

# Détection robuste par modèle probabiliste d'apparence : une approche bayésienne

## Robust detection using probabilistic appearance-based models: a Bayesian approach

par Rozenn DAHYOT<sup>§</sup>\*, Pierre CHARBONNIER<sup>§</sup> et Fabrice HEITZ<sup>\*\*</sup>

<sup>§</sup> ERA 27 LCPC, Laboratoire des Ponts et Chaussées, 11 rue Jean Mentelin, BP 9, 67035 Strasbourg. Pierre.Charbonnier@equipement.gouv.fr

\* Nouvelles coordonnées de R. Dahyot : Department of Electronic and Electrical Engineering, University Trinity College, Dublin 2, Ireland, dahyot@mee.tcd.ie

\*\* LSIIT UMR CNRS 7005, Université Louis Pasteur – Pôle API, Bd Sébastien Brant, 67400 Illkirch. heitz@lsiit.u-strasbg.fr

### *résumé et mots clés*

Dans cet article, nous nous intéressons à la détection d'objets dans des scènes complexes, par des méthodes basées sur des modèles statistiques d'apparence globale. L'approche proposée associe, dans un cadre bayésien, une représentation standard des images d'apprentissage par espace propre à des modèles de bruit et à des modèles *a priori* non gaussiens. Ce modèle permet d'unifier les méthodes de détection classiques rencontrées dans la littérature et conduit, de façon naturelle, à la définition d'une nouvelle classe de détecteurs statistiques, intégrant des modèles de distribution quelconque pour les images d'apprentissage. La comparaison des caractéristiques opérationnelles des récepteurs (courbes COR) sur des bases de données communes, illustre les contributions de l'approche bayésienne. Elle montre également que l'adoption de modèles non gaussiens permet de dépasser significativement les performances des algorithmes faisant actuellement référence dans le domaine [2, 14].

**Représentation par espace propre, ACP, approche bayésienne, modèles non gaussiens, M-estimateurs, algorithmes semi-quadratiques.**

### *abstract and key words*

In this paper, methods are proposed to detect objects in complex scenes using statistical global appearance based models. In our approach, the standard eigenspace representation of a training image database and *a priori* non-Gaussian hypotheses are brought together in a Bayesian framework. This work unifies standard (appearance-based) detection methods already proposed in the literature and leads naturally to the definition of a new family of probabilistic detectors. It allows the use of more general *a priori* assumptions about the distribution on the eigenspace and its orthogonal. Experimental results are illustrated with ROC (Receiver Operating Characteristic) curves and show the major improvement of our Bayesian approach in comparison to the standard methods that have been the reference up to now [2, 14].

**Eigenspace representation, PCA, Bayesian approach, non-Gaussian models, M-estimators, half-quadratic algorithms.**

## 1. introduction

La détection d'objets dans des images est un sujet difficile qui a fait l'objet de nombreux travaux par le passé. Les approches par *modèles d'apparence*, notamment, rencontrent un succès certain depuis le début des années 90. Ce type de modèle est particulièrement adapté lorsque les conditions d'analyse sont rendues difficiles par la variété et la complexité des scènes, parfois prises dans des conditions d'éclairage non contrôlées ou par la variabilité des objets d'intérêt. Les modèles d'apparence permettent la représentation de classes d'objets et l'apprentissage de variations de forme, d'orientation ou d'illumination. L'information de base est liée à la description de l'image comme une surface d'intensité, soit globalement, soit localement. Dans le premier cas, on considère la valeur brute de tous les pixels, dans le second cas, un ensemble de caractéristiques de bas niveau dérivées de l'intensité [19] [20]. D'abord utilisés en reconnaissance de visages [23], ces modèles ont également permis la mise au point de méthodes de reconnaissance d'objets 3D manufacturés [15] en environnement contrôlé, sur des bases de petite dimension. Plus récemment, les modèles d'apparence ont permis la détection de visages ou de véhicules dans des environnements plus complexes [20].

Nous nous intéressons ici au cas particulier des modèles d'apparence globale où toute image est représentée par un vecteur dont chaque composante est la valeur de niveau de gris d'un pixel. Une telle représentation est évidemment peu parcimonieuse : c'est pourquoi on a en général recours à des techniques de réduction de dimension. Celles-ci peuvent être linéaires, comme l'analyse en composantes principales (ACP) ou l'analyse en composantes indépendantes (ACI). Elles peuvent également être non linéaires : courbes et surfaces principales, utilisation de noyaux ou de réseaux neuromimétiques [7]. Nous nous plaçons ici dans le cadre de l'approche linéaire car elle peut conduire à des implantations nettement moins coûteuses en temps de calcul [8, 12]. D'autre part, la linéarité ne semble pas constituer un obstacle à la représentation de variétés complexes d'objets, comme l'ont montré des études expérimentales récentes [12]. Plus particulièrement, nous effectuons la modélisation de notre nuage d'image d'apprentissages par ACP. Les paramètres de ce modèle sont l'image moyenne et un nombre réduit de vecteurs propres, représentant les axes principaux de variation du nuage de points, qu'il est possible de décomposer en valeurs singulières pour accélérer les calculs [8]. Enfin, les vecteurs propres sont orthogonaux, ce qui est également avantageux en matière de reconstruction.

Les algorithmes de détection s'appuyant sur ce type de représentation reposent soit sur le calcul d'une distance entre l'image observée et le nuage des images d'apprentissage, soit, dans un cadre statistique, sur l'évaluation de la probabilité que l'image observée appartienne à la classe des images apprises. Les distances utilisées initialement étaient simples : distance quadra-

tique au centre de masse du nuage d'apprentissage (*Sum of Squared Differences* ou SSD) ou distance orthogonale à l'espace propre (*Distance From Feature Space* ou DFFS) [23, 14]. Aucune des deux approches n'est parfaitement satisfaisante : la première assigne la même distance à toutes les images situées sur une hyper-sphère, la seconde assigne la même valeur à toutes les images situées sur des variétés parallèles à l'espace propre. Il est donc facile de construire des contre-exemples mettant en échec ces méthodes. Moghaddam et Pentland ont proposé plus récemment [13, 14] une approche statistique basée sur l'estimation au sens du maximum de vraisemblance dans le cadre de distributions gaussiennes. L'approximation de la vraisemblance ainsi obtenue a l'avantage de combiner à la fois une information de distance dans l'espace propre et une information de distance orthogonale à celui-ci. Les études expérimentales menées par Moghaddam *et al.* [14, 12] ont montré la supériorité de cette méthode non seulement par rapport à celles utilisant la SSD ou la DFFS, mais aussi par rapport à des méthodes basées sur une représentation non linéaire des nuages d'apprentissage. On peut donc considérer que cette méthode fait actuellement référence dans le domaine. Signalons enfin que Tipping et Bishop ont développé, de façon indépendante dans [22], une autre interprétation probabiliste de l'ACP, qui conduit au même modèle que celui proposé par Moghaddam et Pentland [13].

Les distributions de bruit associées aux méthodes par espace propre sont souvent supposées gaussiennes. Une telle hypothèse n'est cependant vérifiée que très rarement en pratique. Les méthodes classiques de détection et de reconnaissance, basées sur cette représentation [13, 14], sont ainsi sensibles à la présence de données erronées ou *outliers*, induites par exemple, par des occultations partielles. Les M-estimateurs, proposés par Huber [9] dans le cadre de l'estimation statistique, tolèrent des écarts au modèle et permettent donc de gérer les occultations. Ils ont ainsi récemment permis de définir des méthodes robustes de reconnaissance reposant sur l'ACP [2, 6].

Une autre limitation importante des approches précédentes est l'adoption (explicite ou implicite) de modèles gaussiens (voire uniformes) pour représenter la distribution *a priori* des images dans l'espace propre [2, 22]. Ceci conduit à biaiser fortement la détection vers l'image moyenne (centre de masse du nuage des images d'apprentissage), alors que les distributions sont souvent non gaussiennes (voir paragraphe 2.6 et [15]). Une première approche ayant intégré des modèles non gaussiens pour la distribution des images d'apprentissage est l'approche de Murase et Nayar [15], qui représentent de façon *ad hoc* le nuage des images d'apprentissage par des B-splines et utilisent des critères heuristiques de distance par rapport à la variété non linéaire ainsi définie. Des modèles de mélanges gaussiens [14, 21] ainsi que des approches neuromimétiques [12] ont été considérés plus récemment.

Dans cet article, nous abordons le problème de la détection dans un cadre bayésien général qui permet d'unifier la description des différents modèles rencontrés dans la littérature. L'approche

proposée intègre de façon naturelle des contraintes statistiques non linéaires sur la distribution des images d'apprentissage dans le plan propre. Nous montrons expérimentalement l'importance d'une modélisation fine de cette distribution et son impact sur les performances de détection. Par ailleurs, la méthode offre la possibilité de mettre en œuvre des modèles robustes de bruit capables de gérer la présence d'erreurs grossières. Ceci conduit à la définition d'une famille de détecteurs dont les performances expérimentales dépassent celles des méthodes faisant actuellement référence dans le domaine [2, 14].

La suite de l'article est organisée de la façon suivante. Nous développons dans la partie 2 l'approche bayésienne unificatrice de la détection que nous proposons. Nous montrons dans la partie 3 comment cette approche permet de retrouver les méthodes classiques de détection et détaillons l'implantation de nos nouvelles méthodes de détection robuste bayésienne. La partie 4 offre une comparaison de cette approche avec les méthodes classiques en illustrant, sur trois bases d'images, l'apport des différents éléments de modélisation (modèles de bruit, modèles *a priori*). Une comparaison objective entre méthodes est menée à partir des courbes COR des différents détecteurs, montrant l'apport de l'approche proposée.

## 2. détection : une approche bayésienne

### 2.1. représentation par espace propre

Lors de la phase préliminaire d'apprentissage, on collecte un ensemble représentatif (base d'apprentissage)  $\mathcal{B}$  de  $K$  images  $\mathbf{x}^k, k = 1, \dots, K$ , à  $N$  pixels, en sélectionnant différentes vues des objets à modéliser. Les données sont mises sous forme vectorielle par ordonnancement lexicographique. Pour simplifier, nous développons ici le cas des images en niveaux de gris, l'extension au cas des images en couleurs étant détaillée dans [5]. On définit  $\boldsymbol{\mu}$ , la moyenne des échantillons d'apprentissage, par :

$$\boldsymbol{\mu} = \frac{1}{K} \sum_{k=1}^K \mathbf{x}^k \quad (1)$$

La matrice de covariance associée à la base d'apprentissage est estimée par :

$$\Sigma = \frac{1}{K} \sum_{k=1}^K (\mathbf{x}^k - \boldsymbol{\mu}) (\mathbf{x}^k - \boldsymbol{\mu})^T \quad (2)$$

Elle est symétrique, semi-définie positive et se diagonalise :  $\Sigma = U_N \Lambda_N U_N^T$ . Dans cette expression,  $U_N$  est la matrice contenant les  $N$  vecteurs propres orthonormés de  $\Sigma$  et  $\Lambda_N$  est la

matrice diagonale contenant les  $N$  valeurs propres associées, notées  $\{\lambda_j\}_{j \in \{1, \dots, N\}}$ , classées en ordre décroissant. Chaque échantillon  $\mathbf{x}$  peut s'écrire comme la somme de la moyenne des images d'apprentissage,  $\boldsymbol{\mu}$ , et d'une combinaison linéaire de  $J$  vecteurs de base (rangés en colonnes dans une matrice  $U$ ), à une certaine erreur de reconstruction  $\mathbf{w}^r$  près :

$$\mathbf{x} = \boldsymbol{\mu} + U\mathbf{c} + \mathbf{w}^r = \boldsymbol{\mu} + \sum_{j=1}^J c_j \mathbf{u}_j + \mathbf{w}^r \quad (3)$$

Dans (3), les vecteurs propres  $\mathbf{u}_j$  correspondent aux  $J$  plus fortes valeurs propres de la matrice de covariance  $\Sigma$ . Ce modèle est celui de l'Analyse en Composantes Principales (ACP), classique en analyse des données. L'équation 3 est à la base du modèle statistique *non standard* que nous allons développer dans la suite (paragraphe 2.3).

Dans notre implantation, nous utilisons une procédure de centrage modifiée, de façon à améliorer la phase d'apprentissage. En effet, les images des objets d'apprentissage sont généralement acquises sur fond neutre (en pratique un fond noir, voir par exemple figure 2). La présence de ces pixels de *fond* peut entraîner des variabilités supplémentaires, non significatives. Aussi les écartons-nous lors du calcul de la moyenne :

$$\mu_n^v = \frac{1}{\sum_{k=1}^K b_n^k} \sum_{k=1}^K b_n^k x_n^k$$

où  $b_n^k$  est la pondération appliquée au pixel  $n$  de l'image d'apprentissage  $\mathbf{x}^k$ . Celle-ci est égale à 0 si  $x_n^k$  est un pixel de *fond*, 1 si c'est un pixel de l'*objet*. Puis chaque pixel de l'échantillon  $\mathbf{x}^k$  est centré par :

$$\tilde{x}_n^k = b_n^k (x_n^k - \mu_n^v) \quad (4)$$

Ainsi, les pixels *objet* sont centrés tandis que les pixels *fond* sont mis à 0. Cette technique de centrage permet d'obtenir des distributions plus compactes [5]. De plus, comme nous le vérifierons expérimentalement, elle améliore les performances de l'ensemble des méthodes de détection.

### 2.2. principe général de l'algorithme de détection

Le principe général de l'algorithme de détection d'objets d'intérêt dans une scène est illustré figure 1. L'image est parcourue séquentiellement, dans l'ordre lexicographique. En chaque position  $(i, j)$ , on extrait le contenu d'une fenêtre d'observation. Le vecteur  $\mathbf{y}$ , ainsi obtenu, est confronté à la base d'images apprises, en évaluant sa vraisemblance  $\mathcal{P}(\mathbf{y}|\mathcal{B})$ , par rapport au modèle appris sur la base  $\mathcal{B}$ . Cette valeur est affectée à la position où a

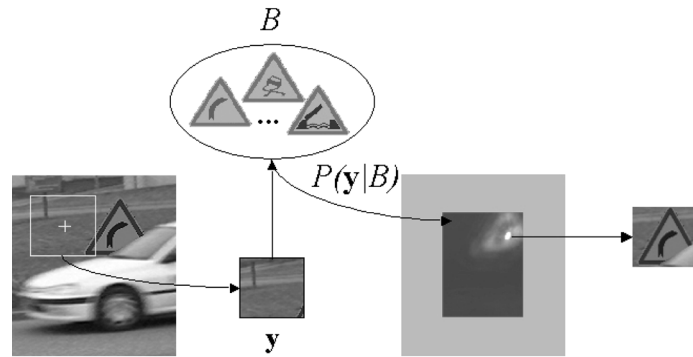


Figure 1. – Algorithme de détection : extraction d'une observation (à gauche) en chaque point de l'image. Calcul du cologarithme de la vraisemblance (au centre). Seuillage de la carte de ressemblance obtenue, permettant d'obtenir la position de l'objet détecté (à droite).

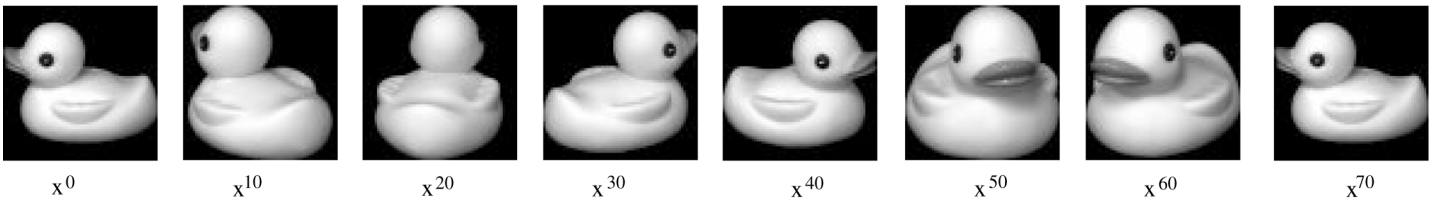


Figure 2. – Exemple d'images d'apprentissage de la base COIL [16].

été extraite l'observation  $\mathbf{y}$ . Lorsque toute la scène a été parcourue de cette manière, on dispose d'une carte de vraisemblance des observations (on peut parler de mesure de *ressemblance* car on se ramène souvent à des critères énergétiques en passant au cologarithme). Un seuillage permet de déterminer si un ou plusieurs objets d'intérêt sont présents dans la scène et, le cas échéant, d'en déterminer la position.

### 2.3. modélisation des observations

Le modèle d'observation que nous considérons correspond à une interprétation statistique *non standard* de l'ACP, inspirée de celle développée par Tipping et Bishop [22] dans le cas gaussien. Plus précisément, nous considérons que toute observation  $\mathbf{y}$  peut se reconstruire à partir des vecteurs propres de la base :

$$\mathbf{y} = \boldsymbol{\mu} + \sum_{j=1}^J c_j \mathbf{u}_j + \mathbf{w}. \quad (5)$$

Ici, le bruit  $\mathbf{w}$  correspond à la somme de l'erreur classique de reconstruction  $\mathbf{w}^r$  (due à la troncature de la représentation) et d'un bruit d'observation  $\mathbf{w}^o$  produit par le système d'acquisition, d'éventuelles occultations, ou la présence d'un fond texturé. Dans l'approche classique de l'ACP probabiliste [14, 22], les

variables latentes  $c_j$  sont décorréelées et suivent une distribution *a priori* gaussienne. Il en est de même pour le bruit  $\mathbf{w}$ . Dans le modèle non standard proposé ici, nous relâchons ces hypothèses, et considérons des modèles non gaussiens pour l'*a priori* ainsi que des modèles robustes, également non-gaussiens, pour le bruit. La représentation proposée ne correspond donc plus à l'approche classique de l'ACP probabiliste. En particulier, les paramètres du modèle (les vecteurs propres  $\mathbf{u}_j$ ) ne s'interprètent plus comme des estimées au sens du maximum de vraisemblance, comme dans [22]. Par contre, l'avantage de cette approche est de permettre une meilleure représentation des distributions complexes pouvant intervenir dans les cas réels.

### 2.4. approche bayésienne pour la détection

Nous nous plaçons dans le cas général où  $\mathbf{c}$  est la réalisation d'une variable aléatoire. La vraisemblance de l'observation,  $\mathcal{P}(\mathbf{y}|\mathcal{B})$ , est obtenue en intégrant par rapport à  $\mathbf{c}$  la vraisemblance conjointe du couple  $(\mathbf{c}, \mathbf{y})$  :

$$\begin{aligned} \mathcal{P}(\mathbf{y}|\mathcal{B}) &= \int \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B}) d\mathbf{c} \\ &= \int \mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B}) \mathcal{P}(\mathbf{c}|\mathcal{B}) d\mathbf{c} \end{aligned} \quad (6)$$

Dans cette expression, on voit apparaître deux termes. Le premier,  $\mathcal{P}(\mathbf{y}|\mathbf{c},\mathcal{B}) = \mathcal{P}(\mathbf{w}|\mathcal{B})^1$ , correspond à la distribution du bruit. Le second terme,  $\mathcal{P}(\mathbf{c}|\mathcal{B})$ , incorpore une connaissance *a priori* sur la distribution des images d'apprentissage dans l'espace propre. Nous rappelons que dans l'algorithme de détection, cette quantité doit être évaluée pour chaque image analysée. Par souci d'efficacité, il est donc souhaitable de disposer d'une expression analytique simple de la vraisemblance (6). Or, mis à part certains cas particuliers comme le cas où le modèle de bruit et l'*a priori* sont tous deux gaussiens [22], il est en général difficile d'obtenir une expression analytique de  $\mathcal{P}(\mathbf{y}|\mathcal{B})$ . Des techniques d'intégration numérique, comme les méthodes de Monte-Carlo, pourraient être utilisées [1]. Cette solution n'a pas été envisagée ici, en raison du coût calculatoire important associé à ce type d'approches.

Une autre solution, adoptée ici, consiste à approcher la distribution  $\mathcal{P}(\mathbf{y}|\mathcal{B})$  par une expression simple. Différentes approximations ont été proposées dans le cadre bayésien, pour calculer les distributions *a posteriori* [1, 11]. Nous adoptons une approximation proposée dans [11], reposant sur le fait que la distribution  $\mathcal{P}(\mathbf{y},\mathbf{c}|\mathcal{B})$  présente souvent un pic important pour les paramètres  $\hat{\mathbf{c}}$  les plus probables. La vraisemblance de l'observation,  $\mathcal{P}(\mathbf{y}|\mathcal{B})$  peut alors être approchée par la hauteur du pic multipliée par sa « largeur »,  $\sigma_{pic}$  :

$$\mathcal{P}(\mathbf{y}|\mathcal{B}) \simeq \sigma_{pic} \max_{\mathbf{c}} \mathcal{P}(\mathbf{y},\mathbf{c}|\mathcal{B}) \propto \mathcal{P}(\mathbf{y}|\hat{\mathbf{c}},\mathcal{B}) \mathcal{P}(\hat{\mathbf{c}}|\mathcal{B}) \quad (7)$$

Cette approximation, qui consiste à confondre la distribution avec son mode, se rattache à d'autres méthodes d'approximation classiquement utilisées en inférence bayésienne, comme la méthode de Laplace [1], ou la méthode du mode ([7], page 92). Elle constitue souvent une bonne approximation (en particulier dans le cas gaussien) et conduit, même dans le cas non gaussien à une distribution approchée, qui permet une bonne détection. Elle est également justifiée *a posteriori* par la qualité des résultats qu'elle permet d'obtenir, par rapport aux détecteurs classiques (voir paragraphe 4).

Dans ce qui suit, nous décrivons différents modèles envisageables pour la distribution du bruit et pour la distribution *a priori* des variables latentes,  $\mathbf{c}$ , dans l'espace propre.

## 2.5. modèles de bruit

La distribution la plus couramment utilisée pour le bruit est bien entendu la distribution gaussienne, avec une matrice de covariance diagonale [22] :

1. Rappel :  $\mathbf{w} = \mathbf{y} - \boldsymbol{\mu} - \mathbf{U}\mathbf{c}$ .

$$\begin{aligned} \mathcal{P}(\mathbf{y}|\mathbf{c},\mathcal{B}) &\propto \exp\left[-\frac{\|\mathbf{y} - \boldsymbol{\mu} - \mathbf{U}\mathbf{c}\|^2}{2\sigma_g^2}\right] \\ &= \exp\left[-\frac{1}{2} \sum_{n=1}^N \left(\frac{w_n}{\sigma_g}\right)^2\right] \end{aligned} \quad (8)$$

L'hypothèse gaussienne est souvent peu satisfaisante, notamment si l'on considère des occurrences dégradées, par exemple, par des occultations. De tels effets provoquent des valeurs importantes du résiduel  $w_n$  auquel le modèle gaussien confère pourtant une probabilité quasi-nulle d'apparition. Ces données erronées sont souvent désignées par le terme anglais d'*outlier*, qui traduit bien le fait qu'elles se situent *en dehors* d'une répartition gaussienne. Leur effet néfaste sur les algorithmes d'estimation est bien connu. Ainsi, la présence d'une seule valeur erronée suffit-elle à biaiser le résultat d'un algorithme de régression au sens des moindres carrés [9]. Afin de prendre en compte la présence d'outliers dans le cadre des modèles d'apparence, Black [2] a proposé d'utiliser des modèles issus de la statistique robuste, conduisant aux désormais classiques M-estimateurs [9]. Ces modèles permettent la prise en compte de données erronées en remplaçant la loi gaussienne par une loi exponentielle de la forme :

$$\mathcal{P}(\mathbf{y}|\mathbf{c},\mathcal{B}) \propto \exp\left[-\frac{1}{2} \sum_{n=1}^N \rho\left(\frac{w_n}{\sigma_\rho}\right)\right] \quad (9)$$

où  $\rho$  est une fonction non quadratique, qui peut également être non convexe [9, 2, 6].

## 2.6. modèles *a priori*

**Distribution uniforme.** Le modèle le plus simple est celui d'une distribution uniforme des variables latentes dans l'espace propre :

$$\mathcal{P}(\mathbf{c}|\mathcal{B}) = cste$$

Cela correspond à une absence de connaissance *a priori* sur la distribution dans l'espace propre.

**Distribution gaussienne.** Une seconde hypothèse [22] consiste à supposer que la distribution des variables latentes est gaussienne :

$$\mathcal{P}(\mathbf{c}|\mathcal{B}) = \frac{\exp\left[-\frac{1}{2} \mathbf{c}^T \boldsymbol{\Lambda}^{-1} \mathbf{c}\right]}{\prod_{j=1}^J \sqrt{2\pi\lambda_j}}$$

où les  $J$  variances  $\lambda_j$ , apparaissant sur la diagonale de  $\boldsymbol{\Lambda}$ , sont les valeurs propres obtenues dans l'apprentissage par ACP. Cette

modélisation consiste à associer un modèle gaussien à l'ACP, comme dans [14, 22].

**Distribution quelconque.** Bien que très utilisés, les modèles uniformes ou gaussiens sont souvent inadaptés pour représenter les distributions observées sur les bases d'images réelles. A titre d'illustration, nous considérons dans la suite trois bases d'apprentissage, sur lesquelles nous comparerons les différents algorithmes. La première base (COIL) est construite à partir d'images de la base de Columbia University, les deux autres à partir d'images de panneaux routiers, dans le cadre d'une application de détection et reconnaissance de la signalisation, en cours de développement au Laboratoire des Ponts et Chaussées. Nous présentons figure 2 quelques échantillons de la base COIL.

Cette base contient 72 images en niveaux de gris de l'objet « canard » de la base COIL [16], avec des points de vue 3D variés. Notre second exemple est une base d'apprentissage constituée d'un seul objet, subissant des rotations dans le plan de l'image, à raison d'une image tous les 2 degrés (cf. figure 3). L'objet en question est l'image moyenne des panneaux de signalisation de danger à fond blanc et la base sera nommée AVG. La troisième base est composée des images en couleurs des 43 panneaux triangulaires, jaunes et blancs, de danger (type « A »), variant suivant des rotations 2D dans le plan de l'image, à raison d'une image tous les 10 degrés (cf. figure 4). Nous la nommerons base A43.

La figure 5 montre la distribution des projections des images d'apprentissage des bases COIL et A43 sur un espace propre

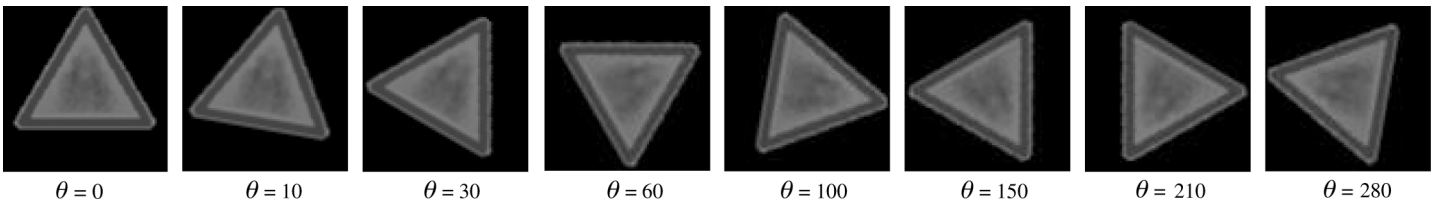


Figure 3. – Base AVG : image moyenne des panneaux blancs de danger, apprise suivant ses rotations dans le plan de l'image, où  $\theta$  est l'angle de rotation.



Figure 4. – Quelques images d'apprentissage de la base A43 [5].

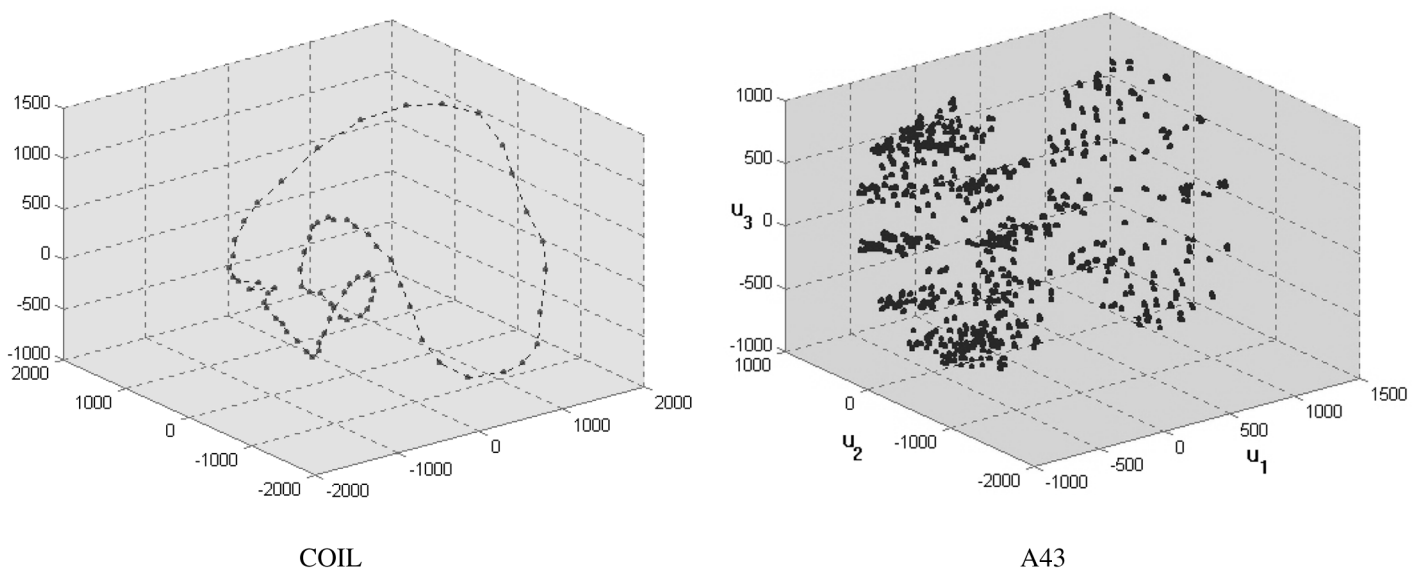


Figure 5. – Distribution des variables latentes  $c$  dans un espace propre réduit à 3 dimensions.

réduit à trois dimensions principales. On constate que la densité de probabilité dans l'espace propre n'est pas gaussienne, ni dans un cas, ni dans l'autre. En ce qui concerne la base COIL, la courbe proposée connecte les coordonnées des échantillons, dans l'ordre des angles de rotation croissants. Murase et Nayar ont proposé dans leur article [15] d'approcher ce type de variété non linéaire par des courbes B-splines. Dans le cas de la base A43, on observe 2 nuages distincts. Ceux-ci correspondent aux panneaux de signalisation usuels à fond blanc et aux panneaux temporaires à fond jaune, moins nombreux. C'est la principale variabilité, captée par la composante principale  $\mathbf{u}_1$ . Les couronnes circulaires observées dans le plan propre défini par  $(\mathbf{u}_2, \mathbf{u}_3)$  sont typiques du cas de l'apprentissage des rotations dans le plan image [17] (voir aussi la figure 6 pour un objet unique, base AVG, et le paragraphe 4.2).

Bien que des extensions à des modèles de mélanges gaussiens [14, 22] puissent apporter une certaine flexibilité à la représentation classique, les modèles paramétriques sont de toutes façons basés sur la connaissance de la forme des densités sous-jacentes et peuvent s'avérer insuffisants pour représenter les distributions rencontrées en pratique [7]. Au contraire, les méthodes non-paramétriques sont un outil efficace pour modéliser des distributions arbitraires, sans faire d'hypothèse sur les formes des densités de probabilité. Nous utilisons des fenêtres de Parzen à noyaux gaussiens, appliquées aux projections des images d'apprentissage dans l'espace propre :

$$\mathcal{P}(\mathbf{c}|\mathcal{B}) = \frac{1}{K} \sum_{k=1}^K \left( \frac{1}{\sqrt{2\pi}\sigma_p} \right)^J \exp - \left[ \frac{\|\mathbf{c} - \mathbf{c}^k\|^2}{2(\sigma_p)^2} \right] \quad (10)$$

On rappelle que  $K$  correspond au nombre d'images d'apprentissage de  $\mathcal{B}$  et  $J$  est la dimension ou le nombre de vecteurs propres retenus pour définir l'espace propre  $F$ . Le paramètre  $\sigma_p$ , qui conditionne la résolution de la représentation de la densité de probabilité, est pour l'instant fixé empiriquement.

### 3. algorithmes de détection

L'approximation (7) conduit à maximiser la distribution  $\mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$  par rapport à  $\mathbf{c}$ . Ceci revient à effectuer une reconstruction de l'image observée, au sens du Maximum A Posteriori, sur l'espace propre :

$$\begin{aligned} \hat{\mathbf{c}} &= \arg \max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B}) \\ &= \arg \max_{\mathbf{c}} \{ \mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B}) \cdot \mathcal{P}(\mathbf{c}|\mathcal{B}) \} \end{aligned} \quad (11)$$

et, d'après (7), la vraisemblance de l'observation est approchée par :

$$\mathcal{P}(\mathbf{y}|\mathcal{B}) \propto \mathcal{P}(\mathbf{y}|\hat{\mathbf{c}}, \mathcal{B}) \cdot \mathcal{P}(\hat{\mathbf{c}}|\mathcal{B})$$

Nous décrivons dans la suite les expressions ou algorithmes nécessaires à l'estimation de  $\hat{\mathbf{c}}$  et le calcul de la mesure de ressemblance,  $-\ln \mathcal{P}(\mathbf{y}|\mathcal{B})$ . Ces expressions sont données pour les différentes distributions du bruit,  $\mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B})$ , et de l'a priori,  $\mathcal{P}(\mathbf{c}|\mathcal{B})$ , considérées plus haut. Nous commençons par le cas du modèle de bruit gaussien, ce qui permet de faire le lien avec les détecteurs classiques associés à l'ACP probabiliste. Nous présentons ensuite le cas des modèles de bruit non gaussiens et proposons des détecteurs robustes. Enfin, le cas le plus général des modèles non gaussiens, à la fois pour la distribution du bruit et pour l'a priori, est traité dans le paragraphe 3.3.

#### 3.1. méthodes classiques de détection

Dans ce paragraphe, on considère l'hypothèse classique de distribution gaussienne du bruit.

##### 3.1.1. modèle de bruit gaussien, a priori uniforme

Lorsque la distribution de  $\mathbf{c}$  dans l'espace propre,  $\mathcal{P}(\mathbf{c}|\mathcal{B})$ , est uniforme, la vraisemblance conjointe se réduit à la distribution de probabilité du bruit (8) et la vraisemblance de l'observation à :

$$\mathcal{P}(\mathbf{y}|\mathcal{B}) \propto \exp \left[ - \frac{\|\mathbf{y} - \boldsymbol{\mu} - \mathbf{U}\hat{\mathbf{c}}\|^2}{2\sigma_g^2} \right] \quad (12)$$

L'estimée correspondante,  $\hat{\mathbf{c}}$ , est la solution des moindres carrés, c'est-à-dire la projection des données sur l'espace propre :

$$\hat{\mathbf{c}} = \mathbf{U}^T (\mathbf{y} - \boldsymbol{\mu})$$

En passant au cologarithme dans (12), on retrouve la mesure classique de ressemblance DFFS (*Distance From Feature Space*) [13], qui correspond à la distance (orthogonale) euclidienne entre l'observation centrée et l'espace propre. Nous nommerons ce détecteur GU pour rappeler nos hypothèses : distribution gaussienne du bruit et distribution a priori uniforme dans l'espace propre.

$$\text{GU}(\mathbf{y}) = \|\mathbf{y} - \boldsymbol{\mu} - \mathbf{U}\hat{\mathbf{c}}\|^2 \quad (13)$$

##### 3.1.2. modèle de bruit gaussien, a priori gaussien

On considère maintenant l'hypothèse d'un a priori gaussien sur les variables latentes : la variable aléatoire  $\mathbf{c}$  de dimension  $J$  suit une loi normale de moyenne nulle et de matrice de covariance  $\Lambda$ . Celle-ci est la matrice diagonale des  $J$  premières valeurs propres  $\lambda_j$  issues de l'apprentissage par ACP. La vraisemblance

s'écrit, cette fois :

$$\mathcal{P}(\mathbf{y}|\mathcal{B}) \propto \exp\left[-\frac{\|\mathbf{y} - \boldsymbol{\mu} - \mathbf{U}\hat{\mathbf{c}}\|^2}{2\sigma_g^2}\right] \cdot \exp\left[-\frac{1}{2}\hat{\mathbf{c}}^T \Lambda^{-1} \hat{\mathbf{c}}\right] \quad (14)$$

L'estimée est obtenue par :

$$\hat{\mathbf{c}} = (\mathbf{I}_J + \sigma_g^2 \Lambda^{-1})^{-1} \mathbf{U}^T \cdot (\mathbf{y} - \boldsymbol{\mu})$$

où  $\mathbf{I}_J$  désigne la matrice identité de dimension  $J \times J$ . Comme  $\mathbf{I}_J$  et  $\Lambda$  sont diagonales, le calcul de l'estimée est direct. En passant au cologarithme dans (14), on obtient l'expression du détecteur associé, que nous nommerons GG, pour gaussien-gaussien :

$$\text{GG}(\mathbf{y}) = \frac{\|\mathbf{y} - \boldsymbol{\mu} - \mathbf{U}\hat{\mathbf{c}}\|^2}{\sigma_g^2} + \sum_{j=1}^J \frac{(\hat{c}_j)^2}{\lambda_j}$$

On remarque qu'on retrouve une expression de forme similaire à celle proposée à la fois par Moghaddam [14] et par Tipping [22].

### 3.2. méthodes robustes de détection

Nous nous intéressons maintenant au cas où le modèle de bruit,  $\mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B})$ , n'est plus gaussien, de façon à prendre en compte les données erronées ou *outliers*.

#### 3.2.1. modélisation robuste du bruit, a priori uniforme

Afin de faciliter la présentation des algorithmes, nous considérons tout d'abord le cas où la distribution de  $\mathbf{c}$  dans le plan propre est supposée uniforme :  $\mathcal{P}(\mathbf{c}|\mathcal{B}) = \text{cste}$ . D'après (9), l'énergie associée au MAP s'écrit à une constante additive près :

$$\mathcal{J}(\mathbf{c}) = \sum_{n=1}^N \rho\left(\frac{w_n}{\sigma_\rho}\right) \quad (15)$$

Cette énergie est minimisée grâce à la théorie semi-quadratique [3], qui introduit une nouvelle fonction d'énergie, notée  $\mathcal{J}^*$ , augmentée d'une variable  $\mathbf{b}$  et ayant le même minimum global que  $\mathcal{J}$  :

$$\mathcal{J}(\mathbf{c}) = \min_{\mathbf{b}} \left\{ \mathcal{J}^*(\mathbf{c}, \mathbf{b}) = \sum_{n=1}^N \rho^*\left(\frac{w_n}{\sigma_\rho}, b_n\right) \right\}$$

L'énergie augmentée est minimisée alternativement en  $\mathbf{c}$  et en  $\mathbf{b}$  par des techniques linéaires [3]. Deux formes d'énergie augmentée ont été proposées [3], conduisant à deux algorithmes de minimisation différents que nous présentons maintenant.

**ARTUR ou minimisation avec des poids modifiés.** La première forme d'énergie augmentée est :

$$\mathcal{J}^A(\mathbf{c}, \mathbf{b}^A) = \sum_{n=1}^N b_n^A \left(\frac{w_n}{\sigma_\rho}\right)^2 + \Psi(b_n^A)$$

Cette expression conduit à un algorithme de type *moindres carrés pondérés itératifs* qui, à l'étape ( $m$ ), s'écrit :

$$\begin{cases} \mathbf{w}^{(m)} = \mathbf{y} - \boldsymbol{\mu} - \mathbf{U} \mathbf{c}^{(m)} \\ \forall n \in \{1 \dots N\} b_n^{A(m+1)} = \frac{\rho'\left(\frac{w_n^{(m)}}{\sigma_\rho}\right)}{2 \frac{w_n^{(m)}}{\sigma_\rho}} \\ \mathbf{c}^{(m+1)} = (\mathbf{U}^T \cdot \mathbf{B}^{A(m+1)} \cdot \mathbf{U})^{-1} \cdot \mathbf{U}^T \cdot \mathbf{B}^{A(m+1)} \cdot (\mathbf{y} - \boldsymbol{\mu}) \end{cases} \quad (16)$$

où  $\mathbf{B}^{A(m+1)}$  est la matrice diagonale qui collecte les pondérations  $b_n^{A(m+1)}$ . Cet algorithme est largement utilisé [8], notamment dans le cadre de la reconnaissance robuste [2, 6]. Le calcul et l'inversion d'une matrice nécessaires à chaque étape sont, cependant, relativement coûteux en temps de calcul [5].

**LEGEND ou minimisation avec des résidus modifiés.** La seconde forme d'énergie augmentée proposée par la théorie semi-quadratique est :

$$\mathcal{J}^L(\mathbf{c}, \mathbf{b}^L) = \sum_{n=1}^N \left(\frac{w_n}{\sigma_\rho} - b_n^L\right)^2 + \xi(b_n^L) \quad (17)$$

Cette expression conduit à un algorithme de *moindres carrés avec des résidus modifiés* :

$$\begin{cases} \mathbf{w}^{(m)} = \mathbf{y} - \boldsymbol{\mu} - \mathbf{U} \mathbf{c}^{(m)} \\ \forall n \in \{1 \dots N\}, b_n^{L(m+1)} = w_n \left(1 - \frac{\rho'\left(\frac{w_n^{(m)}}{\sigma_\rho}\right)}{2 \frac{w_n^{(m)}}{\sigma_\rho}}\right) \\ \mathbf{c}^{(m+1)} = \mathbf{U}^T \cdot (\mathbf{y} - \boldsymbol{\mu} - \sigma_\rho \cdot \mathbf{b}^{L(m+1)}) \end{cases} \quad (18)$$

Ces deux algorithmes sont équivalents à ceux proposés par Huber [9]. Le second est, à notre connaissance, très peu utilisé en pratique. Pourtant, il peut être avantageux dans le cadre de la reconstruction sur une base orthogonale. On peut montrer que chaque étape de LEGEND fait moins décroître l'énergie qu'une étape d'ARTUR, ce qui le rend plus lent à converger. Par contre, chaque étape correspond à une simple projection, sans inversion



de matrice. Globalement, l'algorithme LEGEND s'est avéré beaucoup moins coûteux en temps de calcul dans le cadre de notre application [5].

L'estimée  $\hat{\mathbf{c}}$  est, dans les deux cas, la valeur obtenue à convergence. Celle-ci est calculée pour la fonction  $\rho(x)$  de Geman et McClure (GM, voir tableau 1). Cette fonction n'étant pas convexe, nous utilisons une stratégie de continuation. On introduit graduellement la non convexité en minimisant successivement les fonctions HS (hyper-surfaces, convexe), HL (Hebert et Leahy, non convexe), et enfin GM [6]. Le paramètre d'échelle,  $\sigma_\rho$ , intervenant dans l'expression du potentiel, est estimé préalablement grâce aux images d'apprentissage [6], celles-ci étant considérées sans défaut. La méthode ne requiert donc pas de réglage de paramètre par l'utilisateur.

Tableau 1. – Fonctions robustes utilisées en continuation [6].

	$\rho(x)$	convexité
HS	$2\sqrt{1+x^2} - 2$	convexe
HL	$\log(1+x^2)$	non convexe
GM	$\frac{x^2}{1+x^2}$	non convexe

La mesure de ressemblance à laquelle on aboutit dans ce cas s'écrit :

$$\text{RU}(\mathbf{y}) = \min_{\mathbf{c}} \sum_{n=1}^N \rho\left(\frac{w_n}{\sigma_\rho}\right) = \mathcal{J}(\hat{\mathbf{c}}) \quad (19)$$

Elle peut s'interpréter comme une version robuste de la DFFS, et sera appelée RU (pour robuste-uniforme).

### 3.2.2. modélisation robuste du bruit, a priori gaussien

Dans le cas de l'hypothèse gaussienne pour la distribution *a priori* dans l'espace propre, l'énergie à minimiser s'écrit :

$$\mathcal{J}(\mathbf{c}) = \sum_{n=1}^N \rho\left(\frac{w_n}{\sigma_\rho}\right) + \sum_{j=1}^J \frac{(c_j)^2}{\lambda_j} \quad (20)$$

Les algorithmes de minimisation de l'énergie  $\mathcal{J}$  restent similaires à (16) et (18), mise à part l'étape de calcul de  $\mathbf{c}^{(m+1)}$ , qui devient pour ARTUR :

$$\mathbf{c}^{(m+1)} = (\mathbf{U}^T \cdot \mathbf{B}^{A(m+1)} \cdot \mathbf{U} + \sigma_\rho^2 \Lambda^{-1})^{-1} \cdot \mathbf{U}^T \cdot \mathbf{B}^{A(m+1)} \cdot (\mathbf{y} - \boldsymbol{\mu})$$

et pour LEGEND :

$$\mathbf{c}^{(m+1)} = (\mathbf{I}_J + \sigma_\rho^2 \Lambda^{-1})^{-1} \cdot \mathbf{U}^T \cdot (\mathbf{y} - \boldsymbol{\mu} - \sigma_\rho \cdot \mathbf{b}^{L(m+1)})$$

L'estimée  $\hat{\mathbf{c}}$  est, à nouveau, la valeur obtenue à convergence. La mesure de ressemblance associée à ce détecteur robuste-gaussien (RG) s'écrit :

$$\text{RG}(\mathbf{y}) = \min_{\mathbf{c}} \left\{ \sum_{n=1}^N \rho\left(\frac{w_n}{\sigma_\rho}\right) + \sum_{j=1}^J \frac{(c_j)^2}{\lambda_j} \right\} \quad (21)$$

### 3.3. détection dans le cas non gaussien

Nous nous plaçons maintenant dans le cas le plus général, où ni la distribution du bruit, ni celle des coordonnées dans l'espace propre n'est considérée gaussienne. Dans ce cas,  $\mathcal{P}(\mathbf{c}|\mathcal{B})$  est modélisée à partir des images d'apprentissage, en utilisant des fenêtres de Parzen, comme décrit au paragraphe 2.6. L'estimation de  $\mathbf{c}$  au sens du MAP, d'après (11) :

$$\hat{\mathbf{c}}_{MAP} = \arg \max_{\mathbf{c}} \{\mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B}) \cdot \mathcal{P}(\mathbf{c}|\mathcal{B})\} \quad (22)$$

devient toutefois complexe dans ce cas, en raison de la forme de la distribution *a priori*. Nous avons donc recours à une seconde approximation, qui consiste à estimer tout d'abord  $\mathbf{c}$  au sens du maximum de vraisemblance :

$$\hat{\mathbf{c}}_{MV} = \arg \max_{\mathbf{c}} \{\mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B})\} \quad (23)$$

et à approcher ensuite la vraisemblance de l'image observée par :

$$\mathcal{P}(\mathbf{y}|\mathcal{B}) \propto \mathcal{P}(\mathbf{y}|\hat{\mathbf{c}}_{MV}, \mathcal{B}) \mathcal{P}(\hat{\mathbf{c}}_{MV}|\mathcal{B}) \quad (24)$$

Une telle approximation n'est véritablement justifiée que lorsque les pics de  $\mathcal{P}(\mathbf{y}, \hat{\mathbf{c}}_{MV}|\mathcal{B})$  coïncident avec ceux de  $\mathcal{P}(\hat{\mathbf{c}}|\mathcal{B})$ , ce qui n'est clairement pas le cas en général. Elle est toutefois licite dans notre cas, car les reconstructions obtenues par la méthode du maximum de vraisemblance sont déjà très satisfaisantes, en particulier avec un modèle de bruit robuste [2]. La présence du modèle *a priori* dans (24) permet d'introduire une contrainte très utile dans le calcul de  $\mathcal{P}(\mathbf{y}|\mathcal{B})$ , qui n'est pas prise en compte dans les méthodes actuelles, et qui permet d'améliorer grandement les résultats de détection, comme nous le verrons dans la partie consacrée aux résultats expérimentaux.

Ce type de détecteur sera dénommé robuste – non gaussien (RNG). Dans le cas le plus général où la distribution dans l'espace propre est modélisée par des fenêtres de Parzen à noyau

gaussien, la mesure de ressemblance s'écrit :

$$\text{RNG}(\mathbf{y}) = \min_{\mathbf{c}} \left\{ \frac{1}{2} \sum_{n=1}^N \rho \left( \frac{w_n}{\sigma_p} \right) \right\} - \log \left[ \sum_{k=1}^K \exp - \left[ \frac{\|\hat{\mathbf{c}} - \mathbf{c}^k\|^2}{2(\sigma_p)^2} \right] \right] \quad (25)$$

La variance  $\sigma_p$  permet de pondérer l'influence de l'*a priori* dans l'espace propre par rapport à la vraisemblance de l'erreur.

## 4. résultats expérimentaux

Les détecteurs sont évalués sur un ensemble de scènes synthétiques, où apparaissent des occurrences de nos objets d'intérêt, sur des fonds très texturés, avec des occlusions et parfois des leurres. Trois bases sont utilisées : la base COIL, la base AVG et la base A43, présentées au paragraphe 2.6. Les images de test ont été obtenues en insérant les objets sur divers fonds texturés, avec des occlusions parfois importantes (voir par exemple Figure 7). Les courbes COR, ou Caractéristiques Opérationnelles du Récepteur, sont calculées pour comparer les performances de nos détecteurs. Elles présentent le pourcentage de bonnes détections obtenues en fonction du pourcentage de fausses alarmes. Les positions exactes de nos objets sont connues. La détection, obtenue par simple seuillage de la carte de ressemblance, est considérée correcte lorsqu'elle se situe dans un 8-voisinage de la réponse, ce qui correspond à une tolérance de précision de 1 pixel. Le tableau 2 rappelle les différents détecteurs évalués dans ce paragraphe. Leur dénomination est liée aux hypothèses sur la distribution de l'erreur et celle dans le plan propre. A titre de comparaison, nous avons également réimplanté la méthode de Moghaddam et Pentland [14], s'appuyant sur un modèle gaussien.

Tableau 2. – Rappel des différents détecteurs proposés et des hypothèses sous-jacentes sur la distribution du bruit et l'*a priori* dans l'espace propre.

$\mathcal{P}(\mathbf{c} \mathcal{B})$ \diagdown $\mathcal{P}(\mathbf{y} \mathbf{c}, \mathcal{B})$	Uniforme	Gaussien	Non-Gaussien
Gaussien	GU <sub>(DFFS)</sub>	GG	–
Robuste	RU	RG	RNG

### 4.1. importance du modèle robuste de bruit

Cette série d'expériences est réalisée sur la base COIL dans le but de comparer l'hypothèse classique de distribution gaussien-

ne du bruit et l'hypothèse robuste, qui autorise la présence de données fausses dans les observations. On vérifie aussi l'intérêt de la méthode de centrage présentée au paragraphe 2.1. On considère ici un espace propre de dimension  $J = 5$ . On utilise un jeu de 21 scènes de test, présentant 57 occurrences de l'objet d'intérêt, parfois partiellement occultées, sur des fonds texturés (deux exemples sont donnés figure 7).

**Effet du centrage modifié.** Les courbes COR de la figure 8 comparent les performances des deux détecteurs classique GU et robuste RU suivant la méthode de centrage utilisée pour l'apprentissage. On remarque que la nouvelle méthode de *centrage modifié* (cf. paragraphe 2.1), permet d'améliorer notablement les résultats du détecteur classique et, plus nettement encore, ceux du détecteur robuste. Ceci a été vérifié pour tous nos détecteurs et dans toutes nos expériences, aussi les résultats suivants ont-ils été réalisés en utilisant systématiquement cette nouvelle méthode de centrage des données.

**Apport du terme de vraisemblance robuste de l'erreur  $\mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B})$ .** La figure 7 présente les cartes de vraisemblance obtenues par les détecteurs GU, RU et RNG sur deux scènes test. A titre indicatif, sur un PC AMD 750 MHz, le temps moyen nécessaire au calcul d'une carte de vraisemblance, à l'aide d'un programme C non optimisé (compilé en mode « debug »), est de l'ordre de 1 minute pour le détecteur GG, de 5 minutes pour RU et RNG et de 11 minutes pour RG. Les pixels hors cadre indiquent les positions où l'on ne peut extraire d'observation complète. Visuellement, le détecteur RU produit des pics plus marqués que GU, permettant une meilleure localisation des objets d'intérêt. Le détecteur RNG, qui intègre un *a priori* non gaussien sur les variables latentes  $\mathbf{c}$ , fournit une carte de ressemblance similaire à celle de RU.

Cette impression visuelle est confirmée par les courbes COR de la figure 9 qui permettent de comparer les performances des détecteurs sur l'ensemble des scènes de test. Les performances du détecteur GG s'avèrent identiques à celles du détecteur classique de Moghaddam et Pentland [14] dans nos expériences, nous ne présenterons par la suite qu'une seule courbe pour les deux. Comme on pouvait s'y attendre, le détecteur RU obtient des taux de bonne détection significativement supérieurs à ceux du détecteur à modèle de bruit gaussien GU. Par ailleurs, les hypothèses gaussienne et uniforme *a priori* sur les variables latentes  $\mathbf{c}$  conduisent à des résultats semblables. Ceci n'a rien de surprenant, puisque la distribution dans l'espace propre n'est ni gaussienne, ni uniforme, comme l'illustre la figure 5. Un *a priori* non gaussien est pris en compte dans le cas du détecteur RNG (cf. équation 25), qui combine modèle de bruit non gaussien et estimation par fenêtres de Parzen de la distribution dans l'espace propre. On peut constater que l'introduction de cet *a priori* dans le détecteur RNG améliore encore légèrement les résultats de RU (cf. figures 7 et 9). Au total, le modèle RNG conduit aux meilleurs résultats, surpassant les modèles seulement basés sur des hypothèses gaussiennes.

## 4.2. importance du modèle *a priori* (I)

Cette seconde expérience montre qu'une modélisation adaptée de la distribution dans l'espace propre peut grandement améliorer les performances de détection.

Les tests ont été réalisés en utilisant la base AVG, pour  $J = 20$ , avec 18 scènes en couleurs présentant 29 occurrences de panneaux triangulaires. Deux d'entre elles sont présentées figure 10. Une spécificité de ce test est que l'image moyenne de la base AVG a été introduite dans  $I_{17}$  (motif circulaire situé sur la gauche de l'image).

L'image moyenne ne ressemble évidemment pas à un panneau blanc triangulaire. Cependant, elle représente un leurre pour les détecteurs GU, GG et RU qui ne prennent pas en compte la forme particulière de la distribution *a priori* dans l'espace propre. En effet, l'image moyenne appartient par définition au plan propre, et minimise donc l'erreur de reconstruction, quelle que soit l'hypothèse de distribution choisie pour le bruit (robuste ou gaussienne). De fait, comme le montre la figure 10, l'image moyenne est parfaitement détectée dans l'image  $I_{17}$  par RU et ce, avec une plus forte valeur de ressemblance que les deux panneaux également présents dans la scène ! Le détecteur robuste RU est également leurré par les zones herbeuses de l'image  $I_2$ . Les performances des différents détecteurs sont présentées à la figure 11 : les méthodes classiques (GU, GG ou Moghaddam et Pentland) ont des performances généralement identiques et plutôt médiocres. L'utilisation d'un modèle robuste de vraisemblance (détecteur RU) n'améliore que peu les résultats.

Une amélioration significative est, par contre, obtenue en associant un modèle *a priori* adéquat, à un modèle de bruit robuste. Comme cette base particulière d'apprentissage est obtenue à partir des rotations d'un objet unique dans le plan image, il n'est pas nécessaire d'avoir recours aux fenêtres de Parzen pour modéliser la densité de probabilité de  $\mathbf{c}$ . En fait, l'expression analytique de  $\mathcal{P}(\mathbf{c}|\mathcal{B})$  est connue dans ce cas particulier [17] : les valeurs propres sont doubles ( $\lambda_{2j} = \lambda_{2j-1}$ ) et, dans chaque plan formé par une paire de vecteurs propres, les coordonnées des images d'apprentissage se répartissent selon un cercle dont le rayon est donné par  $R^2 = \lambda_{2j} + \lambda_{2j-1}$ . Plus précisément,

$$\mathcal{P}(\mathbf{c}|\mathcal{B}) \propto \prod_{j=1}^{\frac{J}{2}} \exp - \left[ \frac{1}{2\gamma} |(c_{2j-1})^2 + (c_{2j})^2 - \lambda_{2j} - \lambda_{2j-1}| \right]. \quad (26)$$

On peut le vérifier sur la figure 6. C'est aussi ce qui explique les couronnes circulaires observables sur la partie droite de la figure 5. L'apprentissage des rotations dans le plan image est un cas assez particulier, qui a été utilisé avec succès en détection pour le guidage de robots à partir d'images panoramiques [10]. Dans ce contexte, le détecteur RNG prend la forme suivante :

$$\text{RNG}(\mathbf{y}) = \min_{\mathbf{c}} \left\{ \sum_{n=1}^N \rho \left( \frac{w_n}{\sigma_p} \right) \right\} + \frac{1}{\gamma} \sum_{j=1}^{\frac{J}{2}} |(\hat{c}_{2j-1})^2 + (\hat{c}_{2j})^2 - \lambda_{2j} - \lambda_{2j-1}| \quad (27)$$

où le paramètre  $\gamma$  joue le même rôle que  $\sigma_p$  dans le cas général. Comme on pouvait s'y attendre, les performances du modèle complet sont de loin les meilleures sur ce test (cf. figures 10 et 11).

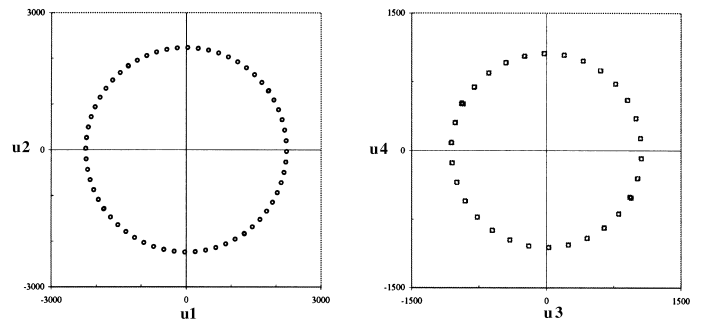


Figure 6. – Distribution des variables latentes  $\mathbf{c}$  dans les deux premiers plans de l'espace propre pour la base AVG. La forme en couronne est typique de l'apprentissage de la rotation dans le plan image [10].

## 4.3. importance du modèle *a priori* (II)

Cette série d'expérience est une autre illustration de l'importance d'une modélisation fine de la distribution des coordonnées dans l'espace propre, dans le cas où la densité sous-jacente est de forme plus générale. Les expériences de détection sont réalisées avec la base A43,  $J = 30$ , sur un jeu de 27 scènes test en couleurs contenant 58 occurrences des panneaux triangulaires (voir exemples figure 12).

Les cartes de vraisemblance obtenues avec le détecteur RU sur les scènes  $I_2$  et  $I_{17}$  paraissent visuellement peu exploitables : les positions des occurrences des objets d'intérêt ne se distinguent pas facilement. Cette impression est confirmée par les courbes COR représentées figure 13 pour les détecteurs GU et RU. Les résultats obtenus avec un *a priori* gaussien (détecteurs GG et RG) sont identiques et ne sont pas représentés ici. Encore une fois, les hypothèses *a priori* uniforme et gaussienne, ne sont de toute évidence pas adaptées à la distribution réelle des variables latentes (cf. figure 5), ce qui explique l'effondrement des performances de ces détecteurs classiques. L'utilisation d'une vraisemblance robuste améliore légèrement les résultats, mais ceux-ci demeurent médiocres : pour détecter 80 % des objets présents dans les scènes, il faut accepter environ 68 % de fausses alarmes.

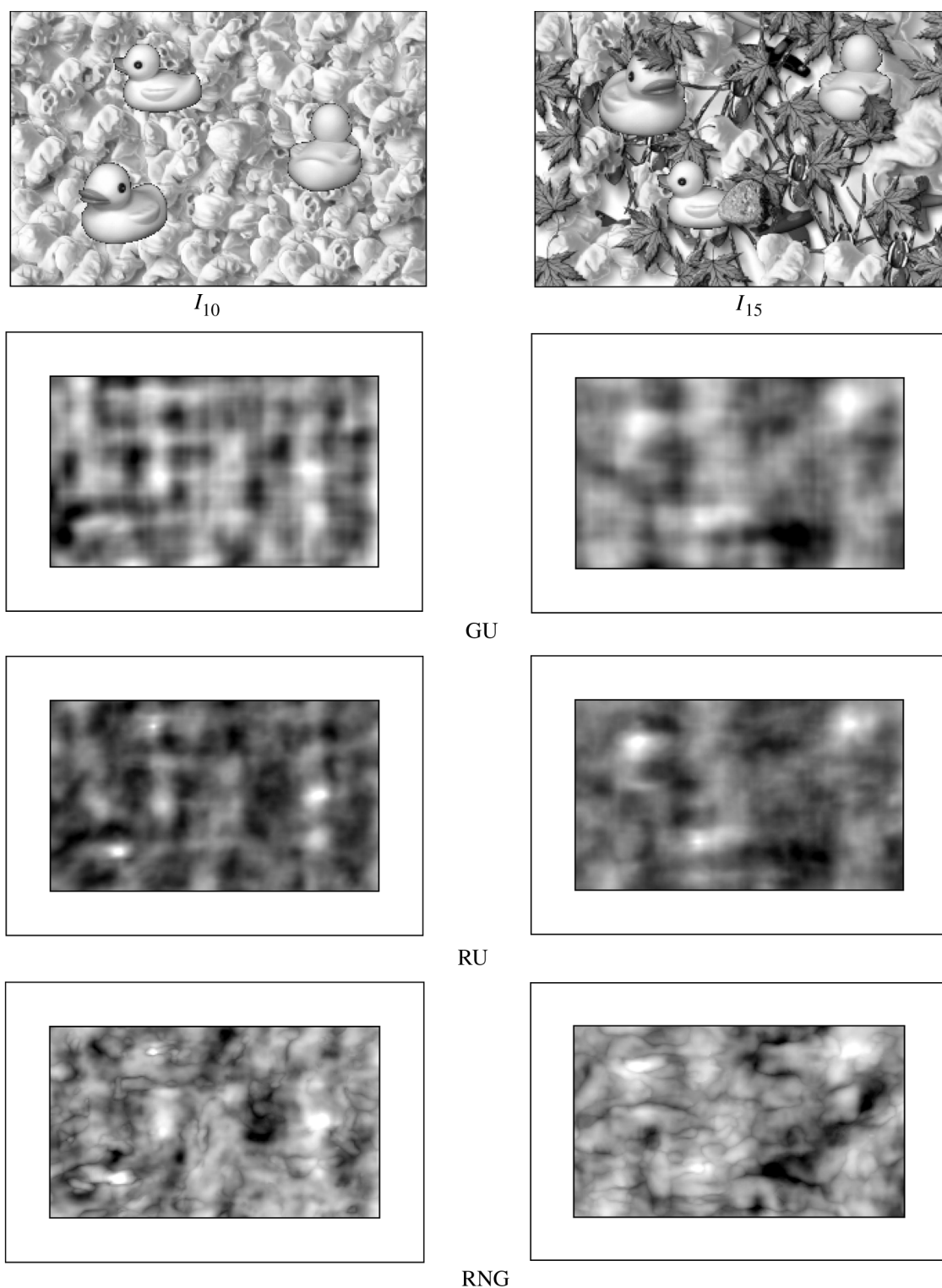


Figure 7. – Exemples de scènes test et cartes de log-vraisemblance obtenues avec les détecteurs GU et RU et le modèle complet RNG, base COIL,  $J = 5$ . Les plus fortes vraisemblances apparaissent comme des points clairs.

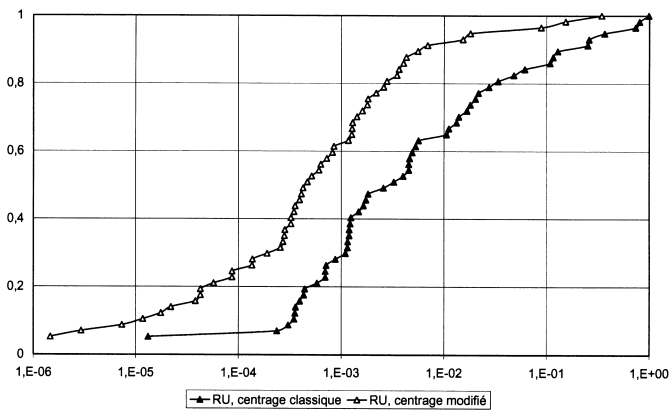
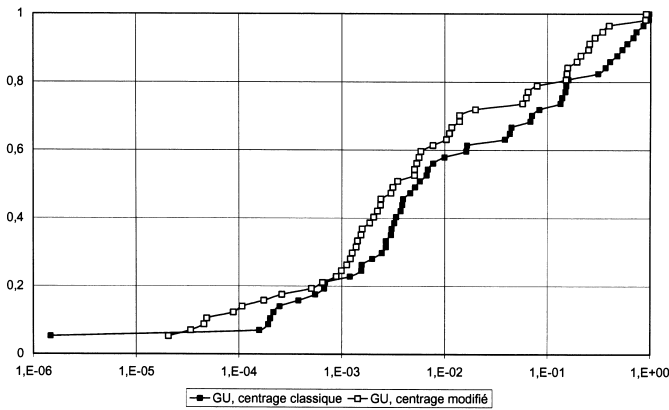


Figure 8. – Effet du centrage : courbes COR calculées avec les détecteurs GU et RU, base COIL  $J = 5$ .

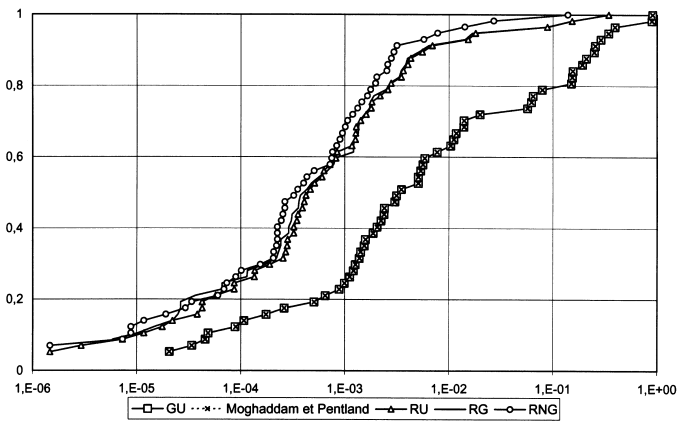


Figure 9. – Courbes COR obtenues avec les détecteurs classiques gaussiens (GU, GG ou Moghaddam et Pentland) ; modèle robuste (RU, RG) ; modèle complet robuste – non gaussien (RNG). Base COIL,  $J = 5$  (voir texte).

En ce qui concerne le détecteur RNG, la distribution de  $\mathbf{c}$  dans l'espace propre est modélisée par la méthode des fenêtres de Parzen, comme cela est présenté à la section 3.3. Nous utilisons l'approximation de l'équation (24) : la ressemblance est calculé à l'aide de l'estimée de la variable latente,  $\mathbf{c}$ , au sens du maxi-

um de vraisemblance robuste. De plus, un poids important est donné au terme *a priori*.

Les cartes de détection calculées avec ce modèle sont présentées figure 12. Elles permettent de localiser facilement les occurrences de nos objets d'intérêt, qui apparaissent comme des points brillants. La figure 13 présente les courbes COR correspondantes. On y remarque immédiatement l'amélioration apportée par le modèle complet RNG : plus de 70 % des objets d'intérêt sont détectés avant que les premières fausses alarmes n'apparaissent. Ces performances sont nettement supérieures à celles des détecteurs classiques. Une modélisation fine de la distribution des coefficients dans l'espace propre est donc essentielle.

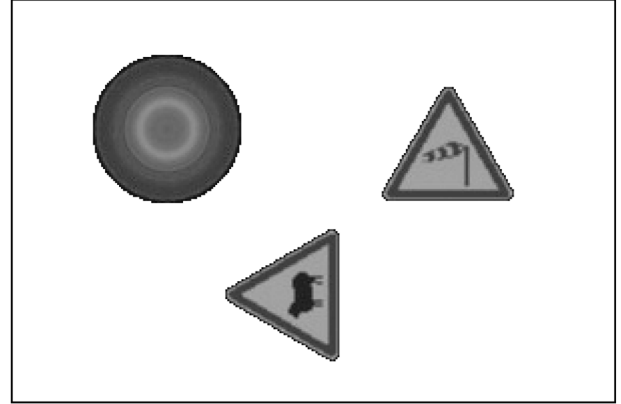
## 5. conclusion

Dans cet article, nous avons présenté une approche bayésienne générale de la détection d'objets d'intérêt par modèle d'apparence. Cette approche permet une description unifiée des méthodes classiques déjà proposées dans ce domaine. Elle ouvre également la voie à la définition de nouveaux détecteurs permettant de prendre en compte des classes complexes d'objets et de gérer des scènes présentant des fonds texturés, des occultations ou des instances dégradées des objets modélisés. Le cadre proposé, ainsi que les approximations introduites, permettent de gérer des distributions non gaussiennes, à deux niveaux : distribution des images d'apprentissage (modèle *a priori*) et distribution du bruit d'observation. Nous montrons que la prise en compte de distributions non gaussiennes pour les images d'apprentissage permet, dans certaines situations critiques, d'améliorer significativement les performances des détecteurs, par rapport aux méthodes faisant actuellement référence [2, 14]. Des modèles non gaussiens de bruit d'observation, conduisent, plus classiquement, à des estimations robustes, moins sensibles aux dégradations des objets modélisés. Chacune de ces composantes a un effet positif sur les résultats de détection et, comme l'ont montré les exemples présentés, aucune des deux n'est à elle seule suffisante : c'est leur combinaison qui permet d'obtenir les meilleurs résultats.

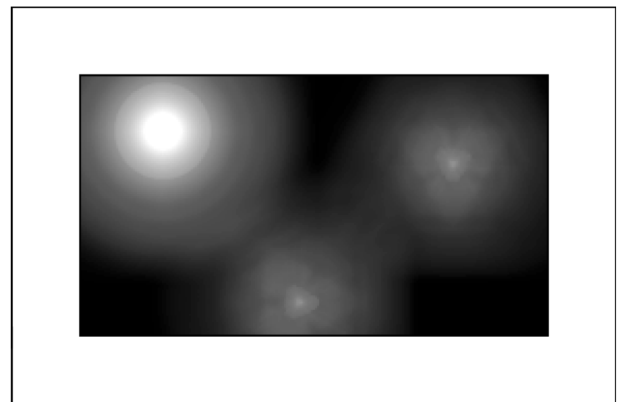
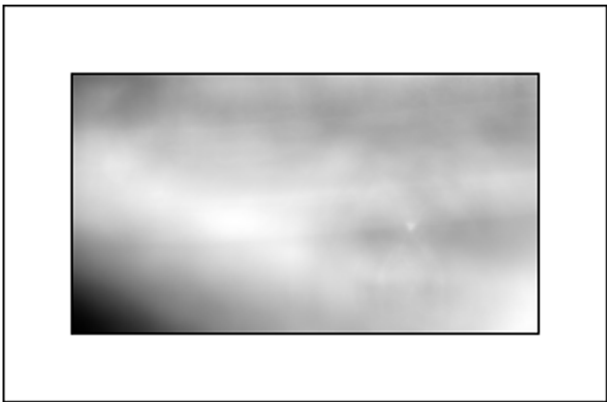
L'étude expérimentale proposée, menée sur des données synthétiques, nous a permis d'illustrer les contributions des différents éléments du modèle proposé. La validation de celui-ci sur des images réelles est une perspective de travail intéressante, qui fera l'objet de publications ultérieures. D'autre part, même si l'approximation (24) s'avère justifiée en pratique, il est possible de prendre en compte le modèle complet de manière exacte, à l'aide d'une technique originale, utilisant des développements récents [4], et que nous présentons dans [24]. Enfin, le temps de calcul inhérent à ce type d'approches demeure relativement important. Des extensions plus rapides pourraient être imaginées en utilisant, par exemple, les résultats de [8] ou en considérant des approches plus locales [19].



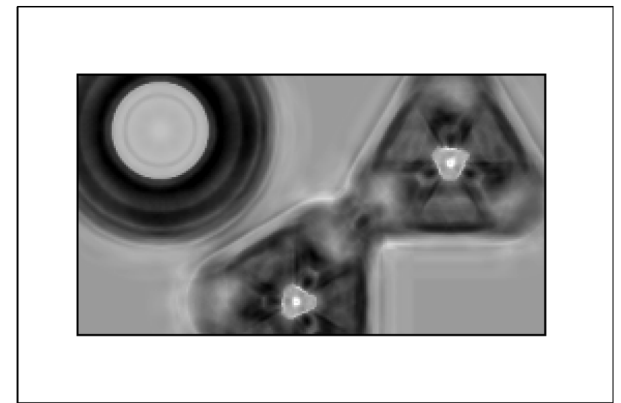
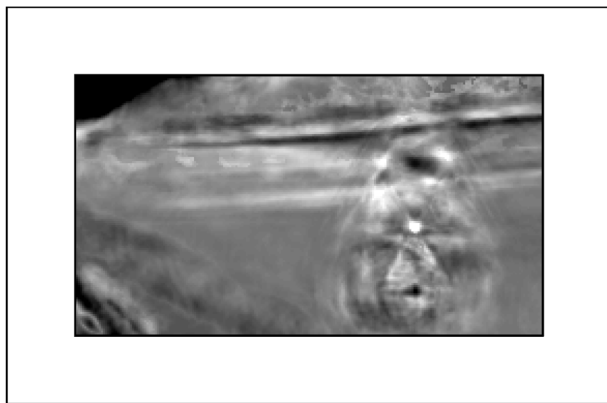
$I_2$



$I_{17}$



RU

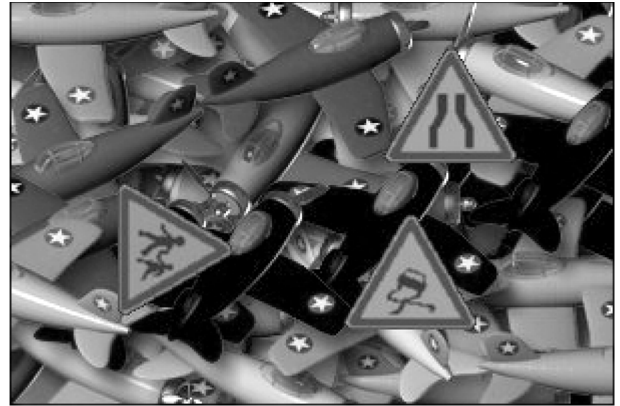


RNG

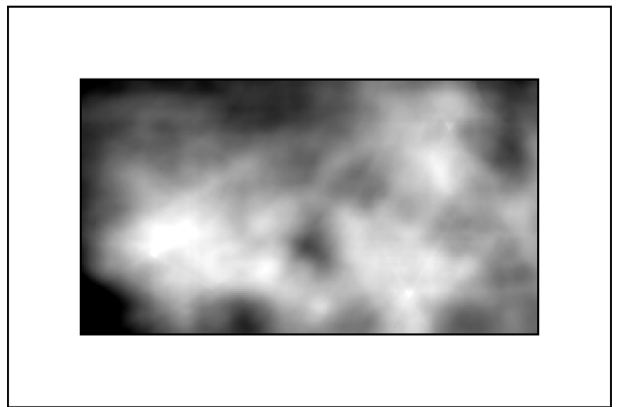
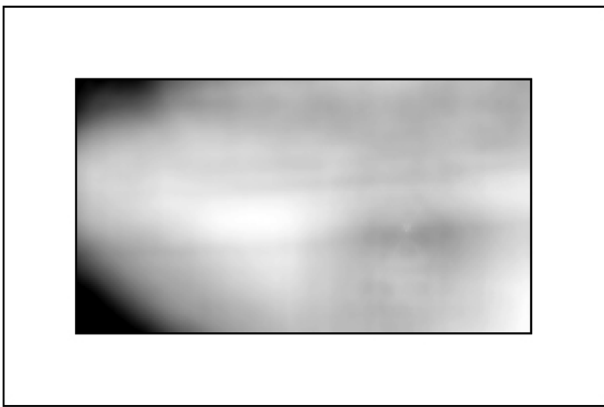
Figure 10. – Cartes de log-vraisemblance (les plus fortes vraisemblance apparaissent en clair) obtenues sur les images  $I_2$  et  $I_{17}$ , base AVG,  $J = 20$ .



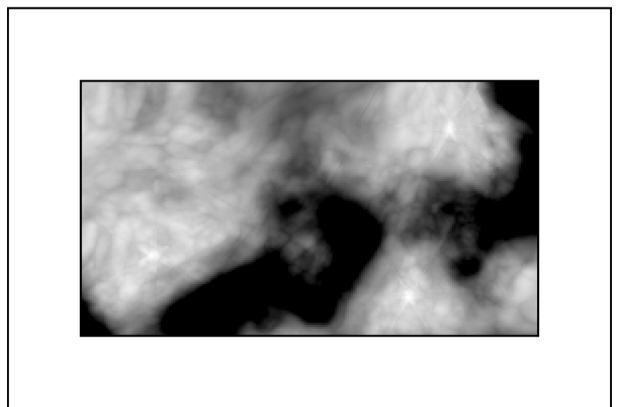
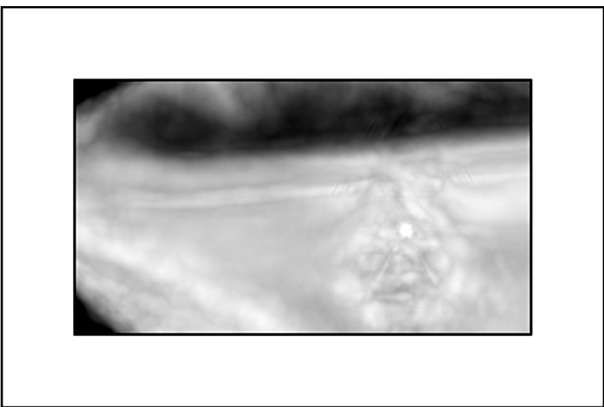
$I_2$



$I_{17}$



RU



RNG

Figure 12. – Exemples de cartes de vraisemblance (les plus fortes vraisemblances apparaissent en clair), base A43,  $J = 30$ .

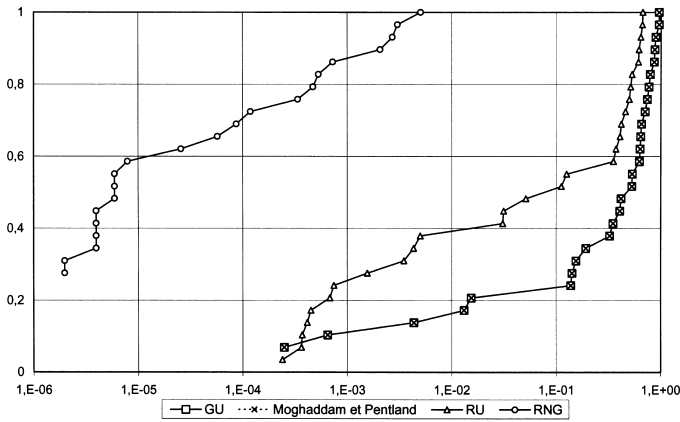


Figure 11. – Courbes COR obtenues pour les détecteurs gaussiens (GU, GG ou Moghaddam et Pentland) le détecteur robuste RU et le modèle complet robuste – non gaussien (RNG), base AVG,  $J = 20$ .

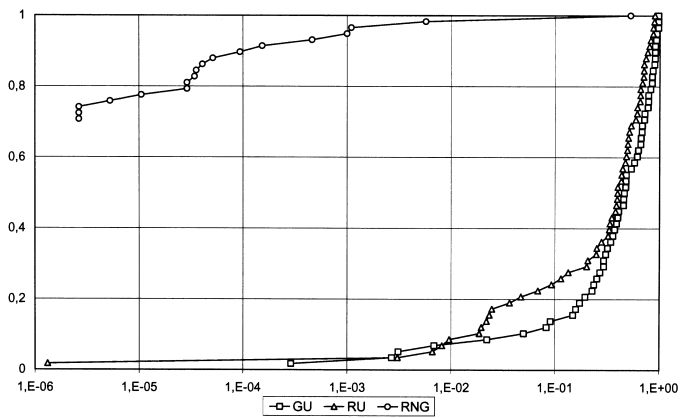


Figure 13. – Courbes COR obtenues avec les détecteurs GU et RU ; modèle complet RNG. Base A43,  $J = 30$ .

## Remerciements

Ce travail a été soutenu par une bourse de thèse du Laboratoire Central des Ponts et Chaussées.

## BIBLIOGRAPHIE

- [1] J.M. Bernardo et A.F. Smith. *Bayesian Theory*. Wiley Series in Probability and Statistics, New York, 2000.
- [2] M. J. Black et A. D. Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. *International Journal on Computer Vision*, 26(1):63-84, janvier 1998.
- [3] P. Charbonnier, L. Blanc-Féraud, G. Aubert, et M. Barlaud. Two deterministic half quadratic regularization algorithms for computed imaging. Dans *International Conference on Image Processing*, pages 168-172, Austin, USA, 1994.
- [4] D. Comaniciu et P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603-619, mai 2002.
- [5] R. Dahyot. *Analyse d'images séquentielles de scènes routières par modèles d'apparence pour la gestion du réseau routier*. Thèse de doctorat, Université Louis Pasteur, Strasbourg, France, novembre 2001.
- [6] R. Dahyot, P. Charbonnier et F. Heitz. Robust visual recognition of colour images. Dans *IEEE conference on Computer Vision and Pattern Recognition*, volume 1, pages 685-690, Hilton Head Island, USA, juin 2000.
- [7] R.O. Duda, P.E. Hart et D.G. Stork. *Pattern Classification (2nd. ed.)*. John Wiley and Sons, New York, 2001.
- [8] R. Hamdan, F. Heitz, et L. Thoraval. Modèles probabilistes d'apparence : une représentation approchée de faible complexité. *Traitement de Signal*, 18(3):147-160, 2001.
- [9] P.J. Huber. *Robust Statistics*. John Wiley and Sons, New York, 1981.
- [10] M. Jogan et A. Leonardis. Parametric eigenspace representations of panoramic images. Dans *International Conference on Advanced Robotics 2001 – Omnidirectional Vision Applied to robotic orientation and nondestructive testing (NDT)*, pages 31-36, Budapest, Hongrie, août 2001. IEEE Computer Society.
- [11] D.J.C. MacKay. Probable network and plausible predictions – a review of practical Bayesian methods for supervised neural networks. *Network*, pages 469-505, 1995.
- [12] B. Moghaddam. Principal manifolds and Bayesian subspaces for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):780-788, juin 2002.
- [13] B. Moghaddam et A. Pentland. Probabilistic visual learning for object detection. Dans *The 5th International Conference on Computer Vision*, Cambridge, USA, juin 1995.
- [14] B. Moghaddam et A. Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696-710, juillet 1997.
- [15] H. Murase et S. K. Nayar. Visual learning and recognition of 3-D objects from appearance. *International Journal of Computer Vision*, 14(1):5-24, janvier 1995.
- [16] S. A. Nene, S. K. Nayar et H. Murase. Columbia object image library (coil-20). Rapport technique CUCS-005-96, Department of Computer Science, Columbia University, 1996.
- [17] R. H. Park. Comments on optimal approximation of uniformly rotated images: Relationship between Karhunen-Loeve expansion and discrete cosine transform. *IEEE Transactions on Image Processing*, 11(3):332-334, mars 2002.
- [18] W. H. Press, S. A. Teukolsky, W. T. Vetterling, et B. P. Flannery. *Numerical Recipes in C - The Art of Scientific Computing*. Cambridge University Press, 1995.
- [19] C. Schmid et R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530-534, 1997.



- [20] H. Schneiderman. *A Statistical Approach to 3D Object Detection Applied to Faces and Cars*. Thèse de doctorat, Carnegie Mellon University, Pittsburg USA, 2000.
- [21] M.E. Tipping et C.M. Bishop. Mixtures of probabilistic principal component analysers. *Neural Computation*, 11(2):443-482, 1999.
- [22] M.E. Tipping et C.M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B*, 61(3):611-622, 1999.
- [23] M. Turk et A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71-86, 1991.
- [24] T. Vik, F. Heitz et P. Charbonnier. Mean shift-based Bayesian image reconstruction into visual subspace. Dans *IEEE International Conference on Image Processing (accepté pour publication)*, Barcelone, Espagne, septembre 2003.

*Manuscrit reçu le 23 septembre 2002*

## LES AUTEURS

### Rozenn DAHYOT



Ingénieur de l'ENSPS (1998), Rozenn Dahyot a préparé, au Laboratoire Régional des Ponts et Chaussées de Strasbourg, une thèse portant sur l'analyse de scènes routières par modèles d'apparence et obtenu en novembre 2001 le titre de Docteur de l'Université Louis Pasteur de Strasbourg. Elle est actuellement en Post-doctorat à Trinity College à Dublin. Ses domaines d'intérêt portent sur l'applications de méthodes statistiques pour l'analyse multimedia.

### Fabrice HEITZ



Ingénieur ENST Bretagne (1984), Docteur Télécom Paris (1988), Fabrice Heitz a été de 1988 à 1994 Chargé de Recherches INRIA à l'IRISA (Rennes). Il est actuellement Professeur de Traitement du Signal et des Images à l'École Nationale Supérieure de Physique de Strasbourg (ENSPS). Il anime un groupe de recherche en analyse multi-images au sein du LSIIT (Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection, UMR CNRS 7005). Ses domaines d'intérêt incluent la modélisation statistique et les modèles déformables appliqués à l'analyse multi-images et à l'imagerie médicale.

### Pierre CHARBONNIER



Ingénieur ESSI (1991), Docteur de l'Université de Nice - Sophia Antipolis (1994), Pierre Charbonnier est actuellement Chargé de Recherches au Laboratoire Régional des Ponts et Chaussées de Strasbourg (ERA 27 du Laboratoire Central des Ponts et Chaussées). Ses domaines d'intérêt incluent les modèles statistiques et les modèles déformables appliqués à l'analyse multi-images.