames, taking into account the apparent translational motion of the background, is also proposed. Simulation results show improved performance for detecting both fast and slow moving objects.

## 1 Introduction

Nous nous intéressons au problème de la détection de cibles ponctuelles dans une séquence d’images. L’application visée est le dimensionnement d’un système de surveillance du ciel par imagerie satellitaire, mais cela constitue également un sujet d’intérêt dans de nombreux autres domaines, notamment en imagerie astronomique ou biomédicale. Nous abordons ce problème comme celui de la *suppression du fond* ou du *rehaussement des cibles* : notre but est d’extraire des objets ponctuels isolés, observables sous la forme de « points brillants », du bruit de fond structuré (ou « clutter ») constitué par le ciel nuageux. Il s’agit donc plutôt d’un problème d’estimation, l’aspect décisionnel étant reporté à une étape ultérieure. De fait, nous nous limitons pour l’instant au cadre de l’optimisation *convexe*, et nous effectuons ensuite un simple seuillage sur l’image estimée des cibles pour obtenir la carte des détections.

La détection d’objets faiblement résolus a donné lieu à de nombreux travaux lors des vingt dernières années. Les techniques classiques utilisent des méthodes de prédiction linéaire du fond corrélé, éventuellement adaptatives, et identifient les cibles ponctuelles comme des anomalies par rapport à la structure du fond [1, 2]. Cependant, l’intensité des cibles est souvent trop faible pour pouvoir séparer suffisamment les deux composantes à partir d’une seule image. De nombreuses approches dites « track before detect » ont alors recours à des techniques de filtrage spatio-temporel pour intégrer l’énergie des cibles sur plusieurs images, sous diverses hypothèses de mouvement [3, 4].

Dans cette communication comme dans [5], nous propo-

sons une approche *régularisée* de ce problème. Nous formulons d’abord le principe de l’estimation *mono-image* des cibles *i.e.*, à partir d’une seule image de la séquence observée. Dans ce cas, notre méthode d’estimation peut être vue comme une méthode de prédiction non linéaire globale du fond, dans laquelle l’image des cibles s’identifie aux résidus de prédiction. Nous proposons ensuite une extension *multi-image* de cette méthode et nous illustrons notre propos par des résultats obtenus sur des images de synthèse physique simulées à l’ONERA. La solution bi-image que nous préconisons s’avère aussi performante qu’un filtrage spatio-temporel adapté mais est nettement plus robuste à de faibles vitesses de défilement des cibles.

## 2 Estimation mono-image

Nous notons  $\mathbf{z} = (z_s)_{s \in \mathcal{S}}$  l’image des données (matrice de taille  $M \times N$ ) et nous modélisons les observations comme la superposition des deux composantes :

$$\mathbf{z} = \mathbf{x} + \mathbf{t}, \quad (1)$$

$\mathbf{x}$  et  $\mathbf{t}$  désignant respectivement l’image du fond nuageux et l’image des objets ponctuels. Ce modèle additif est justifié car nous considérons des cibles de très petite taille dont la présence ne masque pas le rayonnement lumineux émis par le fond. Notre but est d’estimer  $\mathbf{x}$  et  $\mathbf{t}$  à partir des données  $\mathbf{z}$ . Ce problème inverse est indéterminé et nous le résolvons dans le cadre de la régularisation [6] en prenant en compte des informations *a priori* sur les composantes à séparer. Celles-ci se distinguent fondamentalement par le fait que  $\mathbf{x}$  est spatialement coloré, alors que l’image des

cibles  $\mathbf{t}$  est nulle presque partout sauf en certains pixels où l'on observe des pics (points brillants) traduisant la présence de cibles potentielles. Ces caractéristiques sont introduites naturellement dans un contexte bayésien sous la forme de distributions de probabilité. Aussi choisissons nous :

- un modèle corrélé pour  $\mathbf{x}$ , sous la forme d'un champ de Gibbs-Markov *non gaussien* de façon à préserver les discontinuités du fond [7],

$$-\log p(\mathbf{X} = \mathbf{x}) \propto \sum_{r \sim s} \phi(x_r - x_s) + \text{cste} \quad (2)$$

où «  $r \sim s$  » désigne deux plus proches voisins.

- un modèle *i.i.d.* avec une loi « à longue queue » pour  $\mathbf{t}$  afin de modéliser la présence de valeurs extrêmes associées aux points brillants [8],

$$-\log p(\mathbf{T} = \mathbf{t}) \propto \sum_{s \in \mathcal{S}} \chi(t_s) + \text{cste} \quad (3)$$

et nous supposons en outre que les réalisations  $\mathbf{x}$  et  $\mathbf{t}$  sont indépendantes. Nous estimons alors  $\mathbf{t}$  par l'argument  $\hat{\mathbf{t}}$  qui maximise la distribution *a posteriori* sachant  $\mathbf{z}$  (MAP) :

$$\hat{\mathbf{t}} = \arg \max_{\mathbf{t} \in \mathbb{R}^{MN}} p_{\mathbf{T}|\mathbf{Z}}(\mathbf{t}|\mathbf{z}). \quad (4)$$

Cela revient en fait à définir  $\hat{\mathbf{t}}$  comme l'image des résidus :

$$\hat{\mathbf{t}} = \mathbf{z} - \hat{\mathbf{x}} \quad \text{avec} \quad \hat{\mathbf{x}} = \arg \max_{\mathbf{x} \in \mathbb{R}^{MN}} p_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}). \quad (5)$$

$\hat{\mathbf{x}}$  est l'estimée du fond au sens du MAP et minimise donc un critère  $\mathcal{J}_1(\mathbf{x})$  de la forme :

$$\mathcal{J}_1(\mathbf{x}) = \sum_{s \in \mathcal{S}} \chi(z_s - x_s) + \lambda \sum_{r \sim s} \phi(x_r - x_s), \quad (6)$$

où  $\lambda$  est un coefficient strictement positif. Pour des raisons de coût de calcul, nous nous restreignons à un critère *strictement convexe* et différentiable, dont l'optimisation peut être réalisée facilement par une procédure itérative de descente locale de type gradient [9]. Les potentiels  $\chi$  et  $\phi$ , pénalisant respectivement les écarts aux données et les différences entre pixels voisins du fond, sont tous les deux choisis parmi les fonctions «  $L_{21}$  » telles que la fonction hyperbolique :  $u \mapsto \sqrt{u^2 + T^2}$ . Ainsi,  $\mathcal{J}_1(\mathbf{x})$  dépend de trois hyperparamètres, les deux paramètres de seuil  $T_\chi$ ,  $T_\phi$  et le paramètre de pondération  $\lambda$  qui contrôle le compromis entre les deux termes. Ce critère composite fait donc intervenir deux termes convexes mais *non quadratiques*, respectivement robuste aux valeurs aberrantes et préservant les contours. Autrement dit,  $\hat{\mathbf{x}}$  résulte d'un débruitage « non gaussien » à la fois en ce qui concerne le bruit d'observation – identifié à l'image des cibles – et l'*a priori* sur l'objet. Des fonctions de coût similaires à  $\mathcal{J}_1(\mathbf{x})$ , mais souvent non convexes, ont déjà été utilisées pour d'autres applications, comme dans [10] par exemple pour la restauration de films.

Le réglage des hyperparamètres est évidemment crucial. Il est effectué de manière empirique. Toutefois, nous avons proposé dans [5] une étude du comportement local des estimateurs  $\hat{x}_s$  et  $\hat{t}_s$  en un pixel  $s \in \mathcal{S}$  donné. Nous

montrons le caractère effectivement séparateur de la méthode, à condition que le paramètre  $\lambda$  soit supérieur à une valeur minimale explicite :

$$\lambda_{\min} = \frac{\sup |\chi'|}{k \sup |\phi'|}, \quad (7)$$

où  $k$  est le nombre de plus proches voisins (nous prenons  $k = 4$  dans les simulations).

Cependant, comme d'autres techniques de rehaussement de cibles fondées sur des propriétés spatiales uniquement, notre méthode n'est pas satisfaisante pour détecter des cibles de très faible intensité par rapport à la dynamique du fond. Il faut alors compter sur l'apport d'autres données (multi-spectrales ou multi-temporelles) pour améliorer les performances du traitement mono-image.

### 3 Extension à une paire d'images

Nous proposons dans cette partie une méthode d'estimation conjointe des cibles à partir de deux images successives de la séquence, notées  $\mathbf{z}^1$  et  $\mathbf{z}^2$ . Pour ce faire, nous tirons partie de la forte corrélation temporelle entre images du fond. Plus précisément, nous supposons que celles-ci sont identiques au bougé de la ligne de visée près (assimilé à une translation *subpixelique*), et au bruit de mesure et d'échantillonnage près, c'est-à-dire que l'on néglige la déformation du nuage entre deux prises de vues. Par contre, nous ne possédons pas d'information explicite sur le déplacement des cibles et par conséquent nous n'introduisons pas d'interaction entre les images des objets ponctuels. Ainsi, nous considérons que les deux images  $\mathbf{z}^1$  et  $\mathbf{z}^2$  sont chacune la superposition d'une version décalée de l'image du fond nuageux  $\mathbf{x}$  et d'une image distincte des cibles ( $\mathbf{t}^1$  et  $\mathbf{t}^2$ ). En notant  $\delta^1$  et  $\delta^2$  les translations absolues par rapport à un instant de référence donné, le modèle *continu* d'observation bi-image s'écrit :

$$\begin{cases} z^1(u, v) &= x(u - \delta_u^1, v - \delta_v^1) + t^1(u, v) \\ z^2(u, v) &= x(u - \delta_u^2, v - \delta_v^2) + t^2(u, v). \end{cases} \quad (8)$$

Les images observées résultent de l'échantillonnage en un nombre fini de points ( $u = 0, \dots, M-1$  et  $v = 0, \dots, N-1$  par exemple si on considère des périodes d'échantillonnage égales à l'unité) et de la quantification.

La discrétisation des images recherchées implique en outre de remplacer les translations  $\delta^1$  et  $\delta^2$  par des interpolations. Si on pose  $\mathbf{x}^1 = \{x(m - \delta_u^1, n - \delta_v^1)\}_{m,n}$  et  $\mathbf{x}^2 = \{x(m - \delta_u^2, n - \delta_v^2)\}_{m,n}$ , le passage d'une image du fond à la suivante (correspondant théoriquement à une translation relative de  $\delta = \delta^2 - \delta^1$ ) est approchée par un opérateur d'interpolation et de rééchantillonnage  $\mathbf{I}_\delta$  :

$$\mathbf{x}^2 = \mathbf{I}_\delta \mathbf{x}^1. \quad (9)$$

Cela suppose qu'il n'y a pas de repliement dans les images du fond et que l'on peut prévoir  $\mathbf{x}^2$  à partir de  $\mathbf{x}^1$ . En pratique, nous avons recours à un interpolateur bicubique [11], dont le noyau est de taille  $5 \times 5$  si  $\delta \in ]-1, 1[^2$ . Une fois cette approximation faite, et connaissant le décalage relatif  $\delta$  entre les deux fonds, le problème d'estimation des images des cibles  $\mathbf{t}^1$  et  $\mathbf{t}^2$  se réduit à estimer *une seule*

image du fond  $\mathbf{x} = \mathbf{x}^1$ . Les images des cibles  $\mathbf{t}^1, \mathbf{t}^2$  se déduisent alors comme images des résidus :

$$\begin{cases} \hat{\mathbf{t}}^1 &= \mathbf{z}^1 - \hat{\mathbf{x}} \\ \hat{\mathbf{t}}^2 &= \mathbf{z}^2 - \mathbf{I}_\delta \hat{\mathbf{x}} \end{cases} \quad (10)$$

De même qu'en mono-image, nous proposons d'estimer  $\mathbf{x}$  au sens du MAP :

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x} \in \mathbb{R}^{MN}} p_{\mathbf{X}|\mathbf{Z}^1, \mathbf{Z}^2}(\mathbf{x}|\mathbf{z}^1, \mathbf{z}^2), \quad (11)$$

ce qui correspond, sous l'hypothèse supplémentaire d'indépendance entre  $\mathbf{t}^1, \mathbf{t}^2$  et  $\mathbf{x}$ , à minimiser le critère suivant :

$$\begin{aligned} \mathcal{J}_2(\mathbf{x}) &= \sum_{s \in \mathcal{S}} \chi(z_s^1 - x_s) + \sum_{s \in \mathcal{S}} \chi(z_s^2 - [\mathbf{I}_\delta \mathbf{x}]_s) \\ &\quad + \mu \sum_{r \sim s} \phi(x_r - x_s). \end{aligned} \quad (12)$$

Nous choisissons la même fonction de pénalisation  $\chi$  et le même seuil  $T_\chi$  pour les deux termes d'adéquation aux données, et nous prenons toujours pour  $\chi$  et  $\phi$  des fonctions de type «  $L_{21}$  ». À  $\delta$  fixé, le critère  $\mathcal{J}_2(\mathbf{x})$  est donc strictement convexe et dépend de trois paramètres  $T_\chi, T_\phi$  et  $\mu$ . Quant au décalage  $\delta$ , il est estimé au préalable directement sur les données, par moindres carrés non linéaires :

$$\hat{\delta} = \arg \min_{\delta \in \mathbb{R}^2} \|\mathbf{z}^2 - \mathbf{I}_\delta \mathbf{z}^1\|^2 \quad (13)$$

sous l'hypothèse réaliste qu'il y a peu de cibles présentes dans les images observées.

## 4 Résultats expérimentaux

Nous avons mis en œuvre les algorithmes mono et bi-image sur différentes séquences simulées à l'ONERA. Les images du fond traitées représentent une scène nuageuse observée dans le proche infra-rouge. Elles ont été synthétisées par le code Nualum 3D développé à l'ONERA/DOta. Pour tester l'efficacité de la méthode proposée, nous avons incrusté des grilles de cibles pixelliques, régulièrement espacées dans la zone difficile *i.e.*, à l'intérieur du nuage. La figure 1 montre un exemple d'image des données ainsi générée.

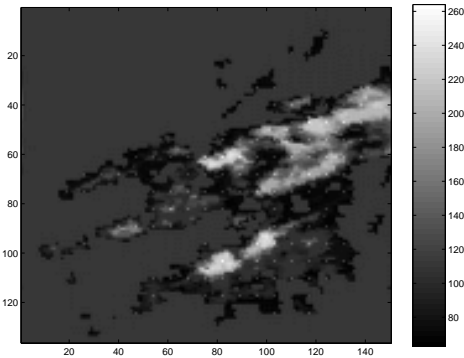


FIG. 1: Exemple d'image de fond nuageux dans laquelle on a inséré, à l'intérieur du nuage, des cibles de la taille du pixel et d'intensité égale à +40 niveaux de gris.

Le résultat de notre modèle mono-image est présenté sur la figure 2. En pratique, nous minimisons le critère (6) par

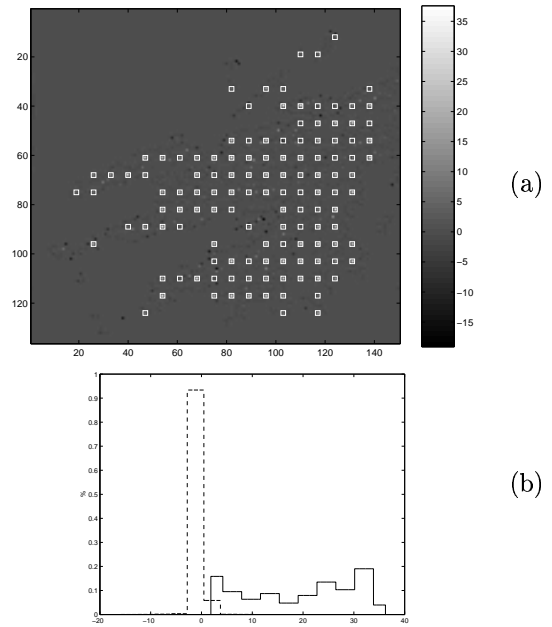


FIG. 2: Résultat d'un traitement mono-image sur l'image de la figure 1 avec les paramètres  $T_\chi = 1, T_\phi = 5$  et  $\lambda = 0.3$ . (a) Image estimée des cibles  $\hat{\mathbf{t}} = \mathbf{z} - \hat{\mathbf{x}}$ , les positions des vraies cibles étant repérées par des carrés blancs (grille  $\mathcal{G}^*$ ). (b) En trait continu, histogramme des résidus sur la grille  $\{\hat{\mathbf{t}}_g, g \in \mathcal{G}^*\}$ , et en pointillés, histogramme des résidus en dehors de la grille  $\{\hat{\mathbf{t}}_r, r \in \mathcal{S} \setminus \mathcal{G}^*\}$ .

une méthode de gradient pseudo-conjugué à pas adaptatif. Les hyperparamètres sont réglés empiriquement. Nous avons choisi la branche d'hyperbole pour les deux fonctions potentiels  $\chi$  et  $\phi$ . Les meilleurs résultats visuels sont obtenus pour des valeurs de  $T_\chi$  et  $T_\phi$  assez petites, et une valeur de  $\lambda$  proche de 0.3 *i.e.*, juste au-dessus de notre borne théorique  $\lambda_{\min} = 1/k = 0.25$ . L'image des résidus, nulle presque partout, fait bien ressortir la majorité des cibles incrustées et présente assez peu de « fausses alarmes ».

Nous avons comparé notre méthode d'estimation à un filtrage adapté classique, spatial et spatio-temporel. Cette technique prédit linéairement les pixels du fond en fonction des données voisines de l'image courante, plus celles de l'image précédente dans le cas bi-image. Les coefficients du filtre sont estimés globalement de sorte à minimiser la norme quadratique de l'erreur de prédiction, et l'image des résidus fournit alors l'estimée de l'image des cibles. Ce type d'algorithmes donne de très bons résultats, en particulier le filtre prédictif spatio-temporel, mais il présente l'inconvénient d'être très sensible à la vitesse des cibles. Lorsque les cibles bougent suffisamment par rapport au fond, les performances de détection sont excellentes, mais si le déplacement relatif des cibles est très petit, le filtre spatio-temporel supprime les cibles en même temps que le fond. Au contraire, notre algorithme bi-image conserve des performances très convenables si la cible ne bouge pas, tout en étant aussi efficace que le filtre adapté quand les cibles défilent rapidement. La figure 3 montre le résultat du traitement dans le cas où les cibles apparaissent ou se déplacent relativement au fond. Nous illustrons également

notre propos par les courbes COR de la figure 4, en distinguant deux cas suivant la vitesse de défilement des cibles. Si les cibles ne bougent pratiquement pas d'une image à l'autre, les performances se maintiennent à un niveau acceptable alors que celles du filtre adapté se dégradent considérablement.

## 5 Conclusion et perspectives

Nous venons de présenter une méthode de traitement mono et bi-image permettant de faire ressortir des objets ponctuels isolés dans un bruit de fond additif fortement structuré. L'approche régularisée proposée passe par une estimation du fond, les images des cibles recherchées correspondant aux résidus. Un modèle non-convexe serait sans doute plus approprié à des fins de détection mais nous nous limitons, pour des raisons de coût de calcul et de simplicité algorithmique, à la minimisation d'une énergie strictement convexe. Les résultats de simulation montrent notamment l'apport de notre méthode par rapport à un filtre adapté standard pour suivre des cibles défilant très lentement d'une image à l'autre. Le travail actuel consiste à mettre en œuvre un modèle de surrésolution qui prenne en compte la réponse impulsionnelle de l'optique et le caractère subpixelique de la position des cibles.

## Références

- [1] T. Soni, J. R. Zeidler et W. H. Ku, « Performance evaluation of 2-D adaptive prediction filters for detection of small objects in image data », *IEEE Trans. Image Processing*, vol. 2, n°3, pp. 327–340, juillet 1993.
- [2] E. A. Ashton, « Detection of subpixel anomalies in multispectral infrared imagery using an adaptive bayesian classifier », *IEEE Trans. Geosci. Remote Sensing*, vol. GE-36, n°2, pp. 506–517, mars 1998.
- [3] I. S. Reed, R. M. Gagliardi et H. M. Shao, « Application of three-dimensional filtering to moving target detection », *IEEE Trans. Aerospace and Electronic Systems*, vol. 19, n°6, pp. 898–905, novembre 1983.
- [4] S. D. Blostein et T. S. Huang, « Detecting small, moving objects in image sequences using sequential hypothesis testing », *IEEE Trans. Signal Processing*, vol. 39, n°7, pp. 1611–1629, juillet 1991.
- [5] V. Samson, F. Champagnat et J.-F. Giovannelli, « Détection d'objets ponctuels en mouvement dans une séquence d'images : une approche régularisée », Rapport technique 1/04005 DTIM, ONERA, février 2001.
- [6] J. Idier, « Regularization tools and models for image and signal reconstruction », in *3rd Intern. Conf. Inverse Problems in Engng.*, Port Ludlow, USA, juin 1999, pp. 23–29.
- [7] C. A. Bouman et K. D. Sauer, « A generalized Gaussian image model for edge-preserving MAP estimation », *IEEE Trans. Image Processing*, vol. 2, n°3, pp. 296–310, juillet 1993.
- [8] P. Ciuciu, J. Idier et J.-F. Giovannelli, « Analyse spectrale non paramétrique haute résolution », in *Actes du 17<sup>e</sup> colloque GRETSI*, Vannes, France, septembre 1999, pp. 721–724.

- [9] D. P. Bertsekas, *Nonlinear programming*, Athena Scientific, Belmont, Massachusetts, 1995.
- [10] S. Geman, D. McClure et D. Geman, « A nonlinear filter for film restoration and other problems in image processing », *CVGIP : Graphical Models and Image Processing*, vol. 54, n°4, pp. 281–289, juillet 1992.
- [11] J. A. Parker, R. V. Kenyon et D. E. Troxel, « Comparison of interpolating methods for image resampling », *IEEE Trans. Medical Imaging*, vol. MI-2, n°1, pp. 31–39, mars 1983.

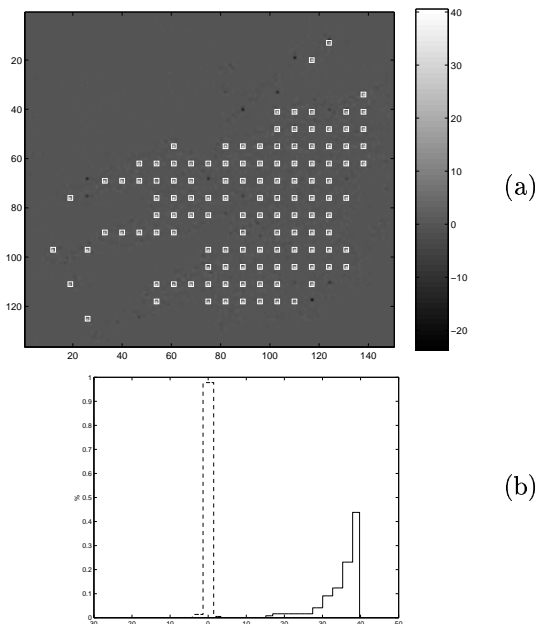


FIG. 3: Résultat d'un traitement bi-image sur deux images successives de la séquence (dont est extraite l'image de la figure 1) dans le cas où les cibles se sont déplacées d'environ un pixel par rapport au fond. Le décalage  $\delta$  est pré-estimé sur les deux images du fond sans cible d'après (13), et les paramètres sont réglés empiriquement,  $T_\chi = 1$ ,  $T_\phi = 5$  et  $\mu = 0.5$ . (a) Image estimée des cibles  $\hat{t}^2 = z^2 - \mathbf{I}_\delta \hat{x}$ . (b) Histogramme des résidus sur la grille (trait continu) et en dehors (pointillés).

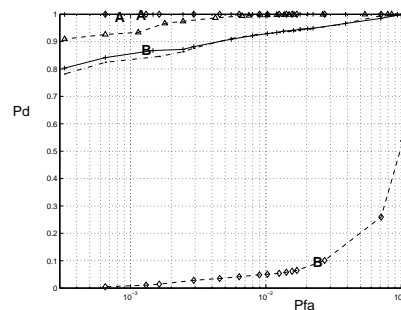


FIG. 4: Courbes COR obtenues respectivement avec le modèle mono-image [---] et le filtre spatial adapté [-Δ-] sur l'image de la figure 1, avec le modèle bi-image [-+-] et le filtre prédictif spatio-temporel [-◇-] sur deux images successives dans le cas où (A) les cibles n'ont pas bougé d'une image à l'autre, (B) les cibles se sont déplacées d'un pixel.