

N°d'ordre :

THÈSE

présentée à

L'Université Louis Pasteur - Strasbourg I

pour obtenir le titre de

Docteur de l'Université Louis Pasteur - Strasbourg I

Domaine : Electronique, Electrotechnique, Automatique

Spécialité : Traitement d'images et vision par ordinateur

par

Rozenn DAHYOT

**Analyse d'images séquentielles de scènes routières  
par modèles d'apparence pour la gestion du réseau routier**

Jury composé de :

M. Jean GALLICE	Rapporteur externe
M. Jack-Gérard POSTAIRE	Rapporteur externe
M. Ernest HIRSCH	Rapporteur interne
M. Fabrice HEITZ	Directeur de thèse
M. Pierre CHARBONNIER	Examinateur
M. Jean-Philippe TAREL	Invité



# Remerciements

Cette thèse est l'aboutissement de travaux réalisés au Laboratoire Régional des Ponts et Chaussées de Strasbourg (LRS) portant sur l'analyse d'images de scènes routières pour la gestion du réseau routier. Ces recherches ont été réalisées sous la direction de M. Pierre CHARBONNIER, chargé de recherche au groupe Métrologie au LRS, et M. Fabrice HEITZ professeur à l'Ecole Nationale Supérieure de Physique de Strasbourg. Je souhaite leur exprimer ma gratitude pour m'avoir encadré avec disponibilité et clairvoyance lors de mon stage de DEA et de ces trois années de thèse.

Je remercie M. Jean GALLICE professeur à l'Université Blaise Pascal de Clermont-Ferrand, M. Jack-Gérard POSTAIRE professeur à l'Université des Sciences et Technologies de Lille, et M. Ernest HIRSCH professeur à l'Ecole Nationale Supérieure de Physique de Strasbourg d'avoir accepté d'examiner ce travail de thèse. Je suis particulièrement reconnaissante aux membres du comité de suivi présidé par M. Pierre-Yves TEXIER, directeur de recherche au Laboratoire Central des Ponts et Chaussées de Paris (LCPC), Jean-Marie CAUSSIGNAC (LCPC), Jean PEYBERNARD (LCPC), Yannick GUILLARD responsable du groupe métrologique (LRS), Jean-Phillipe TAREL (INRETS) et Guy GRATIA (LRPC Lyon), pour leurs conseils avisés et leurs encouragements qu'ils m'ont témoignés tout au long de cette étude.

Je remercie chaleureusement l'ensemble du personnel du LRS et plus particulièrement les personnes du groupe Métrologie pour l'ambiance conviviale qu'ils ont su créer. Je terminerai en remerciant ma famille pour leur soutien de tous les instants.



# Table des matières

<b>Remerciements</b>	<b>3</b>
<b>Table des matières</b>	<b>5</b>
<b>Liste des tableaux</b>	<b>13</b>
<b>Liste des figures</b>	<b>15</b>
<b>Notations</b>	<b>19</b>
<b>1 Introduction</b>	<b>21</b>
1.1 Contexte de l'étude . . . . .	21
1.2 Détection et reconnaissance d'objets . . . . .	22
1.3 Sommaire des chapitres de la thèse . . . . .	23
<b>2 Etat de l'art</b>	<b>25</b>
2.1 Caractéristiques images . . . . .	25
2.1.1 La couleur . . . . .	26
2.1.2 La forme . . . . .	27
2.1.2.1 Représentation <i>surface</i> . . . . .	27
2.1.2.2 Représentation <i>contour</i> . . . . .	27
2.1.3 Apparence . . . . .	28
2.1.3.1 Représentation globale de l'apparence . . . . .	29
2.1.3.2 Représentation locale de l'apparence . . . . .	33
2.1.4 La texture . . . . .	33
2.1.4.1 Approches par filtrage spatio-fréquentiel . . . . .	34
2.1.4.2 Modèles autorégressifs . . . . .	34
2.1.4.3 Décomposition de Wold . . . . .	34
2.1.5 Le mouvement . . . . .	35

2.1.6	Description de vidéo . . . . .	35
2.1.7	Topologie spatiale . . . . .	36
2.1.8	Conclusion . . . . .	36
2.2	Application à la détection et la reconnaissance . . . . .	37
2.2.1	Calcul de similarité . . . . .	38
2.2.2	Méthodes statistiques . . . . .	38
2.3	Systèmes de transports intelligents . . . . .	39
2.3.1	Images monochromes . . . . .	40
2.3.2	Images en couleur . . . . .	40
2.3.3	Remarques . . . . .	44
2.4	Conclusion . . . . .	44
<b>I</b>	<b>Apparence locale</b>	<b>47</b>
<b>3</b>	<b>Mesures locales d'apparence</b>	<b>51</b>
3.1	Mesures locales d'apparence . . . . .	52
3.1.1	Information couleur . . . . .	52
3.1.2	Mesures locales de forme . . . . .	52
3.2	Estimation de la distribution des mesures locales . . . . .	53
3.2.1	Histogrammes classiques . . . . .	54
3.2.2	Histogrammes interpolés . . . . .	55
3.2.3	Méthodes des noyaux . . . . .	56
3.2.4	Comparaison entre les histogrammes calculés avec $\phi_u$ et $\phi_l$ . . . . .	57
3.3	Histogrammes de forme . . . . .	58
3.3.1	Segmentation des contours . . . . .	58
3.3.2	Histogramme pondéré . . . . .	59
3.3.3	Histogramme des mesures $(\theta, \alpha, N)$ . . . . .	59
3.4	Conclusion . . . . .	59
<b>4</b>	<b>Pré-détection non supervisée</b>	<b>61</b>
4.1	Détection de changements statistiques . . . . .	62
4.1.1	Hypothèses relatives à $\mathcal{B}$ et $\bar{\mathcal{B}}$ . . . . .	62
4.1.2	Extracteurs d'évènements . . . . .	63
4.2	Détection par rétroprojection bayésienne . . . . .	64
4.3	Résultats de pré-détection . . . . .	67
4.3.1	Choix du coefficient $\gamma$ . . . . .	67
4.3.2	Pré-détection par mesures de couleur . . . . .	68

## TABLE DES MATIÈRES

7

4.3.3	Pré-détection par mesures de forme . . . . .	70
4.3.4	Fusion d'information . . . . .	71
4.3.5	Résultats quantitatifs sur 5 km . . . . .	72
4.4	Conclusion . . . . .	74
<b>II</b>	<b>Apparence Globale</b>	<b>77</b>
<b>5</b>	<b>Mesures globales d'apparence</b>	<b>81</b>
5.1	Analyse en Composantes Principales . . . . .	82
5.1.1	Estimation classique de la moyenne et de la matrice de covariance . . . . .	82
5.1.1.1	Estimateur classique de la moyenne : $\hat{\mu}^c$ . . . . .	82
5.1.1.2	Estimateur classique de la matrice d'autocovariance : $\hat{\Sigma}^c$ . . . . .	83
5.1.2	Nouveaux estimateurs de la moyenne et de la matrice de covariance . . . . .	83
5.1.2.1	Nouvel estimateur de la moyenne : $\hat{\mu}^n$ . . . . .	84
5.1.2.2	Nouvel estimateur de la matrice d'autocovariance : $\hat{\Sigma}^n$ . . . . .	84
5.1.2.3	Remarque . . . . .	85
5.1.3	Extraction des Composantes Principales . . . . .	85
5.1.3.1	Valeurs Propres . . . . .	86
5.1.3.2	Vecteurs Propres . . . . .	87
5.1.3.3	Conclusion . . . . .	90
5.2	Distribution des images d'apprentissage $\mathcal{P}(x \mathcal{B})$ . . . . .	91
5.2.1	Distribution dans l'espace orthogonal $F^\perp$ . . . . .	93
5.2.2	Distribution dans l'espace propre $F$ . . . . .	94
5.2.2.1	Hypothèses simples : distributions uniforme et gaussienne . . . . .	94
5.2.2.2	Distributions particulières . . . . .	94
5.2.3	Distribution $\mathcal{P}(x \mathcal{B})$ des images d'apprentissage de $\mathcal{B}$ . . . . .	98
5.3	Conclusion . . . . .	99
<b>6</b>	<b>Reconnaissance robuste</b>	<b>101</b>
6.1	Approches statistiques . . . . .	103
6.1.1	Approche classique : estimateur du maximum de vraisemblance . . . . .	103
6.1.1.1	Bruit gaussien . . . . .	104
6.1.1.2	Hypothèse robuste de la distribution du bruit . . . . .	104
6.1.2	Approche Bayésienne . . . . .	104
6.1.2.1	Hypothèses : distribution uniforme des coordonnées $c$ et distribution gaussienne du bruit $w$ . . . . .	105

6.1.2.2	Hypothèses : distribution uniforme des coordonnées $\mathbf{c}$ et distribution robuste du bruit $\mathbf{w}$ . . . . .	105
6.1.2.3	Hypothèses : distribution gaussienne des coordonnées $\mathbf{c}$ et distribution gaussienne du bruit $\mathbf{w}$ . . . . .	105
6.1.2.4	Hypothèses : distribution gaussienne des coordonnées $\mathbf{c}$ et distribution robuste du bruit $\mathbf{w}$ . . . . .	106
6.1.3	Remarques . . . . .	106
6.2	Minimisation d'énergies non quadratiques . . . . .	107
6.2.1	Cas des images en niveau de gris . . . . .	107
6.2.1.1	Théorie semi-quadratique . . . . .	107
6.2.1.2	Approche de Huber . . . . .	109
6.2.1.3	Lien entre l'approche de Huber et la théorie semi-quadratique . . . . .	112
6.2.1.4	Vitesse des deux algorithmes de minimisation . . . . .	113
6.2.2	Généralisation à des images couleurs ou multi-bandes . . . . .	113
6.2.2.1	ARTUR . . . . .	114
6.2.2.2	LEGEND . . . . .	114
6.2.3	Algorithmes d'estimation . . . . .	115
6.2.3.1	Estimation de $\hat{\mathbf{c}}_{MV}$ . . . . .	115
6.2.3.2	Estimation de $\hat{\mathbf{c}}_{RMV}$ . . . . .	115
6.2.3.3	Estimation de $\hat{\mathbf{c}}_{MAP}$ . . . . .	116
6.2.3.4	Estimation de $\hat{\mathbf{c}}_{RMAP}$ . . . . .	116
6.2.4	Conclusion . . . . .	117
6.3	Paramètre d'échelle $\sigma_\rho$ et estimation en continuation . . . . .	117
6.3.1	Fonctions $\rho$ . . . . .	117
6.3.2	Paramètre d'échelle $\sigma_\rho$ . . . . .	118
6.3.2.1	Estimation de $\sigma_\rho$ à partir de l'observation $\mathbf{y}$ . . . . .	119
6.3.2.2	Utilisation de l'apprentissage pour fixer le paramètre d'échelle <i>a priori</i> . . . . .	120
6.3.3	Estimation en continuation . . . . .	121
6.4	Résultats de reconnaissance . . . . .	123
6.4.1	Images en niveaux de gris . . . . .	123
6.4.1.1	Paramètre d'échelle fixé <i>a priori</i> . . . . .	124
6.4.1.2	Estimation conjointe $(\mathbf{c}, \sigma_\rho)$ . . . . .	128
6.4.2	Images en couleurs . . . . .	128
6.4.2.1	Résultats de reconnaissance . . . . .	129
6.4.2.2	Conclusion sur l'expérience de reconnaissance en couleur . . . . .	132
6.5	Conclusion . . . . .	132

<b>7 Détection robuste</b>	<b>135</b>
7.1 Méthodes classiques de détection . . . . .	136
7.1.1 Distance From Feature Space . . . . .	136
7.1.2 Vraisemblance de Moghaddam et Pentland . . . . .	137
7.1.3 Remarques . . . . .	137
7.2 Nouvelles méthodes de détection . . . . .	138
7.2.1 Hypothèses : distribution uniforme des coordonnées $c$ et distribution gaussienne du bruit $w$ . . . . .	139
7.2.2 Hypothèses : distribution uniforme des coordonnées $c$ et distribution robuste du bruit $w$ . . . . .	140
7.2.3 Hypothèses : distribution gaussienne des coordonnées $c$ et distribution gaussienne du bruit $w$ . . . . .	140
7.2.4 Hypothèses : distribution gaussienne des coordonnées $c$ et distribution robuste du bruit $w$ . . . . .	141
7.2.5 Utilisation des pondérations . . . . .	141
7.2.6 Remarques . . . . .	142
7.3 Expériences de détection . . . . .	143
7.3.1 Images en niveaux de gris : base COIL . . . . .	143
7.3.1.1 DéTECTEURS classiques . . . . .	143
7.3.1.2 DéTECTEURS robustes . . . . .	143
7.3.1.3 Remarques . . . . .	146
7.3.2 Expérience de détection avec des images en couleurs . . . . .	148
7.3.2.1 DéTECTION avec la base A43 . . . . .	148
7.3.2.2 DéTECTION avec la base AVG . . . . .	150
7.4 Conclusion . . . . .	153
<b>8 Conclusion et perspectives</b>	<b>157</b>
<b>Bibliographie</b>	<b>173</b>
<b>III Annexes</b>	<b>175</b>
<b>A Influence du bruit gaussien sur les mesures d'angle</b>	<b>177</b>
<b>B Procédure d'extraction</b>	<b>181</b>
B.1 Notations et Hypothèses . . . . .	181
B.1.1 Hypothèses . . . . .	181
B.1.2 Extracteurs . . . . .	182

B.2 Calcul de $\Delta_+(\mathbf{m}, t)$ . . . . .	182
<b>C Expérience de pré-détection</b>	<b>185</b>
C.1 Pré-détection avec des mesures de couleur . . . . .	185
C.1.1 Mesure $(r, g, L), \gamma = 2$ . . . . .	185
C.1.2 Mesure $(r, g), \gamma = 2$ . . . . .	189
C.1.3 Fusion d'information et effet de $\gamma$ . . . . .	192
C.2 Pré-détection avec des mesures de forme . . . . .	196
<b>D Rotation dans le plan image</b>	<b>201</b>
D.1 Valeurs propres et vecteurs propres . . . . .	201
D.2 Distribution des projections des modèles dans l'espace propre . . . . .	202
<b>E Preuve de la convexité de <math>J_{HU}(\mathbf{c}, \sigma_\rho)</math></b>	<b>205</b>
<b>F Liens entre la théorie semi-quadratique et l'approche de Huber</b>	<b>207</b>
F.1 Cas des images en niveaux de gris . . . . .	208
F.1.1 Montrons que $\rho(t) \leq q^\omega(t) \leq q^r(t)$ . . . . .	208
F.1.1.1 Montrons que $\rho(t) \leq q^\omega(t)$ . . . . .	208
F.1.1.2 Montrons que $q^\omega(t) \leq q^r(t)$ . . . . .	209
F.1.2 Remarques . . . . .	209
F.2 Extension aux images multibandes . . . . .	211
F.2.1 Montrons que $\rho(\ w\ ) \leq q^\omega(w)$ . . . . .	211
F.2.2 Montrons que $q^\omega(w) \leq q^r(w)$ . . . . .	213
F.2.3 Conclusion . . . . .	213
<b>G Reconnaissance : expérience avec la base COIL</b>	<b>215</b>
<b>H Reconnaissance : expérience avec la base A43</b>	<b>217</b>
H.1 Images d'apprentissage . . . . .	217
H.2 Test de reconnaissance sur des images présentant des occultations . . . . .	218
<b>I Courbes COR</b>	<b>221</b>
<b>J Détection : expérience sur des images en niveaux de gris</b>	<b>225</b>
<b>K Détection : expérience sur des images en couleur</b>	<b>227</b>
K.1 Test avec la base A43 . . . . .	227
K.2 Test avec la base AVG . . . . .	228

*TABLE DES MATIÈRES*

11

<b>L Publications de l'auteur</b>	<b>231</b>
<b>Résumé-Abstract</b>	<b>233</b>



# Liste des tableaux

4.1	<i>Effet de l'extracteur <math>\Delta_+</math> pour une sensibilité <math>\gamma = 2</math> dans le cas des hypothèses <math>H_4</math> et <math>H'_4</math> avec <math>H_0</math>, <math>H_1</math>, <math>H_2</math> et <math>H_3</math> vraies.</i>	65
4.2	<i>Résultats de pré-détection sur une séquence de 5km.</i>	74
5.1	<i>Variance de l'erreur de reconstruction <math>\mathbf{w}^{0,\mathbf{r}}</math> pour les différents apprentissages, pour un espace propre <math>F</math> de dimension <math>J = 5</math>.</i>	90
6.1	<i>Notations des estimateurs suivant les hypothèses a priori sur les distributions de probabilité de <math>\mathbf{c}</math> et <math>\mathbf{w}</math>.</i>	106
6.2	<i>Expressions des fonctions robustes <math>\rho</math> utilisées.</i>	118
6.3	<i>Paramètres d'échelle des différentes fonctions <math>\rho</math> utilisées.</i>	121
6.4	<i>Part d'inertie expliquée pour la base d'apprentissage COIL, suivant les deux méthodes d'apprentissage.</i>	123
6.5	<i>Résultats de reconnaissance sur les images d'apprentissage.</i>	124
6.6	<i>Résultats de reconnaissance sur des images présentant des fonds variés.</i>	125
6.7	<i>Résultats de reconnaissance sur des occurrences partiellement occultées.</i>	126
6.8	<i>Résultats de reconnaissance sur l'ensemble des 432 images testées.</i>	127
6.9	<i>Comparaison des paramètres <math>\hat{\mathbf{c}}</math> estimés conjointement avec le paramètre d'échelle et avec ceux obtenus avec un paramètre d'échelle fixé a priori.</i>	128
6.10	<i>Résultats de reconnaissance sur des images présentant des fonds texturés.</i>	130
6.11	<i>Résultats de reconnaissance sur les images fortement dégradés de l'annexe H avec un espace propre <math>F</math> de dimension <math>J = 60</math> et <math>J = 100</math>.</i>	131
7.1	<i>Détecteurs classiques et robustes.</i>	142
F.1	<i>Expressions des fonctions robustes <math>\rho</math> utilisées, et des fonctions des variables auxiliaires <math>\Psi</math> et <math>\xi</math>.</i>	211
G.1	<i>Reconnaissance : Observations partiellement occultées.</i>	215

H.1	<i>Panneaux triangulaires de la base d'apprentissage A43.</i>	217
H.2	<i>Observations.</i>	218
H.3	<i>Modèles reconnus par le système de reconnaissance robuste, sur un espace propre <math>F</math> de dimension <math>J = 60</math>, pour les observations du tableau H.2.</i>	219
I.1	<i>Calcul des <math>N_o</math> points <math>(x(s), y(s))</math> de la courbe COR.</i>	222
J.1	<i>Images du test de détection avec la base d'apprentissage COIL.</i>	226
K.1	<i>Détection : Images de test avec la base d'apprentissage A43.</i>	228
K.2	<i>Détection : Images de test avec la base d'apprentissage AVG.</i>	229

# Table des figures

1.1	<i>Exemple d'information à extraire d'une image de scène routière.</i>	22
2.1	<i>Attributs visuels d'un objet.</i>	26
2.2	<i>Détection des éléments d'une classe d'objets.</i>	37
3.1	<i>Mesures de forme l'image <math>L(x, y)</math> : la norme, l'angle du gradient et l'alignement.</i>	53
3.2	<i>Partition en classes de l'espace des mesures : exemple en dimension <math>d = 2</math>.</i>	54
3.3	<i>Fonctions d'interpolation utilisées dans le calcul des histogrammes classiques et interpolés (<math>d = 1</math>).</i>	55
3.4	<i>Densité de probabilité gaussienne théorique.</i>	56
3.5	<i>Différence entre l'approximation de p.d.f. par histogrammes classique et interpolé.</i>	56
3.6	<i>Influence d'un bruit gaussien sur les histogrammes : distance <math>\mathcal{L}_1(\mathbf{h}, \mathbf{h}_b)</math> en fonction de l'écart type du bruit <math>\sigma</math>.</i>	57
3.7	<i>Influence d'un gain sur les histogrammes : distance <math>\mathcal{L}_1(\mathbf{h}, \mathbf{h}_g)</math> en fonction d'un gain <math>g</math>.</i>	58
4.1	<i>Images successives d'une séquence de scènes routières.</i>	61
4.2	<i>Trois images synthétiques successives.</i>	66
4.3	<i>Cartes de pré-détection calculées avec les mesures de couleur (<math>\gamma = 2</math>).</i>	66
4.4	<i>Cartes de pré-détection calculées avec les mesures de couleur (<math>\gamma = 4</math>).</i>	66
4.5	<i>Cartes de pré-détection calculées avec les mesures de forme (<math>\gamma = 2</math>).</i>	67
4.6	<i>Résultats de pré-détection couleur sur l'image <math>I_{75}</math> pour différentes valeurs de <math>\gamma</math>.</i>	67
4.7	<i>Résultat de rétroposition avec des histogrammes classiques calculés sur des mesures couleurs (<math>r, g, L</math>) (<math>\gamma = 2</math>).</i>	68
4.8	<i>Résultat de rétroposition avec des histogrammes interpolés calculés sur des mesures couleurs (<math>r, g, L</math>) (<math>\gamma = 2</math>).</i>	69
4.9	<i>Exemple de résultats de pré-détection couleur (<math>\gamma = 2</math>) lorsque l'hypothèse <math>H_3</math> n'est pas strictement vérifiée.</i>	69

4.10 Résultat de rétroprojection avec des histogrammes classiques calculés sur des mesures de forme $(\theta, \alpha, N)$ .	70
4.11 Résultat de rétroprojection avec des histogrammes interpolés calculés sur des mesures de forme $(\theta, \alpha, N)$ .	70
4.12 Norme du gradient, $N$ , de l'image $I_{75}$ .	71
4.13 Fusion d'information entre la rétroprojection en avant et en arrière pour l'image $I_{75}$ .	71
4.14 Fusion des résultats de rétroprojection avec les mesures de couleur et de forme pour l'image $I_{75}$ .	72
4.15 Régions à mesures de forme verticales.	73
4.16 Pré-détection par mesures de forme verticales : l'ombre portée sur le délinéateur atténue ses contours et empêche sa détection.	73
4.17 Pré-détection par mesures de forme verticales : le délinéateur est détecté.	73
4.18 Suivi : rétroprojection de la distribution $\mathcal{P}(\mathbf{m}, 75 \mathcal{B})$ dans les images précédentes.	75
5.1 Image multi-bande $\mathbf{x}^k$ .	81
5.2 Exemple d'images d'apprentissage de $\mathcal{B}$ (issues de la base COIL [106]).	82
5.3 Image des pondérations $\mathbf{b}^0$ appliquée à l'image d'apprentissage $\mathbf{x}^0$ .	84
5.4 Estimées des moyennes $\hat{\mu}^c$ et $\hat{\mu}^n$ , de la base $\mathcal{B}$ , et image centrée $\tilde{\mathbf{x}}^0$ .	85
5.5 Spectres de $(\hat{\mu}^c, \hat{\Sigma}^c)$ et $(\hat{\mu}^n, \hat{\Sigma}^c)$ .	86
5.6 Spectre de $(\hat{\mu}^n, \hat{\Sigma}^n)$ .	87
5.7 Part d'inertie expliquée $Q_J$ en fonction du nombre $J$ de vecteurs propres retenus, pour les apprentissages $(\hat{\mu}^c, \hat{\Sigma}^c)$ et $(\hat{\mu}^n, \hat{\Sigma}^c)$ .	88
5.8 Cinq premiers vecteurs propres issus des estimateurs $(\hat{\mu}^c, \hat{\Sigma}^c)$ .	88
5.9 Reconstruction de $\tilde{\mathbf{x}}^0$ sur les cinq premiers vecteurs propres issus de $(\hat{\mu}^c, \hat{\Sigma}^c)$ .	89
5.10 Cinq premiers vecteurs propres issus des estimateurs $(\hat{\mu}^n, \hat{\Sigma}^c)$ .	89
5.11 Reconstruction de $\tilde{\mathbf{x}}^0$ sur les cinq premiers vecteurs propres issus de $(\hat{\mu}^n, \hat{\Sigma}^c)$ .	89
5.12 Cinq premiers vecteurs propres issus des estimateurs $(\hat{\mu}^n, \hat{\Sigma}^n)$ .	90
5.13 Reconstruction de $\tilde{\mathbf{x}}^0$ sur les cinq premiers vecteurs propres issus de $(\hat{\mu}^n, \hat{\Sigma}^n)$ .	90
5.14 Représentation dans un espace de dimension $NP = 3$ : $\mathbf{c}_{NP} = \{c_1, c_2, c_3\}$ , $\mathbf{c} = \{c_1, c_2\}$ et l'espace propre $F$ est défini par les vecteurs $(\mathbf{u}_1, \mathbf{u}_2)$ .	91
5.15 Répartition des coordonnées $(c_1, c_2, c_3)$ des images de la base COIL.	95
5.16 Quelques images d'apprentissage de la base A1 avec indication de leur angle de rotation	96
5.17 Couples de vecteurs propres associés à une valeur propre double	96
5.18 Distribution des images d'apprentissage de la base A1 dans $F$	96
5.19 Quelques images d'apprentissage de la base A43	97
5.20 Valeurs propres de la base A43.	97

5.21 Vecteurs propres de la base A43, suivant les deux apprentissages $(\hat{\mu}^c, \hat{\Sigma}^c)$ et $(\hat{\mu}^n, \hat{\Sigma}^c)$ . . . . .	98
5.22 Distribution des coordonnées $c$ des images d'apprentissage de A43 (apprentissage $\hat{\mu}^n, \hat{\Sigma}^c$ ). . . . .	98
 6.1 Reconnaissance par modèle d'apparence globale.	
. . . . .	101
6.2 Polynômes $q^r$ et $q^\omega$ approchant, en un point $t_0$ , la fonction $\rho_{GM}$ . . . . .	110
6.3 Courbes de convergence des algorithmes ARTUR et LEGEND. . . . .	113
6.4 Fonctions quadratique $Q$ et robustes $\rho : HS$ «surfaces minimales», $HL$ (d'Hebert et Leahy) et $GM$ (de Geman et Mc Clure). . . . .	118
6.5 Approximation de la distribution $\mathcal{P}(\epsilon inliers, \mathcal{B})$ par une gaussienne (cas des images en couleur). . . . .	122
6.6 Estimation en continuation avec le paramètre d'échelle fixé a priori grâce aux images l'apprentissage. . . . .	122
6.7 Estimation en continuation avec le paramètre d'échelle estimé conjointement aux paramètres $c$ avec la fonction convexe $HS$ . . . . .	123
6.8 Exemple des variables auxiliaires utilisées dans les algorithmes de minimisation : $J = 10$ , apprentissage $(\hat{\mu}^c, \hat{\Sigma}^c)$ . . . . .	125
6.9 Exemple de résultats de reconnaissance : l'observation $y$ est identifiée par l'image d'apprentissage $x$ , suivant les deux méthodes d'apprentissage avec les estimateurs robustes ( $J = 10$ ). . . . .	127
6.10 Reconnaissance robuste sur une image d'apprentissage à paramètre d'échelle fixé $\sigma_{max}$ . . . . .	129
6.11 Quelques images testées. . . . .	129
6.12 Quelques observations présentant des occurrences fortement dégradées de nos objets d'intérêt et les modèles reconnus par l'estimateur $\hat{c}_{RMV}$ avec l'apprentissage $(\hat{\mu}^n, \hat{\Sigma}^c)$ . . . . .	130
6.13 Masques des pondérations $b^A$ estimés pour l'observation $y^{29}$ ( $J = 60$ ). . . . .	131
6.14 Importance de la couleur pour la reconnaissance. . . . .	132
 7.1 Détection. . . . .	135
7.2 Exemple de cartes de détection DFFS, RDFFS et $\ b^A\ $ obtenues sur les images $I_{10}$ et $I_{15}$ pour un espace propre $F$ de dimension $J = 10$ . . . . .	144
7.3 Courbes COR calculées avec les détecteurs classiques pour un espace propre $F$ de dimension $J = 10$ . . . . .	145
7.4 Influence de l'apprentissage : courbes COR obtenues avec les détecteurs robustes sur un espace propre $F$ de dimension $J = 10$ pour les apprentissages $(\hat{\mu}^c, \hat{\Sigma}^c)$ et $(\hat{\mu}^n, \hat{\Sigma}^c)$ . . . . .	146
7.5 Influence du choix de la fonction $\rho$ : courbes COR calculées avec les fonctions $\rho_{HS}$ , $\rho_{HL}$ et $\rho_{GM}$ pour un espace propre $F$ de dimension $J = 10$ . . . . .	147

7.6	<i>Influence de la dimension de l'espace propre <math>F</math> : courbes COR obtenues avec les détecteurs robustes.</i>	147
7.7	<i>Influence d'un bruit gaussien : courbes COR obtenues par les détecteurs robustes pour un espace propre <math>F</math> de dimension <math>J = 10</math>.</i>	148
7.8	<i>Courbes COR obtenues pour les détecteurs robustes RDFFS et <math>\ \mathbf{b}^A\ </math>, sur un espace propre <math>F</math> de dimension <math>J = 60</math>, avec l'apprentissage <math>(\hat{\boldsymbol{\mu}}^n, \hat{\Sigma}^c)</math>.</i>	149
7.9	<i>Carte de détection de l'image <math>I_2</math> obtenue avec le détecteur robuste <math>\ \mathbf{b}^A\ </math>.</i>	149
7.10	<i>Base AVG : image moyenne des panneaux blancs apprise suivant ses rotations dans le plan de l'image (<math>\theta</math> angle de rotation).</i>	150
7.11	<i>Distribution des images d'apprentissage (en vert) et des panneaux blancs (en rouge) dans les plans propres.</i>	151
7.12	<i>Cartes de ressemblance obtenues sur les images <math>I_2</math> et <math>I_{17}</math>.</i>	155
7.13	<i>Courbes COR obtenues pour les différents détecteurs classique et robustes sur un espace de dimension 20.</i>	156
7.14	<i>Courbes COR obtenues avec <math>\mathcal{P}(\hat{\mathbf{c}}_{RMV}   \mathcal{B})</math> et <math>\mathcal{P}_\gamma(\mathbf{y}, \hat{\mathbf{c}}_{RMV}   \mathcal{B})</math> sur un espace propre <math>F</math> de dimension <math>J = 20</math>.</i>	156
F.1	<i>Fonctions <math>b^A(w)</math>, <math>b^L(w)</math>, <math>\Psi(b^A)</math> et <math>\xi(b^L)</math> de la fonction de Geman et Mc Clure (cas <math>P = 1</math>).</i>	212
I.1	<i>Analyse d'une carte de ressemblance de l'image <math>I_k</math> à un seuil <math>s</math> donné.</i>	222
I.2	<i>Tracé de la courbe COR.</i>	223

# Notations

- $\det(\mathbf{U})$  : déterminant de la matrice  $\mathbf{U}$ ,
- $\mathbf{u}^T$  : transposée du vecteur  $\mathbf{u}$ ,  $\mathbf{U}^T$  transposée de la matrice  $\mathbf{U}$ .
- $\mathbf{u}^T \mathbf{v} = <\mathbf{u} | \mathbf{v}>$  : produit scalaire des vecteurs  $\mathbf{u}$  et  $\mathbf{v}$ ,
- $\|\mathbf{u}\| = \sqrt{\mathbf{u}^T \mathbf{u}}$  : norme euclidienne du vecteur  $\mathbf{u}$ ,
- $|u|$  : valeur absolue du scalaire  $u$ ,
- Le **rappor signal à bruit** est défini comme le rapport des variances entre l'image de référence  $I$  et l'image bruitée  $I + b$  :  
$$SNR = 10 \log_{10} \left( \frac{\sigma_I^2}{\sigma_b^2} \right)$$
où  $\sigma_I^2$  est la variance de l'image de référence  $I$ , et  $\sigma_b^2$  est la variance du bruit.
- $E[u]$  : espérance de  $u$ .
- $\Re(u)$  : partie réelle de  $u$ ,  $\Im(u)$  partie imaginaire de  $u$ .



# Chapitre 1

## Introduction

### 1.1 Contexte de l'étude

Les bases de données d'images de la route et de son environnement sont de plus en plus utilisées pour l'exploitation et les études de sécurité routière. Les images numériques acquises par des véhicules en circulation sur le réseau routier constituent une masse importante d'informations à traiter. Afin de faciliter l'extraction de l'information pertinente, cette thèse a pour objectif de proposer des méthodes d'analyse semi-automatique d'images de scènes routières.

L'information recherchée concerne plus particulièrement le positionnement des objets fixes présents sur l'accotement qui représentent un danger potentiel pour un véhicule quittant la chaussée. Les objets d'intérêt sont en priorité les objets d'orientation verticale possédant une partie *fût* tels que les arbres, les poteaux, les piles de pont, les panneaux et les délinéateurs, et les objets à *fût* horizontaux, tels que les glissières. Ces objets, manufacturés et naturels, doivent être détectés, puis reconnus. La stratégie générale de recherche consiste à effectuer une première détection rapide des objets d'intérêt, pouvant inclure des fausses alarmes mais si possible sans oublier. Ces premiers résultats sont affinés ultérieurement par d'autres techniques, éventuellement sous le contrôle d'un opérateur humain.

Des séquences d'images couleur sont acquises par une caméra montée sur une voiture en circulation sur le réseau routier. L'information de distance parcourue est enregistrée avec les images, ce qui permet de limiter la quantité d'information à traiter en ne considérant qu'une image tous les  $n$  mètres. Les scènes à analyser sont des scènes dynamiques extérieures prises de jour sous un éclairage naturel. La nature des scènes routières peut varier d'un environnement urbain ou rural, provenant de routes départementales ou nationales. La figure 1.1 présente un exemple d'image de scène routière à analyser ainsi que quelques objets d'intérêt à localiser.

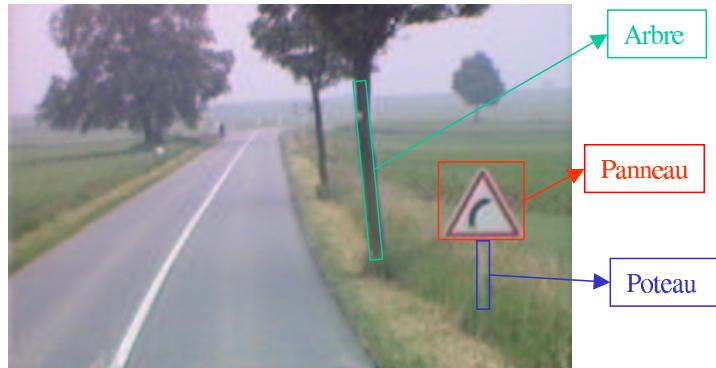


FIG. 1.1: Exemple d'information à extraire d'une image de scène routière.

## 1.2 Détection et reconnaissance d'objets

La détection et la reconnaissance d'objets sont des tâches essentielles dans plusieurs applications telles que l'indexation par le contenu ou la vidéo-surveillance. Les méthodes classiques en vision par ordinateur reposent sur la comparaison de caractéristiques visuelles d'une observation avec des références préalablement apprises par le système. Prendre en compte toutes les variations d'apparence d'un objet dans une image est un défi important. Celles-ci dépendent essentiellement de trois facteurs : la nature de l'objet, la scène et le système d'acquisition.

- ✓ Nos objets d'intérêt peuvent être classés en deux catégories : les éléments naturels comme les arbres, et les objets manufacturés tels que la signalisation verticale ou les poteaux. La première classe a, par essence, une variabilité infinie. La seconde semble plus simple à modéliser : les panneaux répondent à des normes précises concernant la forme et la couleur. Cependant, ceux-ci peuvent présenter des variabilités liées à l'usure des matériaux : peintures effacées, présence de rouille ou autres dégradations.
- ✓ En plus des variations de l'objet, les conditions de prises de vue ne sont pas contrôlées : l'éclairage incident sur la scène dépend des conditions (météorologiques et horaire) d'enregistrement de la séquence. Les changements de position du soleil et de son intensité impliquent des changements d'apparence de l'objet dans les images.
- ✓ Suivant la position du système d'acquisition dans la scène, l'objet apparaît selon différents points de vue, et à différentes échelles. Par ailleurs, il peut être partiellement caché ou dans l'ombre d'un autre objet. En plus des effets d'occultations et d'ombres portées, les reflets et les contre-jours liés à la position de la caméra par rapport au soleil, introduisent aussi des variabilités dans l'apparence d'un objet. Enfin, notre système d'acquisition n'est pas calibré photométriquement, et l'on ne dispose pas de référence de couleur fiable.

## 1.3 Sommaire des chapitres de la thèse

Compte tenu de la variété des objets que l'on cherche à détecter et à reconnaître, on souhaite mettre en œuvre des méthodes génériques, applicables à plusieurs classes d'objets. L'indexation d'images est un thème de recherche qui intègre les techniques d'analyse d'images permettant l'accès rapide à des images ou des objets particuliers, dans les bases d'images et de vidéos, généralistes ou spécialisées. Les systèmes d'indexation fonctionnent en deux étapes. La première, réalisée hors ligne, crée un index contenant un ensemble de signatures descriptives du contenu de la base. La seconde étape, exécutée en ligne, gère la réponse à une requête de l'utilisateur. Notre étude, qui concerne la recherche d'objets particuliers dans de grandes bases d'images de scènes routières, obéit à une problématique similaire à celle de l'indexation. Aussi, le chapitre 2 propose un état de l'art sur les caractéristiques visuelles utilisées en indexation de bases de données d'images et de vidéos, ainsi que sur les méthodes de comparaison de signatures utilisées pour détecter et reconnaître les objets. On s'intéresse particulièrement à la détection et la reconnaissance de panneaux. Les systèmes de transports intelligents proposent des méthodes spécifiques à ces objets, et sont également passés en revue dans l'état de l'art. On s'intéresse particulièrement aux signatures locales et globales de l'apparence utilisées avec des méthodes statistiques.

**Partie 1.** Cette partie présente un système de pré-détection non-supervisée d'objets dans les séquences d'images avec caméra en mouvement. Les techniques classiques de détection et reconnaissance reposent sur la comparaison entre les caractéristiques de l'observation et celles de référence préalablement apprises. Les conditions de prise de vue étant non contrôlées, l'apprentissage de mesures de référence est difficile. Aussi, pour l'éviter, nous proposons une méthode de détection non supervisée d'événements, inspirée des méthodes de détection de changement de plans dans les vidéos. Cette méthode met à profit le mouvement apparent des objets entre images successives. Les événements considérés sont l'apparition, la disparition et le changement d'apparence d'objets dans une séquence d'images acquises par une caméra en mouvement. Ce type de comportement est caractéristique de la plupart de nos objets d'intérêt. La technique proposée est basée sur l'analyse des changements statistiques temporels. Elle réalise une comparaison des distributions statistiques des mesures locales, approchées par des histogrammes multidimensionnels, entre deux images successives dans la séquence. Les statistiques propres aux objets apparaissants, disparaissants ou changeant d'apparence sont extraites sous certaines hypothèses. Leur localisation dans les images est alors réalisée par des techniques de rétroprojection.

Le chapitre 3 précise les caractéristiques locales utilisées, ainsi que la façon de calculer leurs histogrammes. Le chapitre 4 décrit la méthode de détection d'objets non supervisée que nous proposons.

Cette première partie utilise des signatures locales de petite dimension. Leur manipulation pour la détection et la reconnaissance d'objets est rapide et naturellement robuste aux occultations. Cependant, leur utilisation suppose l'existence d'une information locale suffisamment forte pour permettre la discrimination entre les objets recherchés et leur contexte.

**Partie 2.** La seconde partie de ce mémoire propose des méthodes de reconnaissance et de détection robustes basées sur une représentation globale de l'apparence des objets. La signature utilisée est l'image d'une occurrence d'un objet. Elle est de grande dimension. Les variabilités de l'apparence d'une classe d'objets sont apprises en collectant un ensemble de ces caractéristiques. Afin de réduire la dimension de ces signatures, on utilise une technique classique de réduction de données : l'Analyse en Composantes Principales (ACP). Cette méthode nécessite l'estimation des moments d'ordre 1 (la moyenne) et 2 (la matrice d'autocovariance) sur l'ensemble des signatures. Au chapitre 5, on rappelle les estimateurs habituellement utilisés pour l'ACP, puis nous proposons un nouvel estimateur de la moyenne qui permet d'obtenir une représentation plus efficace.

Les techniques de détection et de reconnaissance sur les espaces propres nécessitent, dans une première étape, l'estimation des composantes réduites de l'observation. Chaque image d'apprentissage est alors représentée par une plus petite signature, à une erreur de reconstruction près. Ces coordonnées réduites sont liées à l'image d'apprentissage par une relation linéaire qui correspond à une projection sur l'espace propre. L'estimation classique, basée sur une hypothèse de bruit gaussien, n'est pas robuste à la présence de bruit structuré provoqué, par exemple, par des occultations ou par la présence de fond texturé. Ainsi, les techniques classiques de reconnaissance et de détection requièrent la segmentation préalable de l'objet et ne tolèrent pas les perturbations induites, par exemple, par les occultations, ce qui limite leur utilisation.

Afin de pallier ces limitations, nous proposons d'utiliser une hypothèse robuste de distribution du bruit, permettant une meilleure estimation (*M-estimation*) du représentant de l'observation dans l'espace propre. Cette hypothèse est mise à profit pour proposer de nouvelles méthodes de reconnaissance robustes au chapitre 6, ainsi que de nouvelles méthodes robustes de détection au chapitre 7. L'ensemble de ces techniques est étendu aux images multibandes, et est appliqué à des images en niveaux de gris et en couleur.

Les méthodes proposées dans cette thèse reposent sur une formulation probabiliste. Ce cadre rigoureux permet d'intégrer les incertitudes liées, par exemple, au système d'acquisition ou aux problèmes de données incomplètes (cas des occultations). La décision est directement liée à la probabilité des résultats.

**Conclusion et perspectives.** Nous concluons au chapitre 8 en rappelant l'ensemble des méthodes proposées dans cette thèse : la méthode de pré-détection rapide par mesures locales et les méthodes de détection et de reconnaissance robustes par mesures globales de l'apparence. Quelques perspectives sont également proposées.

# **Chapitre 2**

## **Etat de l'art**

L'objectif de cette thèse est de définir des méthodes permettant de détecter et de reconnaître des objets dans de grandes bases d'images séquentielles de scènes routières. Cet état de l'art recense les principales méthodes utilisées dans des contextes similaires, notamment en indexation par le contenu de bases d'images et de vidéos, et dans les systèmes de transports intelligents.

Lors d'une première phase de pré-traitement, les systèmes d'indexation par le contenu extraient des images un ensemble de signatures pertinentes. La recherche d'un objet particulier dans une base d'images est réalisée par comparaison entre sa signature et celles des objets de la base. Cela suppose qu'un objet ou une classe d'objets puisse être représenté efficacement par un ensemble de caractéristiques. Les signatures sont extraites suivant les propriétés visuelles propres à une image comme sa couleur ou sa texture, et sont présentées à la section 2.1.

Les signatures sont ensuite utilisées pour détecter et reconnaître des objets. Ceci est réalisé par calcul de similarité entre signatures ou par des mesures statistiques de vraisemblance. Ces deux approches sont présentées dans la section 2.2. De façon plus spécifique, les méthodes de détection et de reconnaissance de panneaux routiers, présentes sur les systèmes de transport intelligents (ITS), sont passées en revue dans la dernière section 2.3. Ces systèmes ont la contrainte majeure de devoir fonctionner en temps réel, contrairement aux applications d'indexation de bases de données où la description par le contenu des images de la base est réalisée hors-ligne, et où seule la réponse à l'utilisateur interrogant la base de données doit être rapide.

### **2.1 Caractéristiques images**

Les caractéristiques images sont représentatives des différentes propriétés visuelles d'un objet dans une image ou une vidéo : couleur, forme, texture, apparence, mouvement, ainsi que des contraintes spatiales qui peuvent lier l'objet à son environnement, ou définir des interactions entre les régions le composant [9] (cf. figure 2.1).

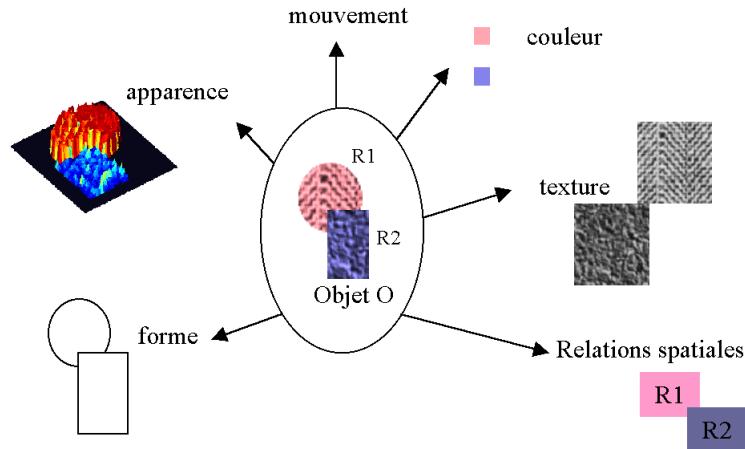


FIG. 2.1: Attributs visuels d'un objet.

Les signatures peuvent se distinguer par leurs propriétés, notamment :

- l'invariance : les signatures restent identiques en présence de certaines transformations de l'image (changement d'éclairage, effets de rotation, ...),
- la robustesse : les signatures varient peu sous l'effet du bruit dû au système d'acquisition, d'occultations, ...
- la dimension : plus les caractéristiques utilisées sont de petites dimensions, moins le temps de calcul nécessaire aux comparaisons est long. De plus, une dimension réduite des signatures permet de limiter la place mémoire nécessaire pour l'indexation.

### 2.1.1 La couleur

L'information couleur d'un pixel est définie par ses valeurs dans les trois canaux couleurs. Plusieurs espaces couleurs permettent de représenter ce triplet. On peut distinguer trois catégories [9] :

- les modèles colorimétriques qui lient la couleur mesurée à l'aide d'un colorimètre à la mesure de réflectance,
- les modèles «physiologiques» qui copient les trois types de cônes de la rétine du système visuel humain (RGB [52, 64, 84], XYZ).
- les modèles «psychologiques» qui considèrent la perception humaine de la couleur. Les espaces HSV [142, 148], OPP [149] et La\*b\* [127] appartiennent à cette catégorie.

L'espace couleur peut être choisi pour sa capacité à reproduire la perception humaine, nécessaire dans les systèmes d'indexation pour correctement répondre à la requête de l'utilisateur. De plus certaines composantes couleurs, notamment les chrominances, présentent l'avantage d'être invariantes aux changements d'éclairage ou à l'orientation de la surface de l'objet coloré. La distribution statistique de la

couleur dans une image, approchée par des histogrammes ou des mélanges de gaussiennes [8], peut être utilisée comme signature. Celle-ci peut être calculée sur chacun des canaux couleurs [148] (distributions mono-dimensionnelles), sur les axes chromatiques (distribution bi-dimensionnelle) ou suivant les trois axes couleurs [149, 84].

## 2.1.2 La forme

### 2.1.2.1 Représentation *surface*

**Attributs géométriques.** Une région peut être caractérisée par des attributs géométriques tels que sa surface, son périmètre, son rectangle englobant, sa compacité définie comme le rapport entre son périmètre au carré et sa surface,... Ainsi, dans le système d'indexation VisualSeek, Smith utilise comme information de forme la position du centre de masse de la région, la surface, la largeur et la hauteur du plus petit rectangle englobant la région [142].

**Moments réguliers.** Soit la fonction  $f(x, y)$  qui caractérise la surface de l'objet :  $f(x, y) = 1$  lorsque  $(x, y)$  est à l'intérieur de l'objet, et  $f(x, y) = 0$  sinon. Le moment régulier d'ordre  $pq$  est défini par :

$$M_{p,q} = \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy$$

Les moments réguliers de la fonction  $f(x, y)$  correspondent à sa projection sur la base des polynômes non orthogonaux  $p_{pq}(x, y) = x^p y^q$ . Le moment  $M_{0,0}$  correspond à la surface de l'objet et,  $M_{1,0}$  et  $M_{0,1}$  sont les coordonnées du centre de gravité.

**Moments de Zernike.** Les moments de Zernike utilisent une base orthogonale de polynômes complexes [71]. Leur amplitude est invariante à la rotation. L'invariance par rapport à la translation est obtenue en prenant le centre de gravité comme origine des coordonnées. Le premier moment régulier est normalisé pour obtenir l'invariance par rapport à l'échelle.

### 2.1.2.2 Représentation *contour*

Un contour se définit comme une courbe fermée et paramétrée  $\Gamma(u) = (x(u), y(u))$ . On peut également le représenter par la fonction de courbure, la distance au centre de masse ou la fonction des coordonnées complexes des contours [67, 86].

**Descripteurs de Fourier.** Les descripteurs de Fourier représentent la forme de l'objet dans le domaine fréquentiel. Les descripteurs de basses fréquences décrivent la forme générale de l'objet tandis que ceux

de fréquences élevées contiennent l'information sur de plus petits détails. Un sous-ensemble des descripteurs de Fourier est souvent suffisant pour discriminer plusieurs types de formes. Les descripteurs de Fourier peuvent être rendus invariants à la translation, la rotation et aux effets d'échelle [67, 86, 2].

**Modèle Autorégressif.** Un modèle linéaire Autorégressif exprime la valeur d'une fonction comme la combinaison linéaire d'un certain nombre de ses valeurs précédentes. Les coefficients AR de cette combinaison linéaire sont estimés et utilisés comme signatures [67, 138].

**Représentation multi-échelle.** C'est une représentation des points d'inflexion d'une courbe plane, lissée et fermée, proposée par Mokhtarian et al.[101]. La courbe est lissée par une gaussienne d'écart type  $\sigma$  et est notée  $\Gamma(u, \sigma)$ . Lorsque  $\sigma$  augmente,  $\Gamma(u, \sigma)$  devient de plus en plus convexe.  $K(u, \sigma)$  désigne la courbure de  $\Gamma_\sigma$ . Les maxima en  $\sigma$  de la courbe définie par l'équation  $K(u, \sigma) = 0$ , sont utilisés comme caractéristiques. Les contours des objets sont normalisés pour que cette représentation soit invariante aux effets d'échelle [101].

**Contours déformables.** Cette méthode, proposée par Sclaroff et Pentland [135, 136], est utilisée par le système de recherche d'images Photobook [112]. Pour comparer les formes de deux objets, il faut décrire les déformations qui les relient. Chaque objet est décrit par un modèle physique de type «éléments finis», où chaque point du contour est caractérisé par ses relations avec les autres points de contour via une équation d'équilibre dynamique. L'énergie de déformation nécessaire à l'objet  $A$  pour correspondre à l'objet  $B$  est calculée et utilisée comme mesure de similarité [112].

**Histogramme de forme.** La forme d'un objet peut être représentée par la distribution des angles de direction des points de contour. Cette distribution est calculée par un histogramme [64]. L'invariance par rapport à l'échelle est obtenue par normalisation de l'histogramme. L'utilisation de la direction des points de contour est invariante aux translations. Pour limiter la sensibilité aux petites rotations dans le plan de l'image, l'histogramme de forme est lissé.

### 2.1.3 Apparence

L'apparence se définit comme la forme de la surface d'intensité de l'image. On peut distinguer les représentations globales de l'apparence qui décrivent l'objet dans son ensemble, des représentations locales qui n'en décrivent qu'une partie ou région. La surface d'intensité d'une image modélisée par les valeurs des pixels représente une grande quantité d'information dans le cas d'une représentation globale. Pour la réduire, deux méthodes ont été proposées : les surfaces déformables et les décompositions suivant des modèles linéaires ou non linéaires. Comme pour les contours déformables, les surfaces déformables utilisent un modèle physique d'interconnectivité entre les noeuds  $(x_i, y_i, I_i)$  d'images en niveaux de

gris  $(x, y) \rightarrow I(x, y)$  dans l'espace 3-dimensionnel  $XYI$  [105]. La décomposition suivant une base modale permet une réduction d'information avec une interprétation purement mécanique. On s'intéresse ici particulièrement à la représentation de l'apparence globale par modèles linéaires et non-linéaire qui permet de lier la surface d'intensité à une caractéristique de plus faible dimension. Ceci est présenté à la section 2.1.3.1. La section 2.1.3.2 présente les mesures locales couramment utilisées en indexation d'images.

### 2.1.3.1 Représentation globale de l'apparence

La caractéristique globale de l'apparence peut se représenter par un vecteur  $\mathbf{x}$  contenant dans l'ordre lexicographique les valeurs des  $N$  pixels de l'image de l'objet considéré. Ce vecteur est de très grande dimension. On se propose de réduire la dimension de  $\mathbf{x}$  en l'exprimant en fonction d'une caractéristique  $\mathbf{c}$  de dimension inférieure telle que :

$$\begin{cases} \mathbf{x} = g(\mathbf{c}) + \mathbf{w}^r \\ \mathbf{c} = f(\mathbf{x}) \end{cases}$$

où  $f$  et  $g$  peuvent être des fonctions linéaires ou non, et  $\mathbf{w}^r$  est l'erreur de reconstruction.  $g$  décrit une surface paramétrée de  $\mathbb{R}^N$  (ou courbe si  $\mathbf{c}$  est de dimension 1).  $f$  est son projecteur associé [87, 69, 22]. Les paramètres  $\mathbf{c}$  sont classiquement estimés en minimisant la distance euclidienne entre  $\mathbf{x}$  et  $g(\mathbf{c})$  :

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c}} \{ \|\mathbf{x} - g(\mathbf{c})\| \}$$

La première difficulté consiste à trouver les relations  $f$  et  $g$  qui permettent une bonne représentation d'un ensemble d'images d'apprentissage  $\mathcal{B} = \{\mathbf{x}^1, \dots, \mathbf{x}^K\}$ , conduisant à une erreur de reconstruction  $\|\mathbf{w}^r\|$  minimale. Plusieurs méthodes de réduction de données ont été proposées utilisant des relations linéaires ou non entre  $\mathbf{x}$  et  $\mathbf{c}$ .

**Modèles linéaires.** On présente trois représentations linéaires : l'analyse en composantes principales, l'analyse discriminante et l'analyse en composantes indépendantes.

**Analyse en Composantes Principales.** L'Analyse en Composantes Principales (ACP) est une méthode classique de réduction de données. Elle consiste à représenter un ensemble  $\mathcal{B} = \{\mathbf{x}^1, \dots, \mathbf{x}^K\}$  d'images d'apprentissage par un sous-espace centré sur leur image moyenne  $\boldsymbol{\mu}$ . Les fonctions de base sont obtenues en calculant les vecteurs et valeurs propres de la matrice de covariance définie par :

$$\Sigma = \frac{1}{K} \sum_{k=1}^K \tilde{\mathbf{x}}^k \tilde{\mathbf{x}}^{k T}$$

où les images d'apprentissage sont centrées ( $\tilde{\mathbf{x}} = \mathbf{x} - \boldsymbol{\mu}$ ). Les vecteurs propres de  $\Sigma$  sont classés par ordre décroissant de leurs valeurs propres. Ils regroupent alors de manière décroissante les caractéristiques générales des images d'apprentissage. Ainsi les premiers vecteurs propres résument les éléments

communs aux modèles et les derniers représentent les particularités et les détails. On peut réaliser une compression d'information avec un minimum d'erreur en choisissant de représenter les modèles sur une base réduite constituée par les vecteurs propres associées aux plus grandes valeurs propres. Si l'on note  $\mathbf{U}$  la matrice des  $J$  vecteurs propres principaux, pour toute image  $\mathbf{x}$  de  $\mathcal{B}$ , on a la relation :

$$\tilde{\mathbf{x}} = \mathbf{U}\mathbf{c} + \mathbf{w}^r$$

où  $\mathbf{w}^r$  est l'erreur de reconstruction orthogonale à l'espace propre  $F$  défini par les vecteurs principaux  $\{\mathbf{u}_1, \dots, \mathbf{u}_J\}$ . Les coordonnées réduites  $\mathbf{c}$  minimisent  $\|\tilde{\mathbf{x}} - \mathbf{U}\mathbf{c}\|$  et se calculent simplement par une projection orthogonale de  $\tilde{\mathbf{x}}$  sur  $F$  :

$$\hat{\mathbf{c}} = \mathbf{U}^T \tilde{\mathbf{x}}$$

Une autre méthode pour calculer les vecteurs propres est de formuler l'ACP comme une estimation par les moindres carrés [158, 74] :

$$\hat{\mathbf{U}} = \arg \min_{\mathbf{U}} \left\{ \mathcal{J} = \sum_{k=1}^K \|\tilde{\mathbf{x}}^k - \mathbf{U}\mathbf{U}^T \tilde{\mathbf{x}}^k\|^2 \right\} \quad (2.1)$$

Cette estimation est peu robuste à la présence d'erreurs grossières dues, par exemple, à la présence d'images d'apprentissage qui ne concernent pas la classe d'objets que l'on souhaite modéliser, ou des images d'apprentissage fortement dégradées. De la Torre et Black ont récemment proposé de calculer les vecteurs propres en utilisant un estimateur robuste (M-estimateur) [74] :

$$\hat{\mathbf{U}} = \arg \min_{\mathbf{U}} \left\{ \mathcal{J} = \sum_{k=1}^K \rho \left( \tilde{\mathbf{x}}^k - \mathbf{U}\mathbf{U}^T \tilde{\mathbf{x}}^k \right) \right\}$$

Cette estimation calcule des pondérations pour chaque pixel de chaque image d'apprentissage : une pondération proche de 0 (respectivement proche de 1), indique que le pixel est considéré comme une donnée erronée (resp. donnée fiable).

Moghaddam et Pentland caractérisent l'ensemble des images de  $\mathcal{B}$  par leur densité de probabilité [97]. En la supposant gaussienne, l'ACP permet de calculer les moments d'ordre 1 (la moyenne  $\boldsymbol{\mu}$ ) et 2 (la matrice d'autocovariance  $\Sigma$ ), qui permet d'écrire la vraisemblance de  $\mathbf{x}$  :

$$\mathcal{P}(\mathbf{x}|\mathcal{B}) = \frac{\exp \left[ -\frac{1}{2} \tilde{\mathbf{x}}^T \Sigma^{-1} \tilde{\mathbf{x}} \right]}{(2\pi)^{\frac{NP}{2}} |\det(\Sigma)|^{\frac{1}{2}}} \quad (2.2)$$

Cette vraisemblance peut être approchée afin de limiter le temps de calcul, pour se résumer au produit de deux probabilités gaussiennes, l'une relative à la distribution dans l'espace propre  $F$  et l'autre relative à la distribution dans son espace orthogonal  $F^\perp$  [97].

**Analyse Discriminante.** Les données d'apprentissage sont partitionnées *a priori* en  $C$  classes, notée  $\{C_l\}_{l \in \{1, \dots, C\}}$ , ce qui nécessite une connaissance additionnelle sur les données d'apprentissage.

L'analyse discriminante calcule la matrice de vecteurs de base  $\hat{V}$  qui maximise le rapport entre la covariance inter-classe,  $\Sigma_{inter}$ , et la covariance intra-classes  $\Sigma_{intra}$  :

$$\hat{V} = \arg \max_V \frac{V^T \Sigma_{inter} V}{V^T \Sigma_{intra} V} \quad \left| \begin{array}{l} \Sigma_{inter} = \sum_{l=1}^C N_l (\boldsymbol{\mu}_l - \boldsymbol{\mu})(\boldsymbol{\mu}_l - \boldsymbol{\mu})^T \\ \Sigma_{intra} = \sum_{l=1}^C \sum_{\mathbf{x} \in C_l} (\mathbf{x} - \boldsymbol{\mu}_l)(\mathbf{x} - \boldsymbol{\mu}_l)^T \end{array} \right.$$

où  $\boldsymbol{\mu}$  est l'image moyenne de toutes les images d'apprentissage,  $\boldsymbol{\mu}_l$  est l'image moyenne de la classe  $C_l$  et  $N_l$  le nombre d'images dans la classe  $C_l$ . Les vecteurs  $\mathbf{c}$  et  $\mathbf{x}$  sont liés par la relation linéaire [6] :  $\mathbf{c} = \hat{V}^T \tilde{\mathbf{x}}$ .

**Analyse en Composantes Indépendantes.** L'Analyse en Composantes Indépendantes permet de calculer des coordonnées  $\mathbf{c}$  telles que  $\{c_1, \dots, c_J\}$  soient les plus indépendantes possibles [95]. Tandis que l'ACP s'appuie sur les moments d'ordre 1 et 2, l'ACI considère des moments d'ordres supérieurs pour calculer des axes de l'espace propre sur lesquels les coordonnées des images d'apprentissage sont indépendantes.  $\mathbf{x}$  et  $\mathbf{c}$  sont liés par la relation :

$$\tilde{\mathbf{x}} = A\mathbf{c} + \mathbf{w}^r$$

associée à la projection  $\mathbf{c} = A^{-1} \tilde{\mathbf{x}}$ . Cette relation est linéaire mais, au contraire de la matrice  $U$  obtenue par l'ACP, la matrice  $A$  n'est pas orthogonale [95].

**Modèles non linéaires.** Les représentations linéaires ne sont pas toujours suffisantes pour représenter un ensemble d'échantillons. On rappelle succinctement quelques méthodes de représentations non-linéaires.

**Approximation par des splines.** Murase et Nayar utilisent une fonction spline  $g$  pour approcher la répartition des projections des images d'apprentissage dans l'espace propre  $F$  modélisé par ACP [103]. Les échantillons d'apprentissage sont des images d'un objet variant suivant deux paramètres : l'angle de prise de vue et des conditions d'illumination  $(g(c_1, c_2))$ . L'expression du projecteur  $f$  associé à  $g$  n'est pas explicite. Après projection orthogonale sur  $F$  de l'observation  $\mathbf{x}$ , l'estimation de  $(c_1, c_2)$  est réalisée en recherchant le point de la surface  $g(c_1, c_2)$  le plus proche.

**Courbes et surfaces principales.** Cette approche permet d'estimer itérativement une courbe ou une surface qui résume la répartition des images d'apprentissage. Chaque point de cette surface correspond à la moyenne des images de  $\mathcal{B}$  qui s'y projettent. On peut montrer que les seules lignes droites qui satisfont cette condition sont les composantes principales calculées par ACP. Les courbes principales sont donc une généralisation de l'ACP [22, 134]. L'apprentissage de surfaces principales  $g$  dépendant d'un paramètre  $c$  de dimension supérieure à 2, peut être réalisé par réseaux de neurones [22, 134], notamment à fonctions de transfert sigmoïdales [95].

**Kernel ACP.** Au contraire des représentations évoquées, cette méthode consiste à extraire des composantes principales d'un espace de dimension plus grande que l'espace des caractéristiques  $\mathbf{x}$  [134]. Les images d'apprentissage de  $\mathbb{R}^N$  sont transformées en  $\mathbf{x}'$  de  $\mathbb{R}^M$  ( $M > N$ ) par une application non linéaire  $\Phi$  telle que  $\mathbf{x}' = \Phi(\mathbf{x})$ . Les données  $\mathbf{x}'$  sont centrées ( $\sum_{k=1}^K \Phi(\mathbf{x}^k) = 0$ ) et la matrice d'autocovariance des données  $\mathbf{x}'$  est calculée par :

$$\Sigma' = \frac{1}{K} \sum_{k=1}^K \Phi(\mathbf{x}^k) \Phi(\mathbf{x}^k)^T$$

La matrice de vecteurs propres  $\mathbf{U}'$  de  $\Sigma'$  est calculée. La projection non linéaire d'une image  $\mathbf{x}$  s'écrit [134] :

$$\mathbf{c} = \mathbf{U}'^T \mathbf{x}' = \mathbf{U}'^T \Phi(\mathbf{x})$$

La transformation non linéaire préalable  $\Phi$ , permet éventuellement de rendre représentable de façon linéaire des données organisées initialement selon des variétés non linéaires.

**Estimation robuste de  $\hat{\mathbf{c}}$ .** Une fois les relations  $g$  et  $f$  calculées sur un ensemble d'échantillons d'apprentissage, estimer les coordonnées  $\mathbf{c}$  sur une nouvelle observation  $\mathbf{y}$  est une tâche difficile si du bruit structuré  $\mathbf{w}^o$  (non blanc, non gaussien), provoqué par exemple par des occultations, la perturbe :

$$\mathbf{y} = \mathbf{x} + \mathbf{w}^o = g(c) + \mathbf{w}$$

Dans ce cas, l'erreur  $\mathbf{w}$  englobe l'erreur due à la troncature  $\mathbf{w}^r$  et le bruit d'observation  $\mathbf{w}^o$ . Les paramètres  $\mathbf{c}$  ne peuvent plus être calculés de façon fiable par le projecteur  $f$ . La présence de données aberrantes dans l'observation implique l'utilisation de techniques robustes d'estimation. Cela équivaut à définir un nouveau projecteur robuste.

**M-estimateurs.** Les M-estimateurs permettent de réduire l'influence des données aberrantes de l'observation dans la régression. Les paramètres  $\mathbf{c}$  sont calculés en minimisant :

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c}} \left\{ \sum_{n=1}^N \rho(y_n - g(\mathbf{c})_n) \right\}$$

A partir d'une représentation  $g$  linéaire modélisée par ACP,  $g(\mathbf{c}) = \mathbf{U}^T \tilde{\mathbf{x}}$ , Black utilise cet estimateur pour estimer les coordonnées  $\mathbf{c}$  sur l'espace propre et reconnaître des objets occultés [15]. La minimisation est réalisée par un algorithme itératif des moindres carrés pondérés.

**Test d'hypothèses.** Leonardis et Bischof proposent d'estimer les paramètres  $\mathbf{c}$  sur les vecteurs propres en utilisant un sous-ensemble des données observées  $\mathbf{y} = \{y_1, \dots, y_N\}$  [10, 80]. Plusieurs estimations sont réalisées sur différents sous-ensembles, générant plusieurs hypothèses de reconnaissance. La reconnaissance est réalisée en sélectionnant la meilleure hypothèse [80].

### 2.1.3.2 Représentation locale de l'apparence

A l'exception des méthodes robustes, plus coûteuses en temps de calcul, la représentation globale des objets ne permet pas de gérer les effets d'occultations partielles. Celles-ci sont moins gênantes dans le cas de la représentation locale. L'objet étant représenté par un ensemble de caractéristiques propres à chacune de ses régions, il suffit qu'une seule d'entre elles ne soit pas occultée, pour permettre sa reconnaissance.

**ACP locale.** Pour être plus robuste aux occultations, plusieurs auteurs réalisent un apprentissage par ACP sur des sous-régions des objets d'intérêt [38, 108, 165]. Lorsque l'ensemble d'apprentissage est très vaste, les vecteurs propres convergent naturellement vers une base de dérivées de gaussienne qui est la meilleure représentation locale de la surface image [124].

**Dérivées de Gaussiennes.** Cette méthode proposée par Manmatha et Ravela [90] et Schmid [132], caractérise localement la surface d'intensité de l'image par son développement de Taylor pour une image en niveaux de gris. On peut définir un *jet local* (ou signature locale) en chaque point d'une image  $I$ . Des signatures invariantes aux transformations telles que la rotation sont proposés [132]. A l'ordre 2 et à une échelle donnée, l'ensemble des invariants à ces transformations sont :

$$\begin{aligned} d_0 &= I && \text{intensité} \\ d_1 &= I_x^2 + I_y^2 && \text{amplitude du gradient} \\ d_2 &= I_{xx} + I_{yy} && \text{laplacien} \\ d_3 &= I_{xx}I_xI_x + 2I_{xy}I_xI_y + I_{yy}I_yI_x \\ d_4 &= I_{xx}^2 + 2I_{xy}^2 + I_{yy}^2 \end{aligned}$$

Manmatha et Ravela calculent en chaque point de l'image le vecteur  $\Delta_\sigma = \langle d_1, \dots, d_4 \rangle_\sigma$  à trois échelles différentes. La caractéristique  $d_0$  n'est pas utilisée pour que le système soit invariant aux effets d'illumination. Le vecteur invariant multi-échelle s'écrit  $D = \langle \Delta_{\sigma_1}, \Delta_{\sigma_2}, \Delta_{\sigma_3} \rangle$ . Un objet peut donc être caractérisé par plusieurs vecteurs  $D_j$ ,  $j$  indiquant la position sur l'objet.

Schiele propose d'utiliser comme signature, la distribution d'un ensemble de mesures locales approchée par des histogrammes multidimensionnels [131].

### 2.1.4 La texture

Quelques modèles de représentation de l'information de texture sont présentés comme les méthodes spatio-fréquentielles utilisant les filtres de Gabor, le modèle auto-régressif, la décomposition de Wold qui est liée aux trois caractéristiques psychovisuelles de la texture : la répétitivité, la directivité et la granularité.

### 2.1.4.1 Approches par filtrage spatio-fréquentiel

Manjunath et Ma utilisent les filtres de Gabor pour l'analyse des textures dans le système de recherche d'images *Netra* [89, 86, 85]. Les fonctions de Gabor sont des exponentielles complexes modulées par des gaussiennes. Elles forment une base complète mais non-orthogonale, et toute image  $I(x, y)$  peut être décomposée dans cette base. Les caractéristiques utilisées pour l'indexation par Manjunath et Ma sont la moyenne et la variance, de la convolution entre l'image  $I$  et les fonctions de base de Gabor.

Dans le système de recherche d'images VisualSEEK, John Smith décompose le plan de Fourier grâce à une pyramide de filtres QMF ou filtres miroirs à quadrature [142]. Le vecteur de mesures de texture est composé de l'énergie locale de réponse de l'image pour chacun des filtres fréquentiels. Une région texturée peut être caractérisée par un histogramme de mesures locales de texture.

### 2.1.4.2 Modèles autorégressifs

Dans le modèle autorégressif simultané (SAR), la valeur d'intensité d'un pixel est estimée comme une combinaison linéaire des valeurs de ses voisins, plus un terme de bruit. Les paramètres du modèle autorégressif sont utilisés comme signature. Le modèle autorégressif multirésolution (MRSAR) calcule ces paramètres pour plusieurs résolutions de la texture analysée en utilisant une pyramide gaussienne [91]. Les caractéristiques utilisées dans l'indexation sont les coordonnées du modèle MRSAR [86]. Le modèle STAR pour *Spatio-Temporal AutoRegressive* proposé par Szummer et Picard [151], permet d'étendre le modèle précédent pour l'analyse des textures «mouvantes» dans les séquences vidéo : la valeur d'intensité d'un pixel est estimée comme une combinaison linéaire des valeurs de ses pixels spatialement et temporellement voisins.

### 2.1.4.3 Décomposition de Wold

La théorie de Wold permet de décomposer un signal aléatoire  $y$  en trois signaux orthogonaux entre eux, et ceci de manière unique :  $y = w + p + e$ .  $w$  est un champ purement aléatoire,  $e$  est la composante évanescante, et  $p$  est déterministe [83]. Ces trois signaux, ou *composantes de Wold*, se rapportent aux caractères aléatoire, directionnel et périodique du signal.  $p$  est la composante harmonique relative à la périodicité,  $e$  est la composante évanescante relative à la directivité, et  $w$  est la composante aléatoire relative à la granularité. Le signal harmonique est caractérisé par les fréquences et les amplitudes des pics harmoniques détectés dans le domaines de Fourier. Le signal évanescant est extrait par une transformée de Hough de l'amplitude de la transformée de Fourier de l'image. Un histogramme de la pente des droites est calculé. La variance de cet histogramme et la variance de l'énergie de Fourier le long des droites correspondant aux pics de l'histogramme sont utilisées comme caractéristiques discriminantes du signal évanescant [83]. Le signal aléatoire est modélisé par un modèle MRSAR. En pratique, le caractère évanescant des textures étant faible, la composante évanescante est également caractérisée par un modèle

MRSAR avec la composante aléatoire [83].

### 2.1.5 Le mouvement

L'information de mouvement est souvent utilisée pour décrire les documents vidéos [112, 48, 65, 39]. Les travaux de recherche ont pour objectif de caractériser les mouvements de caméra et des différents objets composants la scène. Les travaux récents se sont intéressés à la caractérisation de mouvement dans des cadres génériques peu contraints (textures temporelles, notions d'activité de mouvement) [45].

Dans le système de recherche d'images PHOTobook, Pentland et al. utilisent le mouvement pour caractériser les objets [112]. L'estimation du champ de vitesse ou flot optique est réalisée sous certaines hypothèses restrictives comme notamment la constance des niveaux de gris entre deux images successives et est généralement coûteuse en temps de calcul.

Pour reconnaître le type de mouvement de la caméra, Joly utilise une technique dite "X-ray" [65]. Celle-ci consiste à étudier la variation temporelle des images projetées sur les deux axes  $x$  (image "X-ray") et  $y$  (image "Y-ray"). La projection X-ray correspond alors à la moyenne de l'intensité des pixels suivant chaque colonne de l'image (respectivement suivant chaque ligne pour la représentation Y-ray). On extrait alors les contours des images X-ray et Y-ray calculés sur une durée de 1 seconde. Lorsqu'on connaît les angles de ces contours, les divers types de mouvements de la caméra peuvent être identifiés. Cette méthode est robuste à la présence de petits objets mobiles.

Certains travaux proposent d'exploiter l'information directement codée en format MPEG. Manjunath et al. analysent le mouvement grâce à un histogramme bi-dimensionnel calculé pour les vecteurs de mouvement  $(v_x, v_y)$  directement extraits des macro-blocs MPEG [39]. L'analyse de cet histogramme de mouvement permet de séparer le mouvement de la caméra du mouvement d'un objet de la scène.

### 2.1.6 Description de vidéo

Les méthodes de partitionnement temporel de vidéo sont basées sur la détection des transitions entre deux plans [65]. Une méthode de détection des transitions consiste à mesurer les variations statistiques entre les images. Ceci est réalisé par calcul d'une mesure de similarité entre deux histogrammes couleurs calculés sur deux images successives de la séquence. Lorsque cette mesure dépasse un certain seuil [17], la présence d'un changement de plan est détectée.

Zabih et al. proposent de détecter les transitions en calculant une distance entre deux images successives de contours dans la séquence [163]. Un contour qui apparaît dans l'image à l'instant  $t$  loin des contours présents à l'instant  $t - 1$ , est qualifié de *contour entrant*. Inversement, les contours à l'instant  $t - 1$  loin des contours de l'image  $t$ , sont considérés comme des *contours sortants*. Le nombre des pixels de contours sortants ou entrants est utilisé pour détecter une transition.

### 2.1.7 Topologie spatiale

Das et al. indexent les régions d'images qui ne se superposent pas, par des histogrammes de mesures locales [35]. Chaque région est caractérisée par les pics prédominants de son histogramme. Les images de la base dont les régions présentent des caractéristiques identiques à celles de l'image recherchée, sont retenues. Puis la cohérence spatiale des régions dans l'image recherchée et celles de la base est analysée par l'utilisation d'un *graphe des couleurs voisines* (CNG : Color Neighborhood Graph). Ce graphe vérifie la cohérence spatiale des régions, notamment si elles sont voisines ou non dans les images.

### 2.1.8 Conclusion

Cette section a présenté les principales caractéristiques utilisées pour décrire les bases d'images et de vidéo. On peut discerner deux types de signatures : la mesure locale qui ne décrit qu'une partie d'un objet, et la mesure globale qui caractérise l'objet dans son ensemble. On remarque que l'apparence locale caractérisée par son développement de Taylor, correspond au degré 0 à l'information de couleur. L'information relative à un ensemble de mesures (locales ou globales) peut aussi être résumée par leur distribution de probabilité. Deux méthodes de calcul de fonction de densité de probabilité ont été utilisées :

- une méthode non paramétrique : l'histogramme largement utilisé sur les mesures locales de forme, vitesse, apparence, couleur, texture, ...,
- une méthode paramétrique avec une hypothèse *a priori* de distribution gaussienne ou multi-gaussienne, appliquée aux caractéristiques globales de l'apparence [97].

L'histogramme calculé à partir des mesures locales d'un objet peut être considéré comme une mesure globale de l'objet. Cette méthode de calcul des densités de probabilité est étudiée plus en détail au chapitre 3. Le modèle gaussien de Moghaddam et Pentland évoqué à la section 2.1.3.1 est une caractéristique globale à l'ensemble d'une base d'images d'apprentissage.

Différents types d'information peuvent être utilisées conjointement. Par exemple, Schneiderman propose d'utiliser une signature composée d'une mesure d'apparence locale et de sa position (topologie spatiale), et de calculer leur distribution avec des histogrammes multidimensionnels [133]. De même, Schmid caractérise ses objets d'intérêt par un ensemble de mesures locales d'apparence et les angles qui les lient entre elles [132].

Différents critères peuvent motiver le choix des caractéristiques images utilisées, par exemple le type d'objets d'intérêt, le contexte dans lequel ils apparaissent, l'application demandée, la taille de la signature (pour son stockage en mémoire), la complexité de son calcul ... Les mesures locales de couleur, de mouvement et de texture [112, 48, 84] peuvent être utilisées pour segmenter puis indexer les régions homogènes des images ou des vidéos. Le paragraphe suivant aborde l'utilisation des signatures pour des tâches de détection et de reconnaissance.

## 2.2 Application à la détection et la reconnaissance

La section précédente a présenté différentes signatures utilisées en indexation de base de données d'images. Cette section présente les méthodes utilisées pour comparer deux signatures. Comme illustré à la figure 2.2, la détection d'un objet ou d'une classe d'objets est réalisée en évaluant la ressemblance entre la caractéristique extraite de la fenêtre d'observation et celles des objets recherchés. Cette ressemblance est calculée pour chaque position  $(i, j)$  de la fenêtre d'observation et permet d'obtenir une carte de détection. Les positions présentant la meilleure ressemblance indiquent la présence des objets recherchés (pixels sombres sur la carte de détection).

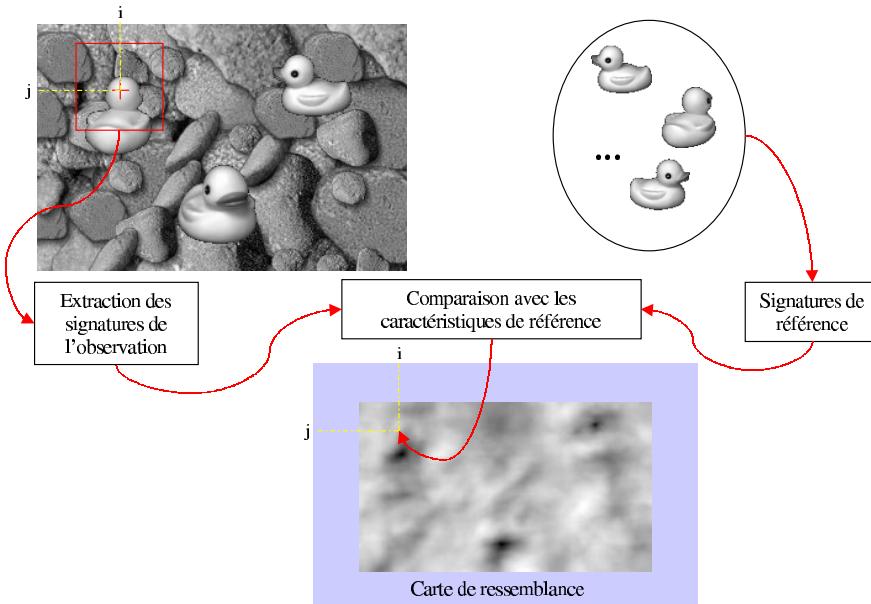


FIG. 2.2:Détection des éléments d'une classe d'objets.

Afin de limiter le temps de calcul nécessaire à cette recherche systématique, Schmid et al. proposent pour les représentations locales de l'apparence, de sélectionner préalablement les positions dans l'observation avec un détecteur de points d'intérêt (détecteur de Harris). Les mesures ne sont calculées et comparées aux références, seulement sur les positions sélectionnées [98]. On distingue deux méthodes de calcul de ressemblance entre signatures : les similarités présentées dans la section 2.2.1 et les mesures de vraisemblances dans la section 2.2.2. Ces dernières se rapportent directement à des mesures de probabilités. Si la ressemblance est forte entre l'observation et la classe d'objets de référence, une étape supplémentaire de reconnaissance est nécessaire pour identifier l'objet de la classe auquel l'observation correspond.

### 2.2.1 Calcul de similarité

On considère deux signatures  $\mathbf{h}$  et  $\mathbf{k}$  de dimension  $N$  dont on cherche à évaluer la ressemblance.

**Mesures de distance.** Les mesures de distance sont couramment utilisées pour comparer des signatures notamment, dans le cadre de l'appariement de mesures locales [38, 132, 99], et dans le cadre d'une représentation d'une classe d'objets par espace propre. Elles peuvent être définies par :

$$\mathcal{L}_r(\mathbf{h}, \mathbf{k}) = \left( \sum_{j=1}^N |h_j - k_j|^r \right)^{1/r}$$

On peut également considérer la distance euclidienne pondérée [54] :

$$\mathcal{L}_2^S(\mathbf{h}, \mathbf{k}) = \sqrt{(\mathbf{h} - \mathbf{k})^T S (\mathbf{h} - \mathbf{k})}$$

où  $S$  est une matrice symétrique dont les valeurs  $s_{ij}$  permettent de prendre en compte les corrélations entre les composantes  $h_i$  et  $h_j$  des signatures [132]. Celles-ci peuvent être apprises [132], ou fixées par l'utilisateur suivant l'importance qu'il accorde à chacune des mesures [142].

**Distance à l'espace propre (Distance from Feature Space).** Cette méthode de détection, basée sur une représentation globale de l'apparence par ACP (cf. section 2.1.3.1), a été proposée pour évaluer la ressemblance entre une observation  $\mathbf{y}$  (extraite dans la fenêtre d'observation) et l'ensemble  $\mathcal{B}$  d'images d'apprentissage [97]. La DFFS consiste à calculer la distance euclidienne entre l'observation et l'espace propre :

$$DFFS(\mathbf{y}) = \|\tilde{\mathbf{y}} - \mathbf{U}\mathbf{U}^T\tilde{\mathbf{y}}\|$$

La reconnaissance d'une observation est réalisée en comparant, par distance euclidienne, les coordonnées réduites de l'observation  $\mathbf{c} = \mathbf{U}^T\tilde{\mathbf{y}}$ , et celles qui indexent chaque image d'apprentissage.

**Comparaison d'histogrammes et rétroprojection d'histogrammes locaux.** Plusieurs mesures de similarité entre histogrammes ont été proposées dans la littérature [127], telles que l'intersection d'histogrammes par Swain [149], la mesure du  $\chi^2$  [131, 4] et les mesures de distances [45]. Sur la région définie par la fenêtre d'observation, un histogramme local est calculé et sa similarité avec l'histogramme de l'objet recherché est mesurée. Cette opération répétée pour toutes les positions  $(i, j)$  de la fenêtre permet d'obtenir une carte de similarité. Cette méthode de détection est appelée la rétroprojection par histogrammes locaux (local histogramming backprojection [142]).

### 2.2.2 Méthodes statistiques

**Rétroprojection de vraisemblance.** Dans cette section, la ressemblance d'une observation est évaluée par sa vraisemblance par rapport à un modèle statistique de référence. Par exemple, calculer un histogramme sur un ensemble de mesures locales  $\mathbf{m}$  prélevées sur un objet  $\mathcal{O}$  revient à modéliser la distribution de la probabilité de  $\mathbf{m}$  :  $\mathcal{P}(\mathbf{m}|\mathcal{O})$ . Les positions  $(i, j)$  dont les mesures,  $\mathbf{m}_{i,j}$ , présentent les plus fortes vraisemblances,  $\mathcal{P}(\mathbf{m}_{i,j}|\mathcal{O})$ , permettent de localiser l'objet  $\mathcal{O}$ .

De même, dans le cadre de la représentation globale de l'apparence, Moghaddam et Pentland ont proposé un modèle probabiliste pour la distribution des images d'apprentissage de la base  $\mathcal{B}$  dans l'espace des images (cf. équation (2.2)). Cette méthode statistique a permis d'obtenir de meilleurs résultats que la DFFS pour localiser des visages [97].

**Rétroprojection bayésienne.** Au lieu de rétroprojeter la vraisemblance *a priori*  $\mathcal{P}(\mathbf{m}|\mathcal{O})$  on peut rétroprojeter la vraisemblance *a posteriori* définie grâce à la formule de Bayes par :

$$\mathcal{P}(\mathcal{O}|\mathbf{m}) = \frac{\mathcal{P}(\mathbf{m}|\mathcal{O}) \mathcal{P}(\mathcal{O})}{\mathcal{P}(\mathbf{m})}$$

L'histogramme des mesures locales  $\mathbf{m}$  de l'image observée permet d'approximer  $\mathcal{P}(\mathbf{m})$ , et  $\mathcal{P}(\mathcal{O})$  est une constante [142].

**Rétroprojection d'un rapport de vraisemblance.** Pour localiser l'objet recherché, Schneiderman rétroprojette le rapport de vraisemblance [133] :

$$\mathcal{R} = \frac{\mathcal{P}(\mathcal{O}|\mathbf{m})}{\mathcal{P}(\overline{\mathcal{O}}|\mathbf{m})}$$

En utilisant le théorème de Bayes,  $\mathcal{R}$  se réécrit :

$$\mathcal{R} = \frac{\mathcal{P}(\mathcal{O}) \mathcal{P}(\mathbf{m}|\mathcal{O})}{\mathcal{P}(\overline{\mathcal{O}}) \mathcal{P}(\mathbf{m}|\overline{\mathcal{O}})} \propto \frac{\mathcal{P}(\mathbf{m}|\mathcal{O})}{\mathcal{P}(\mathbf{m}|\overline{\mathcal{O}})}$$

Ce rapport nécessite l'apprentissage de la distribution de la classe *non-objets*  $\overline{\mathcal{O}}$ .

## 2.3 Systèmes de transports intelligents

Le domaine de recherche concernant la reconnaissance des panneaux routiers, consiste à la détection et la classification de panneaux dans des scènes routières filmées par un véhicule en circulation. C'est un problème qui paraît simple car les panneaux apparaissent dans des positions normalisées dans les scènes et, leur forme, leur couleur et leurs pictogrammes sont connus. Cependant, les conditions de prises de vue ne sont pas contrôlées : les vibrations du véhicule qui filme, les variations d'illumination ou les phénomènes d'occultations induits par d'autres véhicules, compliquent le problème.

Cette section présente différents travaux réalisés pour la détection et la reconnaissance de la signalisation routière. Cet état de l'art s'appuie sur ceux réalisés par Lalonde et Li en 1995 [76], Estable en

1996 [44] et Erckert en 1997 [43]. Plusieurs systèmes de reconnaissance des panneaux routiers ont été développés, principalement en Europe, mais aussi au Japon, au Canada et aux USA. La plupart d'entre eux utilisent les techniques du traitement de l'image.

### 2.3.1 Images monochromes

- L'étude menée par Peugeot s'intéresse à la détection des panneaux de couleur rouge (stop, interdiction, ...) en temps réel. Les images de la route sont acquises par une caméra munie d'un filtre passe-bande rouge et une segmentation couleur n'est alors pas nécessaire. Les images sont traitées par une détection de contour. La reconnaissance consiste en une extraction de caractéristiques et à leur classification par un système expert et un réseau de neurones. La détection et la reconnaissance sont réalisés en 0,7 s pour 94% de bonne reconnaissance [76].

- Plusieurs travaux ont été menés à l'université de Gênes sur la reconnaissance des panneaux de circulation dans des images en niveaux de gris. La détection repose sur une extraction de droites pour les panneaux triangulaires et d'anneaux pour les panneaux circulaires. La reconnaissance est réalisée par inter-corrélation avec les panneaux de la base de données. Le taux de détection pour les panneaux triangulaires atteint 92% avec 11 fausses alarmes sur 600 images, pour un temps de calcul de 6 secondes par image sur une station SPARC ELC. Pour les panneaux circulaires, le taux de bonne détection est de 93% pour un temps de calcul de 15s. La reconnaissance atteint 98% de bonne classification pour 0,5s de temps de calcul pour des bases de données contenant 60 panneaux circulaires et 47 triangulaires. Ce travail a été complété par une étude parallèle utilisant l'information couleur pour la détection. La détection est alors réalisée par une segmentation couleur suivie d'une analyse de formes [76].

### 2.3.2 Images en couleur

La plupart des systèmes de détection et reconnaissance exposés dans cette section fonctionne suivant deux étapes successives :

- segmentation des images par des mesures couleurs pour définir des zones d'intérêt,
- puis analyse de mesures de forme (sur les zones d'intérêt) pour reconnaître le panneau.

**Projet Daimler Benz.** L'ensemble des méthodes de détection et de reconnaissance de panneaux développées par Daimler-Benz durant le projet européen PROMETHEUS, a été utilisé par S. Estable lors de sa thèse pour la réalisation d'un prototype [44]. Deux types de segmentation ont été testés pour la détection des panneaux routiers : une segmentation de type contour et une segmentation de type région sur des critères de couleur. L'approche contour a été testée au début du projet avec un algorithme de segmentation hiérarchique de contour. Les résultats obtenus en laboratoire par cette méthode ont vite montré

ses limites. La primitive contour est en effet facilement perturbée par des recouvrements d'objets. Aussi, il n'a finalement été utilisé que deux segmentations de type région.

**Première méthode de segmentation couleur : Color Connected Component (CCC).** L'algorithme se déroule en deux phases. D'abord, l'image couleur originale est transformée en une image étiquetée. Un label est attribué à chaque pixel de l'image couleur de composante RGB. Ensuite, un algorithme de regroupement de pixels opère sur l'image « labellisée » pour en extraire la topologie de l'image. Les topologies qui correspondent à celles des panneaux routiers sont retenues.

Chaque pixel de l'image couleur d'origine est classifié selon les classes de couleur rouge, bleu, jaune, blanc, gris et noir. La classification est réalisée par un réseau neuronal entraîné à partir d'un ensemble d'images de référence. L'espace couleur normalisé ( $r,g,b$ ) sert de représentation des couleurs afin de rendre la classification robuste par rapport au changement d'éclairage.

$$l = \frac{R+G+B}{3} \quad r = \frac{R}{3l} \quad g = \frac{G}{3l}$$

Une table de correspondance (LUT) est remplie dans une première phase. Celle-ci contient pour chaque combinaison RGB le label correspondant. La classification des pixels d'une image se réduit donc à l'accès à cette table.

Cet algorithme regroupe les cellules de voisinage 4 en régions ayant le même label couleur. Il génère de manière séquentielle les frontières de chaque région. Simultanément, l'algorithme extrait pour chaque région la liste des régions voisines rencontrées en parcourant le bord extérieur, fournissant ainsi les relations topologiques complètes entre les régions de l'image. Les régions d'une taille inférieure à un seuil ne sont pas prises en considération. Pour la reconnaissance des panneaux routiers, des mesures grossières sont introduites comme le rapport de la surface entre la région et son rectangle circonscrit, le rapport hauteur/largeur du rectangle circonscrit, l'excentricité de la région ou la surface totale de la région. Ces mesures permettent d'éliminer les candidats dont les caractéristiques correspondent le moins à des panneaux. L'analyse de l'adjacence des régions n'est pas exploitée car le coût en calcul est important.

**Seconde méthode de segmentation couleur : Color Structure Code (CSC).** Cette méthode est proposée par l'université de Koblenz-Landau qui travaille en collaboration avec Daimler-Benz. Cette approche utilise le système de couleur HSV et s'appuie sur une maille hexagonale. Elle regroupe les avantages des traitements locaux simples et rapides, et ceux des méthodes de croissance de région. Au niveau le plus bas, une croissance de région locale est exécutée. Elle extrait des régions connectées de faible taille. Celles-ci sont ensuite traitées comme des pixels pour les niveaux supérieurs. Le CSC extrait les attributs de classe par l'analyse de la couleur et exploite les primitives contours qui sont définies comme la frontière des régions. Dans cette méthode, un candidat désigné par l'analyse couleur sera confirmé ou infirmé selon la nature de sa forme (la forme étant utilisée comme attribut de classe).

**Première méthode de reconnaissance de formes : reconnaissance structurelle de formes géométriques par combinaison de témoignage.** Une forme géométrique est décrite par des codes de Freeman à 4 ou 6 directions issus des deux segmentations couleurs précédentes. Cette méthode consiste à extraire la symétrie, la surface, la compacité, l'histogramme, le nombre de faces, les faces parallèles et colinéaires. La reconnaissance est alors réalisée à partir de ces caractéristiques.

**Seconde méthode de reconnaissance de formes : Reconnaissance de formes géométriques et de pictogrammes par classification statistique.** La reconnaissance des pictogrammes s'appuie sur une classification multiréférence sous la forme d'un arbre de décision où les feuilles sont les pictogrammes et les nœuds sont des classes de panneaux. Chaque classe de panneau correspond à une couleur et une forme particulière. La reconnaissance de formes géométriques et de pictogrammes doit être robuste par rapport aux variations d'éclairage, de taille et d'orientation. La taille et la couleur des candidats sont donc normalisées puis comparées à un ensemble de référence de formes géométriques et de pictogrammes sélectionnés au cours de la phase d'apprentissage. La première phase du processus de reconnaissance de forme consiste à déterminer la classe du panneau éventuellement présent à partir de l'examen de sa forme. La deuxième phase doit mener à la classification du pictogramme. La représentation normalisée est corrélée à un ensemble de masques se référant chacun à un pictogramme particulier.

**Méthodes utilisées.** La première méthode de segmentation couleur, plus rapide [44], et la seconde méthode d'analyse de forme ont été intégrées au prototype. Le système fonctionne en temps réel, pour la détection et la reconnaissance de certains panneaux autoroutiers. Le suivi des objets d'intérêt est aussi réalisé par un modèle estimation-prédiction de la fenêtre d'intérêt dans l'image suivante. Le taux de bonne détection obtenu est de 90% et le taux de reconnaissance dépasse 95%.

**Autre méthode de reconnaissance de forme de Koblenz Landau.** La détection et la reconnaissance s'effectuent en trois étapes constituant le TSC (Traffic Sign Classifier) [121]. La première réalise une segmentation couleur grâce à la CSC (Color Structure Code) [121, 125]. La seconde étape réalise une classification géométrique des couleurs des objets des différentes listes. On calcule pour chaque objet, sa probabilité d'avoir une des formes recherchées (triangulaire, circulaire, carrée). Si cette probabilité est inférieure à un certain seuil, la forme est rejetée. Cette probabilité est calculée à partir d'une approximation du contour par une enveloppe convexe de 24 points. La reconnaissance de la forme est invariante aux légères rotations. A ce niveau, il y a encore beaucoup de fausses alertes. Dans une troisième étape, un histogramme est calculé pour l'intérieur de l'objet analysé. Cet histogramme a des classes de couleurs spécifiques. Pour analyser les classes «noir» et «blanc», on évalue la luminosité par exemple en analysant les valeurs du rouge pour les panneaux d'interdiction.

**Japon.** Akatsuka et Imai ont pour but de détecter et de reconnaître les panneaux de vitesse de couleurs rouge et blanche avec des chiffres bleus en temps réel [76]. Des techniques classiques sont utilisées telles que le seuillage de l'espace des couleurs, un «template matching» pour l'analyse de formes, et une reconnaissance de chiffre par corrélation. La performance temps réel est permise grâce à l'utilisation de matériels spécifiques pour le seuillage et la reconnaissance de formes.

**USA.** Une étude a été menée sur la reconnaissance du panneau STOP dans le cadre du projet ADIS (Advanced Driver Information System) composante de IVHS (Intelligent Vehicle and Highway Systems). Ces panneaux sont détectés grâce à une segmentation couleur (groupement des pixels rouges) et par une analyse géométrique qui confirme ou rejette la possibilité de la présence d'un panneau [76]. L'espace de couleur utilisé est l'espace HSI.

Un autre projet permet de reconnaître différentes formes de panneaux de couleur rouge [66]. Le système proposé repose sur une segmentation couleur et une analyse géométrique. Les images RGB sont transformées en images  $L^*u^*v^*$ . Les composantes chromatiques  $u^*v^*$  permettent de s'affranchir des fortes variations d'illumination dans la scène. L'image obtenue est segmentée par l'algorithme des k-moyennes. L'analyse de forme est réalisée par la signature fournie par la transformée de Fourier de l'image contenant l'objet centré. Cette signature est invariante aux effets d'échelle et de rotations. La reconnaissance est réalisée par réseau de neurones.

**Canada.** Une étude est menée au Centre de recherche informatique de Montréal sur la détection de panneaux de circulation [75]. Une caméra CCD est montée à l'intérieur d'une voiture. Chaque image RGB est transformée en image HSI. Un label est attribué à chaque pixel de l'image ayant une couleur possible d'un panneau (rouge, bleu, blanc,...). Une méthode de rétroprojection d'histogramme calculé avec ces labels et non avec les valeurs des pixels, est utilisée pour la détection. Les régions d'intérêt présentant des distributions de couleurs similaires aux panneaux, peuvent ensuite être analysées suivant des critères de forme. Pour l'analyse de forme, plusieurs solutions sont proposées telles que le chaînage, suivi d'une interpolation de lignes et une transformée de Hough, ou la programmation dynamique, ou la distance de Hausdorff [77, 63]. La première approche est difficile au vu de la mauvaise qualité des contours obtenus. Les chaînes sont souvent cassées ou incomplètes, des contours surnuméraires sont détectés et ont une mauvaise influence sur la robustesse du système. La seconde approche est intéressante mais n'a pas été testée. Seule la dernière méthode est utilisée par l'équipe de CRIM.

Une étude menée à l'Université Concordia de Montréal propose de détecter et reconnaître les panneaux en quatre étapes [164] : la segmentation des images RGB par une classification des couleurs et définition des zones d'intérêt, une recherche des contours dans les zones d'intérêt, puis une analyse géométrique, et enfin une reconnaissance par des histogrammes. La segmentation des images suivant les couleurs caractéristiques des panneaux recherchés, est réalisée après l'apprentissage de leurs couleurs. Ainsi, les couleurs sont apprises à partir de quelques images de panneaux prises sous différentes intensi-

tés lumineuses. Cet apprentissage permet une quantification adaptée de l'espace RGB pour la segmentation. La reconnaissance est réalisée par histogramme dans la région intérieure du panneau (comptage du nombre de pixels de chaque couleur). Les panneaux symétriques, du type «tourner à droite» et «tourner à gauche», sont divisés en plusieurs régions pour lesquelles un histogramme est calculé afin de pouvoir les différencier.

### 2.3.3 Remarques

Les différents systèmes proposés dans cette section, utilisent unanimement l'information de couleur et de forme pour détecter et reconnaître les panneaux routiers. Ce type d'information est efficace pour caractériser ces objets et pour les discriminer de leurs environnements. La segmentation basée sur l'information couleur, est généralement la première étape d'analyse d'images qui permet de définir des zones d'intérêt. Elle nécessite la calibration des systèmes d'acquisition et les performances du système sont entièrement conditionnées par cette première étape.

L'utilisation des histogrammes de mesures de couleur pour la reconnaissance de pictogrammes, par l'université de Concordia à Montréal, rejoint des techniques déjà évoquées avec les méthodes d'indexation. Cependant, la robustesse des méthodes proposées, face à des occultations, aux ombres portées, à des changements de conditions d'illuminations, ne sont pas clairement évaluées. Ces systèmes ont pour objectif de fournir un système d'aide à la conduite. Aussi, ont-ils la contrainte majeure de devoir fonctionner en temps réel. Dans cette thèse, notre application n'a pas cette contrainte.

## 2.4 Conclusion

Cet état de l'art a présenté au premier paragraphe, plusieurs signatures utilisées pour caractériser les images et la vidéo dans les systèmes d'indexation. Les mesures locales (couleur, apparence,...), présentent l'avantage d'être robuste aux occultations ou aux perturbations locales (spécularités, ombres portées), d'être de faible dimension et généralement facilement extractible. Cependant, cette représentation locale ne conserve pas ou peu d'information topologique relative à la position des mesures les unes par rapport aux autres. Au contraire, la représentation globale de l'apparence permet de conserver à une erreur de reconstruction près toute l'information contenue dans les images des objets d'intérêt utilisées en apprentissage. La signature globale est de plus grande taille et est plus difficile à extraire lorsque l'on souhaite être insensible aux occultations partielles.

Le second paragraphe a présenté les méthodes de comparaison nécessaires aux applications de détection et de reconnaissance. La ressemblance entre signatures peut être évaluée par des mesures de similarité ou de vraisemblance.

Au dernier paragraphe, un intérêt particulier a été porté aux systèmes de détection et de reconnaissance de panneaux routiers disponibles sur les véhicules de transports intelligents. On remarque que les

informations de couleur et de forme sont unanimement utilisées. La segmentation couleur qui sert de première étape de sélection de zones d'intérêt (détection), conditionne fortement la réussite de ces systèmes. Cela implique qu'ils sont sensibles aux erreurs de classification.

**Choix des caractéristiques.** Les méthodes utilisées dans cette thèse reposent sur des modèles d'apparence. Deux représentations ont été utilisées : les mesures locales de l'apparence et les mesures globales.

La représentation d'un ensemble de mesures locales par leur histogramme, est simple à mettre en œuvre. La simplicité, la robustesse et la vitesse des méthodes de rétroprojection et de comparaisons d'histogrammes pour détecter et reconnaître les objets, représentent leurs atouts majeurs. On peut remarquer que la couleur correspond au descripteur d'ordre 0 de l'apparence locale, et l'information de contour est généralement extraite par segmentation des descripteurs d'ordre 1 ou 2 de l'apparence locale. On priviliege les mesures qui ne nécessitent pas une étape de segmentation dont les performances conditionneraient entièrement les résultats de notre application. Les mesures locales d'apparence utilisées dans cette étude correspondent à l'information de couleur et des mesures de forme calculées à partir des dérivées d'ordre 1 de l'image.

La signature d'apparence globale est un vecteur composé des valeurs des pixels de l'image de l'objet. Cette représentation globale contient plus d'information que la représentation par histogramme de mesures locales, car elle conserve l'information topologique. La sensibilité naturelle de cette représentation aux perturbations locales telles que les occultations partielles ou les effets spéculaires, peut être réduit par l'utilisation de méthodes robustes.

**Choix des méthodes.** On s'intéressera plus particulièrement aux méthodes statistiques qui formalisent de manière explicite l'*a priori* utilisé pour évaluer la ressemblance d'une signature.



**Première partie**

**Apparence locale**



## Introduction

Cette première partie s'intéresse à la représentation de l'apparence par des caractéristiques locales. Le caractère local implique que leur calcul est peu sensible aux perturbations (occultations ou effets de spécularité), qui peuvent apparaître dans leur voisinage. La représentation de l'apparence d'un objet ou d'une classe d'objets, par un ensemble de mesures locales  $\mathcal{B} = \{\mathbf{m}_1, \dots, \mathbf{m}_K\}$ , est alors naturellement robuste à de tels effets, car quelques mesures non perturbées suffisent à le détecter et le reconnaître.

L'ensemble  $\mathcal{B}$  peut contenir beaucoup de signatures locales si l'on considère la totalité de la surface image d'un objet, ou d'une classe d'objets. Afin de réduire la dimension de représentation,  $\mathcal{B}$  peut être efficacement représenté par la distribution de ses mesures :  $\mathcal{P}(\mathbf{m}|\mathcal{B})$ . Cette densité de probabilité peut être facilement approchée par un histogramme multi-dimensionnel. Cette méthode, non-paramétrique, est bien adaptée à des mesures de petites dimensions.

Le chapitre 3 présente les signatures locales utilisées pour décrire nos objets d'intérêt, ainsi que le calcul de leurs histogrammes. Plusieurs méthodes de détection et de reconnaissance, utilisant une distribution de mesures locales, ont été proposées dans l'état de l'art (paragraphe 2.2). Cependant, l'apprentissage préalable de signatures de référence  $\mathcal{P}(\mathbf{m}|\mathcal{B})$ , est difficile car elles peuvent changer suivant :

- les conditions d'illumination,
- les réglages du système d'acquisition,
- les conditions de prises de vue : l'angle ou distance de la caméra par rapport à l'objet implique des variations d'échelle, ou de perspective des objets dans les images,
- les variabilités intrinsèques des objets, induites, par exemple, par l'usure des matériaux (panneaux), ou par leur nature (arbres).

Détecter et reconnaître efficacement nos objets d'intérêt par ces méthodes, oblige à prendre en compte l'ensemble de ces facteurs.

Afin d'éviter cet apprentissage exhaustif, nous proposons au chapitre 4, d'utiliser le comportement statistique temporel particulier de nos objets d'intérêt dans les séquences de scènes routières, pour les détecter. En effet, nos objets apparaissent à petite échelle dans une première image, puis grossissent rapidement pour subitement disparaître. L'analyse des changements statistiques dans la séquence permet d'isoler de tels événements et de localiser dans les images, les objets qui en sont la cause.



## Chapitre 3

# Mesures locales d'apparence

Ce chapitre présente les mesures locales choisies pour caractériser nos objets d'intérêt. Le choix des caractéristiques dépend de plusieurs facteurs tels que :

- ✓ le ou les capteurs considérés : dans notre application, on dispose d'une caméra couleur,
- ✓ la tâche à réaliser : on cherche à détecter et reconnaître des objets fixes sur l'accotement routier,
- ✓ la classe  $\mathcal{B}$  d'objets d'intérêt : on s'intéresse à des objets manufacturés (panneaux, poteaux, etc.) et naturels (arbres),
- ✓ du contexte dans lequel ils apparaissent : les scènes routières sont filmées sous un éclairage naturel de jour.

Le choix des mesures locales s'effectue naturellement en choisissant celles qui, conjointement, caractérisent la classe  $\mathcal{B}$  et la discriminent de son contexte (noté  $\bar{\mathcal{B}}$ ). La couleur est un attribut classiquement utilisé pour détecter les panneaux routiers. En effet, leurs couleurs vives permettent généralement de les discriminer de leur environnement moins coloré. L'information couleur n'est toutefois pas suffisante pour qualifier efficacement nos objets d'intérêt. Ceux-ci sont manufacturés et présentent des contours rectilignes qui les distinguent des autres éléments naturels de la scène. La seconde mesure choisie est donc une mesure de forme qualifiant cette information. Ces deux mesures reposent sur les dérivées d'ordre 0 et 1 jugées plus stables au bruit [102]. La section 3.1 explicite ces mesures de couleur et de forme.

Une approche classique pour résumer l'information contenue par une collection de mesures locales, est d'approcher leur distribution par un histogramme [131, 133]. Cependant, cette représentation est sensible au bruit notamment dû à la présence de discontinuités aux frontières de classes. Une autre méthode de calcul de distribution, moins sensible au bruit, est proposée à la section 3.2.

L'information de forme est plus pertinente lorsqu'elle est extraite sur les contours des images et présente peu d'intérêt lorsqu'elle est extraite sur les zones uniformes. C'est pourquoi les techniques classiques proposent de segmenter l'image pour ne retenir que l'information relative à ses contours. Afin d'éviter cette segmentation, un intérêt particulier est porté aux histogrammes de forme à la section 3.3.

## 3.1 Mesures locales d'apparence

### 3.1.1 Information couleur

Notre système d'acquisition fournit des images avec des pixels codés dans l'espace couleur *RGB*. Les méthodes de pré-détection proposées au chapitre 4 comparent les histogrammes de mesures couleurs calculés sur deux images successives d'une même séquence. Si l'on suppose que les conditions lumineuses ne changent pas entre deux images successives, et que le système d'acquisition reste stable, on peut envisager de coder les mesures couleur en *RGB*. Cependant, le système d'acquisition n'est pas stable et induit des gains différents sur les images successives. Cette variation est principalement perceptible sur la composante de luminance. Aussi avons nous choisi l'espace *rgL* qui sépare les deux chrominances (*r*, *g*) de la luminance *L* :

$$\begin{cases} r(x, y) = \frac{R(x, y)}{R(x, y) + G(x, y) + B(x, y)} \\ g(x, y) = \frac{G(x, y)}{R(x, y) + G(x, y) + B(x, y)} \\ L(x, y) = R(x, y) + G(x, y) + B(x, y) \end{cases}$$

L'instabilité de la composante *L* sera mise en évidence expérimentalement au chapitre suivant. Finalement, seules les deux chrominances seront utilisées.

### 3.1.2 Mesures locales de forme

La forme, qualifiée par trois mesures, la norme, l'angle du gradient et l'alignement, est choisie pour caractériser les contours rectilignes présents sur les objets manufacturés (poteaux panneaux,...) recherchés. Ces mesures sont issues des dérivées de degré 1 de la surface d'intensité de l'image, restreinte à la luminance *L*(*x*, *y*). On note *L*<sub>*x*</sub>(*x*, *y*) et *L*<sub>*y*</sub>(*x*, *y*) ces dérivées premières dans les directions *x* et *y*. Elles sont calculées rapidement de manière récursive [40]. La figure 3.1 présente une image d'un losange blanc sur un fond noir bruitée par un bruit gaussien. Les images des mesures de norme, d'angle et d'alignement *y* sont présentées également.

**Norme du gradient.** En chaque position (*x*, *y*) de l'image *L*(*x*, *y*), la norme du gradient est définie par :

$$N(x, y) = \sqrt{L_x^2(x, y) + L_y^2(x, y)}$$

Les fortes valeurs de la norme permettent de localiser les contours des objets dans les images (cf. figure 3.1).

**Angle du gradient.** La direction  $\theta$  du gradient est définie par :

$$\theta(x, y) = \arctan_2 \left( \frac{L_y(x, y)}{L_x(x, y)} \right)$$

La fonction notée  $\arctan_2$ , permet de calculer la valeur de l'angle  $\theta$  comprise entre  $-\pi$  et  $+\pi$ . Sur la figure 3.1, on remarque que cette mesure est très bruitée sur les zones uniformes où le gradient a une direction instable. On montre à l'annexe A que la distribution de l'angle sur une zone uniforme perturbée par un bruit gaussien, est une distribution uniforme.

**Alignement.** Si les deux positions  $p_1 = (x_1, y_1)$  et  $p_2 = (x_2, y_2)$  appartiennent au même contour rectiligne de l'image, alors l'angle de contour en  $P_1$  et  $P_2$  est le même  $\theta_1 = \theta_2 = \theta$ , et les deux points sont alignés suivant la droite d'équation :

$$\cos \theta x_1 + \sin \theta y_1 = \cos \theta x_2 + \sin \theta y_2 = \text{constante} = \alpha$$

$\alpha$  est le coefficient d'alignement et se calcule par :

$$\alpha(x, y) = \frac{x L_x(x, y) + y L_y(x, y)}{N(x, y)}$$

car  $\cos \theta(x, y) = \frac{L_x(x, y)}{N(x, y)}$  et  $\sin \theta(x, y) = \frac{L_y(x, y)}{N(x, y)}$ .

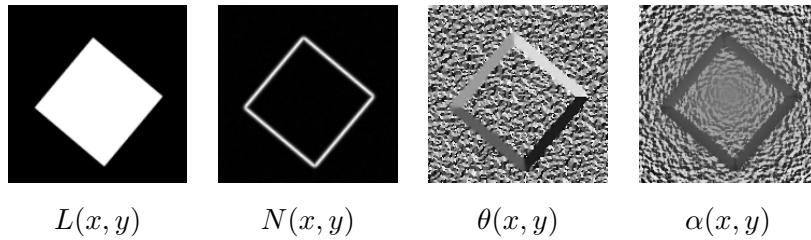


FIG. 3.1: Mesures de forme l'image  $L(x, y)$  : la norme, l'angle du gradient et l'alignement.

## 3.2 Estimation de la distribution des mesures locales

Les méthodes considérées pour calculer la fonction de densité de probabilité sont non paramétriques c'est-à-dire que le type de la fonction de distribution n'est pas spécifié à l'avance et qu'il dépend uniquement des données observées [11].

Etant donné un ensemble de mesures  $\mathcal{B} = \{\mathbf{m}_1, \dots, \mathbf{m}_K\}$  prélevé dans un espace de dimension  $d$ , on cherche à estimer la distribution qui les a générés. L'espace des mesures est quantifié en  $C$  classes de tailles égales (cf. figure 3.2). On note  $D$  la matrice diagonale :

$$D = \begin{bmatrix} D_1 & & 0 \\ & \ddots & \\ 0 & & D_d \end{bmatrix}$$

où  $\frac{1}{D_i}$  correspond à la largeur des classes sur chaque axe  $i \in \{1, \dots, d\}$ . Le déterminant  $\det(D) = \prod_{i=1}^d D_i$  correspond à l'inverse du volume d'une classe.

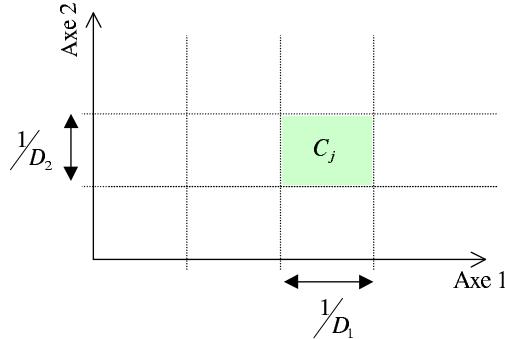


FIG. 3.2: Partition en classes de l'espace des mesures : exemple en dimension  $d = 2$ .

### 3.2.1 Histogrammes classiques

L'histogramme  $\mathbf{h}$  des  $K$  mesures de  $\mathcal{B}$  est calculé  $\forall j \in \{1, \dots, C\}$  par :

$$h_j = \frac{1}{K} \sum_{k=1}^K \begin{cases} 1 & \text{si } \mathbf{m}_k \in C_j \\ 0 & \text{sinon} \end{cases} \quad (3.1)$$

$h_j$  est la proportion de mesures de  $\mathcal{B}$  qui appartiennent à la classe  $C_j$ . L'équation 3.1 peut se réécrire :

$$h_j = \frac{1}{K} \sum_{k=1}^K \phi_u(D(\mathbf{m}_k - \mathbf{c}_j)) \quad (3.2)$$

où  $\mathbf{c}_j$  est le centre de la classe  $C_j$  et  $\phi_u$  est la fonction rectangle de  $\mathbb{R}^d$  dans  $\mathbb{R}$  telle que :

- $\phi_u(\mathbf{x}) = 1$  si  $\mathbf{x}$  appartient à l'hypercube de côté égal à 1 et centré sur  $0_{\mathbb{R}^d}$ ,
- $\phi_u(\mathbf{x}) = 0$  sinon.

La figure 3.3 présente la fonction  $\phi_u$  pour la dimension  $d = 1$ .  $h_j$  estime la probabilité que la mesure  $\mathbf{m}$  appartienne à la classe  $C_j$  de volume  $\frac{1}{\det(D)}$  [11]. La densité de probabilité d'une mesure  $\mathbf{m}$  connaissant  $\mathcal{B}$  s'écrit :

$$\mathcal{P}(\mathbf{m}|\mathcal{B}) = \det(D) \sum_{j=1}^C h_j \phi_u(D(\mathbf{m} - \mathbf{c}_j)) \quad (3.3)$$

et on retrouve facilement :

$$h_j = \int_{C_j} \mathcal{P}(\mathbf{m}|\mathcal{B}) d\mathbf{m}$$

L'histogramme  $\mathbf{h}$  permet de représenter de manière compacte ( $C$  valeurs) la distribution d'un ensemble de mesures locales. Cependant quelques difficultés existent telles que [11] :

- le choix du nombre de classes  $C$  qui conditionne le caractère plus ou moins lisse de la distribution estimée,
- dans le cas où la dimension  $d$  des mesures est grande, l'estimation de la distribution nécessite un nombre  $K$  de mesures élevé,
- l'histogramme présente des discontinuités aux frontières de classes.

### 3.2.2 Histogrammes interpolés

La fonction  $\phi_u$  peut être interprétée comme une fonction d'interpolation [130]. Le choix de fonctions d'interpolation plus lisses que la fonction rectangle (fonction linéaire ou gaussienne) permet d'obtenir des distributions plus lisses. Afin de limiter les problèmes de discontinuité aux frontières des classes, on propose d'utiliser une fonction d'interpolation linéaire notée  $\phi_l$ . Dans ce cas, l'histogramme interpolé se calcule par :

$$h_j = \frac{1}{K} \sum_{k=1}^K \phi_l(D(\mathbf{m}_k - \mathbf{c}_j))$$

et la densité de probabilité de  $\mathbf{m}$  s'écrit :

$$\mathcal{P}(\mathbf{m}|\mathcal{B}) = \det(D) \sum_{j=1}^C h_j \phi_l(D(\mathbf{m} - \mathbf{c}_j)) \quad (3.4)$$

La fonction  $\phi_l$  est représentée à la figure 3.3 pour la dimension  $d = 1$ . D'autres fonctions d'interpolation peuvent être utilisées [130] :  $\phi_l$  a été choisi pour sa simplicité d'utilisation pour le calcul des coefficients  $\mathbf{h} = \{h_1, \dots, h_C\}$ .

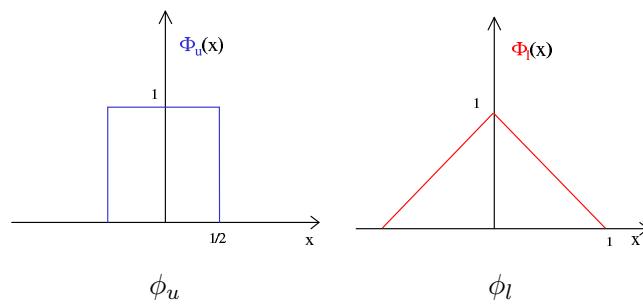


FIG. 3.3:Fonctions d'interpolation utilisées dans le calcul des histogrammes classiques et interpolés ( $d = 1$ ).

**Exemple.** On considère une distribution gaussienne de moyenne 128 et d'écart type 10. La densité de probabilité théorique est représentée à la figure 3.4.

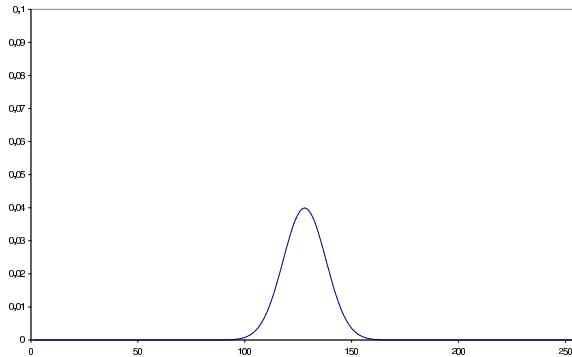


FIG. 3.4: *Densité de probabilité gaussienne théorique.*

La figure 3.5 présente les résultats d'approximation de cette densité de probabilité par un histogramme classique (cf. équation (3.3)) et un histogramme calculé avec une fonction d'interpolation linéaire (cf. équation (3.4)), pour deux quantifications différentes de l'espace des mesures. On remarque

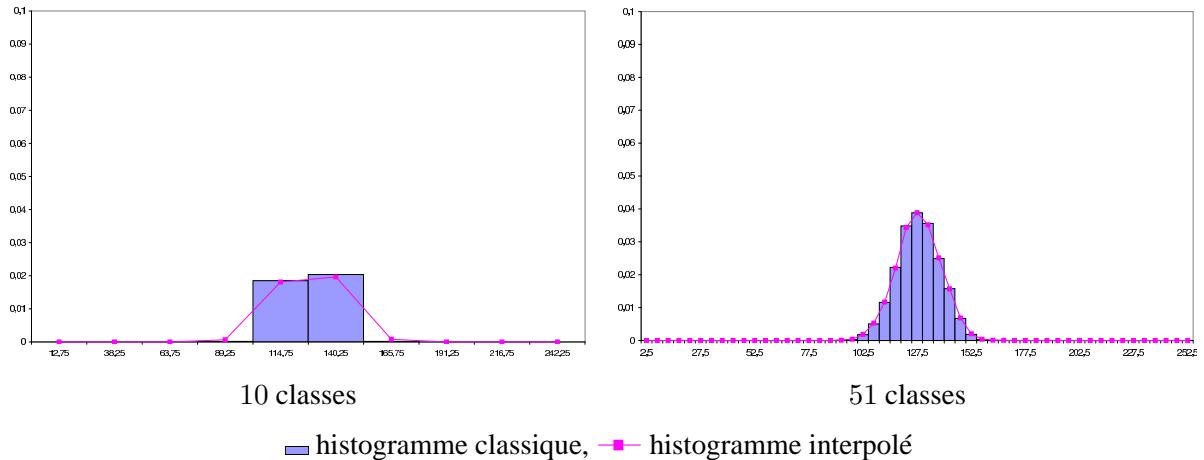


FIG. 3.5: *Différence entre l'approximation de p.d.f. par histogrammes classique et interpolé.*

que l'approximation par histogramme interpolé est plus lisse que l'histogramme classique et est plus proche de la distribution théorique.

### 3.2.3 Méthodes des noyaux

Les fenêtres de Parzen sont une réminiscence des histogrammes à la différence près que les classes ne sont pas définies à l'avance. La densité de probabilité  $\mathcal{P}(\mathbf{m}|\mathcal{B})$  est calculée par l'expression :

$$\mathcal{P}(\mathbf{m}|\mathcal{B}) = \frac{|\mathcal{D}|}{K} \sum_{k=1}^K \phi_u (\mathcal{D}(\mathbf{m} - \mathbf{m}_k))$$

où les centres des classes sont définis par les données. Utiliser la fonction rectangle  $\phi_u$  implique toujours l'apparition de discontinuités aux frontières de classes. Pour les éviter, d'autres formes de noyaux peuvent être utilisées [11].

La méthode des noyaux permet d'approcher la densité de probabilité  $\mathcal{P}(\mathbf{m}|\mathcal{B})$ , mais nécessite de conserver en mémoire l'intégralité des  $K$  mesures de  $\mathcal{B}$ . On n'a donc plus une représentation compacte de la distribution de probabilité à  $C$  paramètres comme pour les histogrammes. C'est pourquoi cette méthode n'a pas été utilisée pour calculer les distributions des mesures locales.

### 3.2.4 Comparaison entre les histogrammes calculés avec $\phi_u$ et $\phi_l$

Les images des séquences de scènes routières, sont perturbées par la présence de bruit gaussien, et par un gain changeant entre images successives. Pour évaluer l'amélioration apportée par des histogrammes calculés avec des fonctions d'interpolation linéaires, deux tests sont réalisés. Le premier s'intéresse à l'influence d'un bruit gaussien sur la comparaison d'histogrammes [131]. La distance  $\mathcal{L}_1$  est calculée entre l'histogramme  $\mathbf{h}$  d'une image de luminance  $L$  et l'histogramme  $\mathbf{h}_b$  de l'image bruitée  $L + b$  où  $b$  est un bruit gaussien de moyenne nulle et d'écart-type  $\sigma$ . Le second test évalue l'influence d'un gain sur la comparaison d'histogramme. La distance  $\mathcal{L}_1$  est calculée entre l'histogramme  $\mathbf{h}$  de l'image  $L$  avec l'histogramme  $\mathbf{h}_g$  de l'image obtenue en appliquant un gain  $g$  à l'image  $L$ , notée  $g \times L$ . Ces deux tests sont réalisés avec des histogrammes calculés avec  $\phi_u$  et  $\phi_l$ .

**Bruit Gaussien.** Le test a été réalisé sur des histogrammes de luminance quantifiés à 256 et 8 classes. Les résultats obtenus sont présentés à la figure 3.6. La distance  $\mathcal{L}_1(\mathbf{h}, \mathbf{h}_b)$  est reportée sur l'axe des ordonnées en fonction de l'écart type du bruit  $\sigma$ . Chaque point de la courbe correspond à la valeur moyenne de  $\mathcal{L}_1(\mathbf{h}, \mathbf{h}_b)$  obtenue sur 40 images d'une séquence de scènes routières.

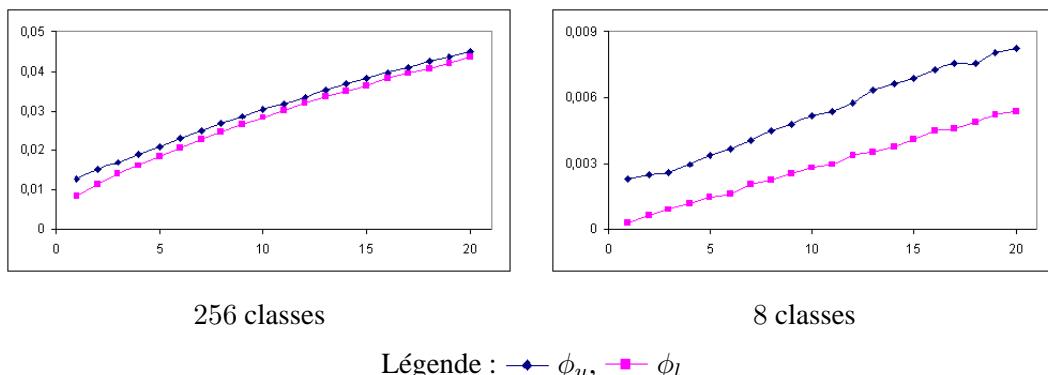


FIG. 3.6: Influence d'un bruit gaussien sur les histogrammes : distance  $\mathcal{L}_1(\mathbf{h}, \mathbf{h}_b)$  en fonction de l'écart type du bruit  $\sigma$ .

Les histogrammes calculés avec la fonction d’interpolation linéaire donnent de meilleurs résultats de similarité ( $\mathcal{L}_1$ ) pour une perturbation de type bruit gaussien.

**Gain.** Les résultats obtenus sont présentés à la figure 3.7. La distance  $\mathcal{L}_1(\mathbf{h}, \mathbf{h}_g)$  est reportée sur l’axe des ordonnées en fonction du gain. Chaque point de la courbe correspond à la valeur moyenne de  $\mathcal{L}_1(\mathbf{h}, \mathbf{h}_g)$  calculée sur 40 images. Les histogrammes calculés avec la fonction d’interpolation  $\phi_l$  sont

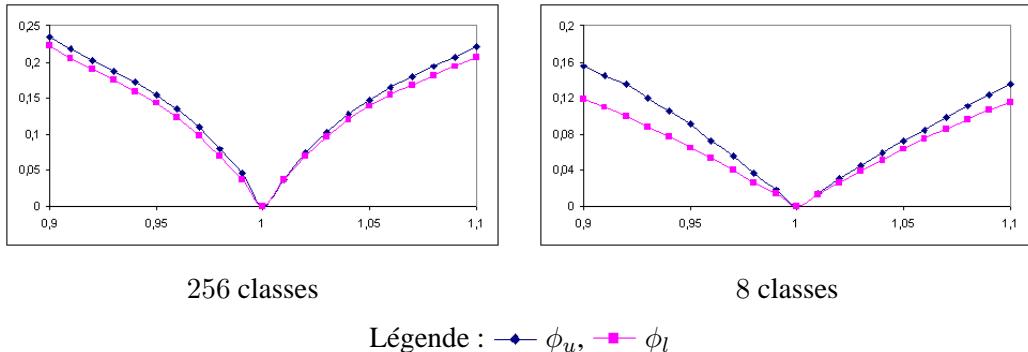


FIG. 3.7: *Influence d’un gain sur les histogrammes : distance  $\mathcal{L}_1(\mathbf{h}, \mathbf{h}_g)$  en fonction d’un gain  $g$ .*

moins sensibles à la présence d’un gain entre deux images. La présence d’un gain provoque des changements de classes, surtout pour les grandes valeurs de luminance où la variation  $\delta = |L - g L| = L |1 - g|$  est très importante. Ce phénomène sera mis en évidence sur des données expérimentales dans le chapitre 4.

### 3.3 Histogrammes de forme

Cette section rappelle deux approches utilisées pour calculer des histogrammes de mesures d’angle. Cette mesure est pertinente lorsqu’elle est calculée sur les contours des images. C’est pourquoi les deux premières méthodes proposent de limiter l’influence des zones uniformes dans le calcul des histogrammes d’angle. La pertinence des mesures d’angle étant proportionnelle à la norme du gradient, nous proposons dans une troisième méthode, de calculer des histogrammes de forme en utilisant la mesure de norme du gradient.

#### 3.3.1 Segmentation des contours

L’approche usuelle présentée dans l’état de l’art, consiste à segmenter l’image pour ne retenir que les informations de contour, et de réaliser un histogramme des mesures de forme (angle) que sur les points retenus (cf. section 2.1.2.2). Cependant, le choix d’un seuil pour déterminer les contours est délicat et peut conduire à des erreurs de classification entre les contours et les zones uniformes.

### 3.3.2 Histogramme pondéré

Afin de limiter ou d'augmenter l'influence des pixels appartenant au contour dans le calcul des histogrammes couleurs, Vertan et Boujema propose d'utiliser des histogrammes pondérés [156]. Dans les histogrammes classiques, chaque mesure pondère la classe à laquelle elle appartient, de manière équivalente par une pondération  $\frac{1}{K}$  (cf. équation (3.1)). Dans les histogrammes pondérés, on peut choisir des pondérations qui contrôlent l'influence de chaque mesure. Ainsi en pondérant faiblement les mesures d'angle calculées sur les zones uniformes, et en pondérant fortement les mesures d'angle calculées sur les contours de l'image, on peut définir des histogrammes de forme. Cette approche est une extension naturelle à la méthode présentée au paragraphe 3.3.1 où les pondérations sont binaires : pondération 0 pour les mesures des zones uniformes et 1 pour celles des zones contours. Cependant, le choix d'une fonction de pondération reste délicat car il nécessite le réglage des paramètres qui conditionnent la classification contour/non-contour au même titre que le seuil évoqué au paragraphe 3.3.1.

### 3.3.3 Histogramme des mesures $(\theta, \alpha, N)$

Les mesures d'angle (et donc d'alignement) qualifient efficacement les contours d'une image, mais présentent peu d'intérêt lorsqu'elles sont calculées sur des zones uniformes. A l'annexe A, on montre que l'angle  $\theta(x, y)$  calculée à la position  $(x, y)$  d'une image, a une variance qui dépend de la mesure de norme du gradient  $N(x, y)$ . Sur les zones uniformes bruitées par un bruit gaussien, la distribution de l'angle est une distribution uniforme, ce qui confirme que cette mesure n'est pas pertinente pour caractériser ce type de régions [38]. Dans cette thèse, nous avons pris la partie de ne pas recourir aux méthodes de segmentation qui conditionneraient l'ensemble des résultats. On propose de calculer l'histogramme de forme en considérant conjointement les trois mesures d'angle, d'alignement et de norme du gradient :  $(\theta, \alpha, N)$ . Cet histogramme tri-dimensionnel conserve l'information relative aux deux de régions : uniformes pour les faibles valeurs de la norme du gradient, et contours pour les fortes valeurs de  $N$ .

## 3.4 Conclusion

Ce chapitre a présenté les mesures locales utilisées dans cette thèse. La couleur et la forme ont été choisies pour qualifier nos objets d'intérêt et les discriminer de leur environnement. L'information relative à un ensemble de mesures locales est résumée par leur histogramme. Les histogrammes classiques présentent des discontinuités aux frontières de classes ce qui implique une forte sensibilité au bruit gaussien ou aux effets de gain. C'est pourquoi nous avons proposé de les calculer à l'aide d'une fonction d'interpolation linéaire qui fournit une meilleure approximation de la densité de probabilité.



## Chapitre 4

# Pré-détection non supervisée

Nos séquences d'images de scènes routières sont indexées par la distance et le mouvement de la caméra correspond grossièrement à un travelling avant. Les objets *fixes* sur le bord de la chaussée ont un mouvement apparent relativement important. Ils apparaissent à une petite échelle dans une première image, grossissent rapidement et disparaissent subitement. Leurs changements d'apparence dans la séquence sont principalement dus au changement d'échelle (apparition, grossissement, disparition) et au changement de leur position dans les images. La figure 4.1 présente trois images successives de séquence de scènes routières où un panneau grossit puis disparaît. Dans cet exemple, la distance entre les images est de cinq mètres.

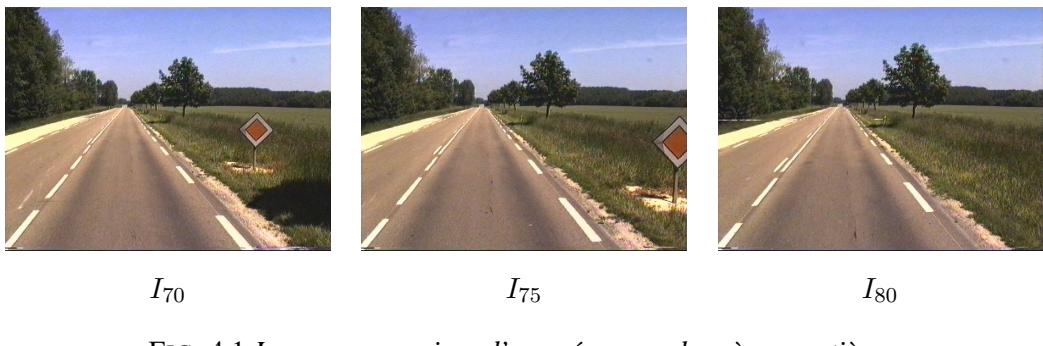


FIG. 4.1: Images successives d'une séquence de scènes routières.

La méthode que nous proposons s'appuie sur la distribution statistique de mesures locales (de couleur et de forme) pour détecter les changements d'apparence relatifs aux objets d'intérêt. Elle repose sur l'analyse des changements statistiques dans la séquence vidéo. A partir d'hypothèses relatives aux comportements statistiques des objets d'intérêt et du reste de la scène, nous proposons un système non supervisé de détection d'événements complexes dans des séquences d'images acquises par une caméra en mouvement [32, 30]. De tels événements peuvent correspondre à la sortie, l'entrée ou au changement d'apparence d'objets dans une scène en mouvement global.

La section 4.1 présente des extracteurs permettant d'isoler les statistiques relatives aux objets d'intérêt.

rêt. Celles-ci sont utilisées en section 4.2 pour localiser les objets dans la séquence grâce à la technique de rétroposition bayésienne. Cette méthode originale de détection est illustrée sur une séquence d'images de scènes routières, avec les mesures locales de couleur et de forme présentées dans le chapitre 3. Les résultats obtenus apparaissent en section 4.3.

## 4.1 Détection de changements statistiques

Pour détecter les changements statistiques, nous définissons les extracteurs :

$$\left| \begin{array}{l} \Delta_+(\mathbf{m}, t) = \max(\mathcal{P}(\mathbf{m}, t) - \gamma \mathcal{P}(\mathbf{m}, t+1), 0) \\ \Delta_-(\mathbf{m}, t) = \max(\mathcal{P}(\mathbf{m}, t) - \gamma \mathcal{P}(\mathbf{m}, t-1), 0) \end{array} \right.$$

où  $\mathcal{P}(\mathbf{m}, t)$  est la distribution des mesures locales  $\mathbf{m}$  dans l'image  $I(t)$ . On suppose que  $\mathcal{P}(\mathbf{m}, t)$  est un mélange de deux lois,  $\mathcal{P}(\mathbf{m}, t|\mathcal{B})$  distribution des mesures appartenant à la classe d'objets d'intérêt, et  $\mathcal{P}(\mathbf{m}, t|\bar{\mathcal{B}})$  distribution des mesures appartenant au fond de l'image. Sous certaines hypothèses *a priori* sur le comportement temporel des densités  $\mathcal{P}(\mathbf{m}, t|\bar{\mathcal{B}})$  et  $\mathcal{P}(\mathbf{m}, t|\mathcal{B})$ , nous montrons que les deux extracteurs  $\Delta_+(m, t)$  et  $\Delta_-(m, t)$  permettent d'isoler les densités de probabilité  $\mathcal{P}(\mathbf{m}, t|\mathcal{B})$  des objets d'intérêt de  $I(t)$  qui ont significativement changé respectivement entre les instants  $t$  et  $t-1$  pour  $\Delta_-$  et entre  $t$  et  $t+1$  pour  $\Delta_+$ .

### 4.1.1 Hypothèses relatives à $\mathcal{B}$ et $\bar{\mathcal{B}}$

$H_0$  : **Composition des images  $I(t)$ .** On suppose que chaque image  $I(t)$  de la séquence est composée d'un ensemble d'objets d'intérêt,  $\mathcal{B}$ , et d'un arrière plan,  $\bar{\mathcal{B}}$ . La densité de probabilité  $\mathcal{P}(\mathbf{m}, t)$  de mesures locales  $\mathbf{m}$  dans l'image  $I(t)$  peut se décrire par la loi de mélange :

$$\mathcal{P}(\mathbf{m}, t) = (1 - \epsilon(t)) \mathcal{P}(\mathbf{m}, t|\bar{\mathcal{B}}) + \epsilon(t) \mathcal{P}(\mathbf{m}, t|\mathcal{B})$$

$H_1$  : **Proportion de  $\mathcal{B}$ .** L'hypothèse  $H_1$  contraint la proportion  $\epsilon$  des objets d'intérêt et est liée au choix du paramètre  $\gamma$  dans les extracteurs. On suppose que  $\epsilon$  appartient à l'intervalle  $[0, \frac{\gamma-1}{\gamma}]$ .

$H_2$  : **Variations statistiques de  $\bar{\mathcal{B}}$ .** On suppose que les statistiques du fond  $\bar{\mathcal{B}}$  varient faiblement au cours du temps, soit :

$$\mathcal{P}(\mathbf{m}, t|\bar{\mathcal{B}}) \simeq \mathcal{P}(\mathbf{m}, t+1|\bar{\mathcal{B}})$$

$H_3$  : **Différence statistique entre  $\mathcal{B}$  et  $\bar{\mathcal{B}}$ .** Cette hypothèse suppose que l'on peut choisir une mesure locale  $\mathbf{m}$  qui permet de discriminer les objets de  $\mathcal{B}$  du fond  $\bar{\mathcal{B}}$ , ou :

$$\mathcal{P}(\mathbf{m}, t | \bar{\mathcal{B}}) \cdot \mathcal{P}(\mathbf{m}, t | \mathcal{B}) \simeq 0 \quad \forall \mathbf{m}$$

Comme les densités de probabilité sont des fonctions positives, cette hypothèse est équivalente à :

$$\begin{cases} \text{si } \mathcal{P}(\mathbf{m}, t | \bar{\mathcal{B}}) > 0 \text{ alors } \mathcal{P}(\mathbf{m}, t | \mathcal{B}) \simeq 0 \\ \text{si } \mathcal{P}(\mathbf{m}, t | \mathcal{B}) > 0 \text{ alors } \mathcal{P}(\mathbf{m}, t | \bar{\mathcal{B}}) \simeq 0 \end{cases}$$

$H_4$  : **Variations statistiques de  $\mathcal{B}$ .** En pratique, nous sommes intéressés par les objets changeants qui apparaissent ou disparaissent de la scène, ou ceux qui ont des mesures locales qui changent significativement au cours du temps. Ainsi deux cas peuvent être envisagés. Le premier,  $H_4$ , considère que  $\mathcal{P}(\mathbf{m}, t | \mathcal{B})$  change fortement au cours du temps :

$$\mathcal{P}(\mathbf{m}, t | \mathcal{B}) \cdot \mathcal{P}(\mathbf{m}, t + 1 | \mathcal{B}) \simeq 0 \quad \forall \mathbf{m}$$

équivalent à :

$$H_4 : \begin{cases} \text{si } \mathcal{P}(\mathbf{m}, t | \mathcal{B}) > 0 \text{ alors } \mathcal{P}(\mathbf{m}, t + 1 | \mathcal{B}) \simeq 0 \\ \text{si } \mathcal{P}(\mathbf{m}, t + 1 | \mathcal{B}) > 0 \text{ alors } \mathcal{P}(\mathbf{m}, t | \mathcal{B}) \simeq 0 \end{cases}$$

Le second cas  $H'_4$  suppose au contraire que :

$$H'_4 : \mathcal{P}(\mathbf{m}, t | \mathcal{B}) \simeq \mathcal{P}(\mathbf{m}, t + 1 | \mathcal{B})$$

mais que c'est la proportion  $\epsilon(t)$  de l'objet qui varie *fortement*. Cette notion de forte variation est précisée au paragraphe suivant.

### 4.1.2 Extracteurs d'évènements

Lorsque les hypothèses  $H_0$ ,  $H_1$ ,  $H_2$ ,  $H_3$  et  $H_4$  sont vérifiées entre les instants  $t$  et  $t + 1$ , et entre  $t$  et  $t - 1$  alors on montre dans l'annexe B que :

$$\left| \begin{array}{l} \Delta_+(\mathbf{m}, t) = \epsilon(t) \mathcal{P}(\mathbf{m}, t | \mathcal{B}) \\ \Delta_-(\mathbf{m}, t) = \epsilon(t) \mathcal{P}(\mathbf{m}, t | \mathcal{B}) \end{array} \right. \quad (4.1)$$

Les densités de probabilité devant s'intégrer à 1, on peut facilement calculer  $\epsilon(t)$  et extraire  $\mathcal{P}(\mathbf{m}, t | \mathcal{B})$ . Ce résultat ne dépend pas du paramètre  $\gamma$  choisi. Dans le cas où c'est l'hypothèse  $H'_4$  qui est vérifiée, on montre (cf. annexe B) :

$$\left| \begin{array}{l} \Delta_+(\mathbf{m}, t) = \max(\epsilon(t) - \gamma \epsilon(t + 1), 0) \mathcal{P}(\mathbf{m}, t | \mathcal{B}) \\ \Delta_-(\mathbf{m}, t) = \max(\epsilon(t) - \gamma \epsilon(t - 1), 0) \mathcal{P}(\mathbf{m}, t | \mathcal{B}) \end{array} \right. \quad (4.2)$$

Le tableau 4.1 illustre dans un cas simple, le fonctionnement de l'extracteur  $\Delta_+$  sur deux densités de probabilité approchées par des histogrammes monodimensionnels :  $\mathbf{m}$  en abscisse et  $\mathcal{P}(\mathbf{m})$  en ordonnée. On montre les différentes étapes de calcul de  $\Delta_+$  : le calcul de  $\gamma \mathcal{P}(m, t+1)$ , et la soustraction  $\mathcal{P}(m, t) - \gamma \mathcal{P}(m, t+1)$  dont la partie positive (en rouge) est retenue par l'extracteur  $\Delta_+$ . Quel que soit le paramètre  $\gamma$  choisi, si l'hypothèse  $H_4$  est vérifiée, les détecteurs permettent d'extraire la densité de probabilité de l'objet d'intérêt. Par contre, lorsque  $H'_4$  est vérifiée, l'extraction n'est possible que si la variation de proportion est suffisamment importante soit pour  $\Delta_+ : \epsilon(t) > \gamma \epsilon(t+1)$  et  $\Delta_- : \epsilon(t) > \gamma \epsilon(t-1)$ .

## 4.2 Détection par rétroprojection bayésienne

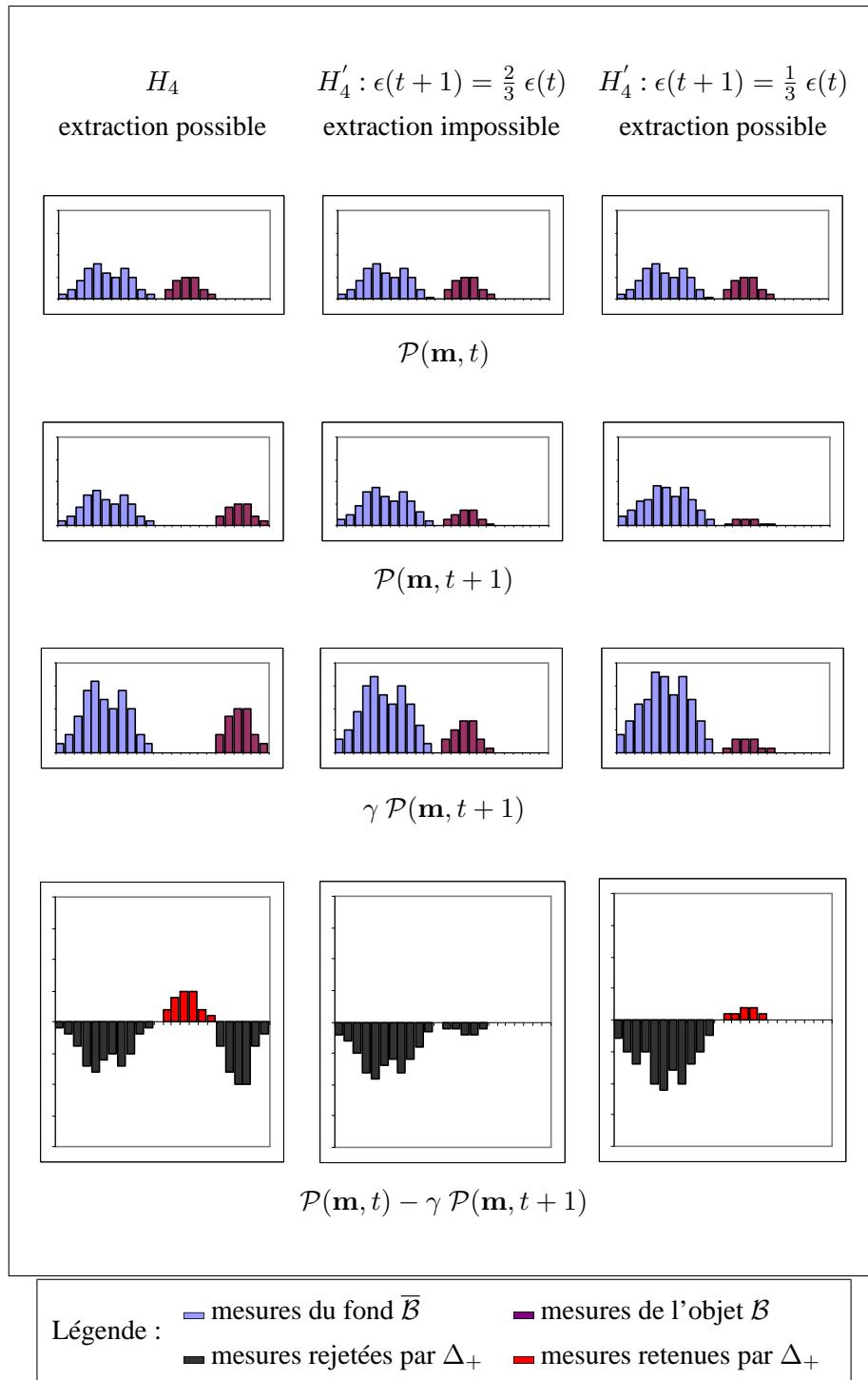
Ayant estimé la distribution des mesures locales de l'objet d'intérêt  $\mathcal{P}(\mathbf{m}, t|\mathcal{B})$ , il convient ensuite de localiser les objets de  $\mathcal{B}$  dans l'image  $I(t)$ . Pour cela, nous utilisons la technique de rétro-projection bayésienne [142, 149]. Pour chaque position  $(x, y)$  de l'image  $I(t)$ , une mesure locale  $\mathbf{m}_{x,y}$  est calculée et sa probabilité *a posteriori*  $\mathcal{P}(\mathcal{B}|\mathbf{m}_{x,y}, t)$  d'appartenir à  $\mathcal{B}$  calculée grâce à la formule de Bayes, est reportée dans une carte de confiance :

$$\mathcal{P}(\mathcal{B}|\mathbf{m}_{x,y}, t) = \frac{\mathcal{P}(\mathbf{m}_{x,y}, t|\mathcal{B}) \mathcal{P}(\mathcal{B})}{\mathcal{P}(\mathbf{m}_{x,y}, t)}$$

La probabilité a priori  $\mathcal{P}(\mathcal{B})$ , constante, n'intervient pas dans la décision. Pour chaque image  $I(t)$  de la séquence, deux cartes complémentaires de rétro-projection sont calculées, l'une dite *en arrière* calculée grâce à  $\Delta_+$ , et l'autre dite *en avant* calculée grâce à  $\Delta_-$ . La rétroprojection *en arrière* permet d'analyser une image grâce à sa suivante dans la séquence : le présent est interprété grâce au futur (d'où la terminologie *en arrière*). La rétroprojection *en avant* permet d'analyser une image grâce à sa précédente dans la séquence : le présent est interprété grâce au passé (d'où la terminologie *en avant*).

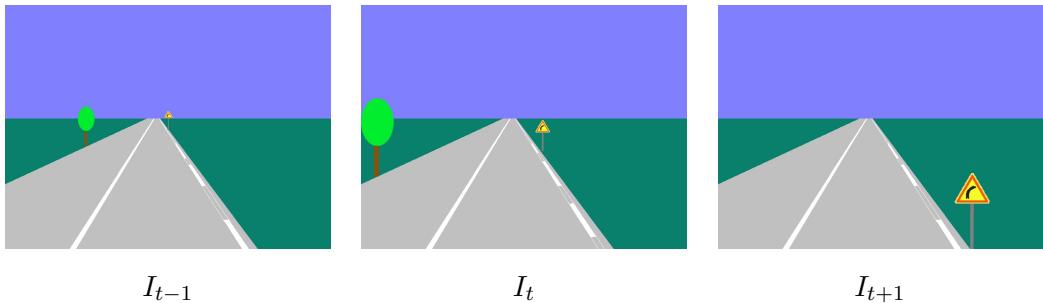
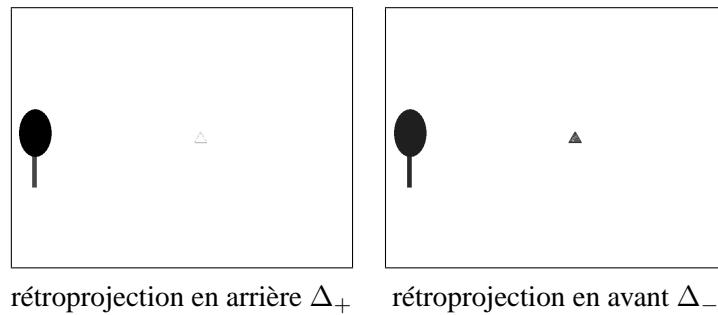
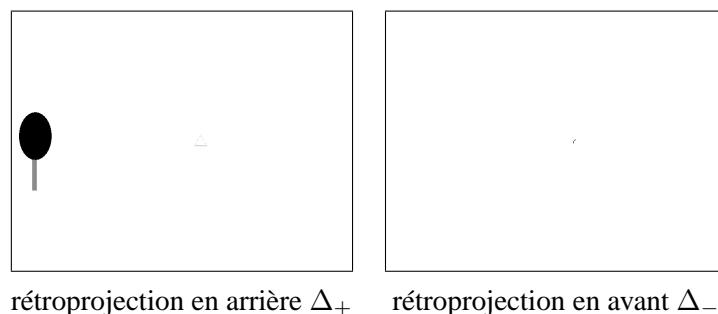
Notre méthode est appliquée à des images synthétiques présentées à la figure 4.2, qui reproduisent trois images successives de scènes routières où un arbre grossit puis disparaît, et un panneau grossit. Les figures 4.3 et 4.4 présentent les résultats de pré-détection obtenus avec les mesures couleur par les extracteurs  $\Delta_+$  et  $\Delta_-$  pour les valeurs  $\gamma = 2$  et  $\gamma = 4$ . On remarque que dans les deux cas la rétroprojection *en arrière* permet de détecter parfaitement la disparition de l'arbre. De la même façon, quelques pixels du contour du panneau sont détectés car leurs couleurs ont changé au changement d'échelle. Ces pixels mixtes (situés sur la frontière entre deux régions de couleurs homogènes) ont une couleur instable. Leur existence a motivé, pour Vertan et al. [156], l'usage des histogrammes pondérés présentés en section 3.3.2. Pour  $\gamma = 2$ , la carte de pré-détection *en avant* permet de localiser le panneau et l'arbre. Ceci n'est pas le cas pour  $\gamma = 4$  car le grossissement de ces deux objets conduit à un facteur d'échelle supérieur à 2 mais inférieur à 4.

La figure 4.5 présente les résultats de pré-détection obtenus avec les mesures de forme décrites à la section 3.1.2, et qui visent à détecter les alignements. Dans ce cas, la localisation des contours du pan-



TAB. 4.1: *Effet de l'extracteur  $\Delta_+$  pour une sensibilité  $\gamma = 2$  dans le cas des hypothèses  $H_4$  et  $H'_4$  avec  $H_0$ ,  $H_1$ ,  $H_2$  et  $H_3$  vraies.*

neau et de l'arbre, est liée aux variations de la mesure  $\alpha$  sensible à la position des objets dans l'image.

FIG. 4.2:*Trois images synthétiques successives.*FIG. 4.3:*Cartes de pré-détection calculées avec les mesures de couleur ( $\gamma = 2$ ).*FIG. 4.4:*Cartes de pré-détection calculées avec les mesures de couleur ( $\gamma = 4$ ).*

L'hypothèse  $H_4$  est alors vérifiée. On remarque également que les contours horizontaux des marquages routiers sont détectés pour la même raison (changement de la mesure  $\alpha$ ), tandis que ceux qui sont parallèles au déplacement du véhicule ne le sont pas : les mesures  $(\theta, \alpha, N)$  ne varient pas entre les images ni en valeur, ni (suffisamment) en proportion.

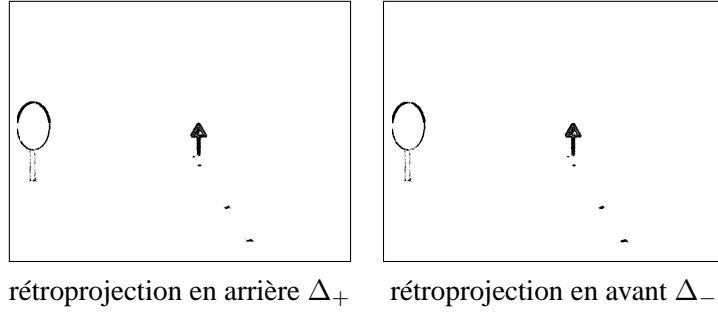


FIG. 4.5: Cartes de pré-détection calculées avec les mesures de forme ( $\gamma = 2$ ).

## 4.3 Résultats de pré-détection

Des tests de pré-détection ont été réalisés sur une base de 1000 images séquentielles de scènes routières, soit 5 km de route de la nationale 4 dans le département 51 (Marne). L'annexe C présente les résultats de pré-détection obtenus pour les 100 premiers mètres. Les cartes de pré-détection sont très discriminantes bien qu'elles soient seuillées à 0 ( i.e. toutes les positions  $(x, y)$  telles que  $\mathcal{P}(\mathcal{B}|\mathbf{m}_{x,y}, t) > 0$  sont retenues).

### 4.3.1 Choix du coefficient $\gamma$

Le paramètre  $\gamma$  permet de régler la sensibilité des détecteurs aux changements statistiques. L'annexe C présente quelques résultats de pré-détection couleur avec différentes valeurs de  $\gamma$  (1.5, 2 et 3). La figure 4.6 présente un résultat de pré-détection obtenu sur l'image  $I_{75}$  (cf. figure 4.1). Plus  $\gamma$  est grand, moins le détecteur est sensible aux changements statistiques dus aux variations de proportion (hypothèse  $H'_4$ ). Par contre, lorsque l'hypothèse  $H_4$  est strictement vérifiée, ce qui est le cas du panneau, la densité  $\mathcal{P}(\mathbf{m}, t|\mathcal{B})$  est extraite quelle que soit la valeur de  $\gamma$  utilisée. NB : Les dernières lignes des images de la

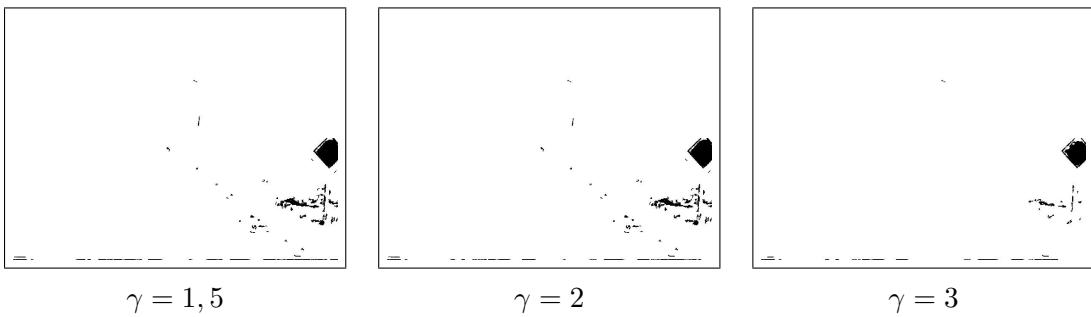


FIG. 4.6: Résultats de pré-détection couleur sur l'image  $I_{75}$  pour différentes valeurs de  $\gamma$ .

séquence sont instables suite à un problème de numérisation et leurs variations statistiques sont perçues presque à chaque image.

On rappelle que  $\gamma$  est lié au domaine de variation de  $\epsilon$  par l'hypothèse  $H_1$ . Pour  $\gamma = 3$ ,  $\epsilon$  est limité à l'intervalle  $[0, \frac{2}{3}]$  et pour  $\gamma = 1,5$ ,  $\epsilon \in [0, \frac{1}{3}]$ . La valeur  $\gamma = 2$  a finalement été choisie car elle est un bon compromis entre la sensibilité des détecteurs à des changements d'échelle, et la résistance au bruit.

### 4.3.2 Pré-détection par mesures de couleur

La figure 4.7 présente le résultat de pré-détection obtenu sur l'image  $I_{75}$  de la séquence, calculé avec des histogrammes classiques de mesures couleurs ( $r, g, L$ ). La figure 4.8 montre le résultat obtenu avec des histogrammes interpolés. La rétroposition *en avant* pour les histogrammes classiques indique qu'une partie importante du ciel a changé de couleur entre les images  $I_{70}$  et  $I_{75}$ . Ce changement est dû à un gain en luminance de la caméra entre les deux images, qui provoque un changement de classe des mesures couleurs du ciel. L'utilisation d'histogrammes interpolés permet de limiter de tels effets liés à l'instabilité du système d'acquisition mais cela ne suffit pas toujours (cf. résultats de l'annexe C). Ce gain perturbateur est principalement présent sur l'information de luminance  $L$ . Aussi, la pré-détection couleur s'est finalement limitée à l'information des chrominances ( $r, g$ ), jugée plus stable.

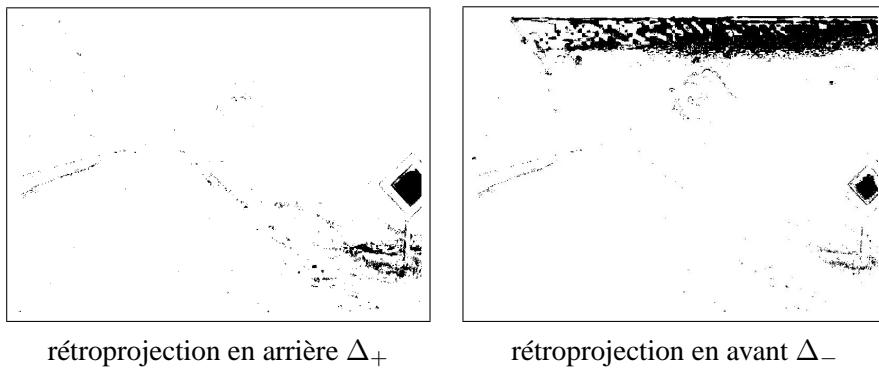


FIG. 4.7: Résultat de rétroposition avec des histogrammes classiques calculés sur des mesures couleurs ( $r, g, L$ ) ( $\gamma = 2$ ).

La partie colorée jaune du panneau est détectée grâce à son changement d'échelle important (hypothèse  $H'_4$ ) dans le cas de la rétroposition *en avant*, et à sa disparition dans le cas de la rétroposition *en arrière* (hypothèse  $H'_4$ ). Quelques mesures des contours des arbres situées dans l'arrière plan sont également détectées car leur couleur est changeante (pixels mixtes).

**Remarque.**  $H_3$  suppose que les objets d'intérêt de  $\mathcal{B}$  ne ressemblent statistiquement pas au fond  $\bar{\mathcal{B}}$ . Cette hypothèse n'est pas vérifiée dans le cas des images présentées à la figure 4.9. La maison au toit rouge est présente dans les deux images successives  $I_{4310}$  et  $I_{4315}$  : c'est un élément de  $\bar{\mathcal{B}}$ . Sa couleur est identique (de même classe) à celle des panneaux disparaissants. La variation statistique des pixels rouges est suffisamment importante pour être extraite par le détecteur  $\Delta_+$  : la rétroposition *en arrière* permet

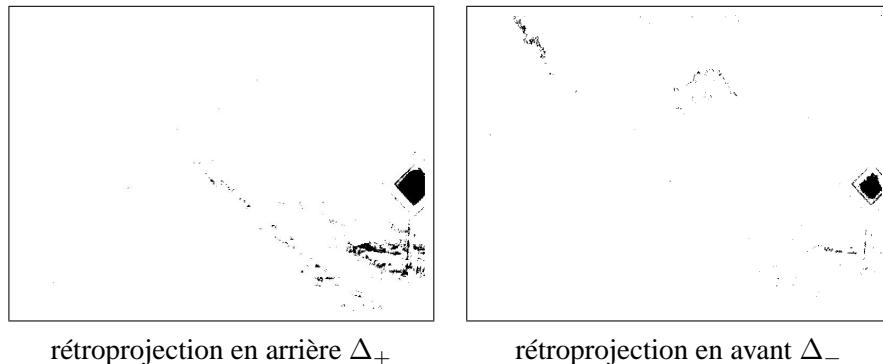


FIG. 4.8: Résultat de rétroprojection avec des histogrammes interpolés calculés sur des mesures couleurs  $(r, g, L)$  ( $\gamma = 2$ ).

de localiser les panneaux mais ne les distingue pas du toit qui a les mêmes statistiques couleurs.

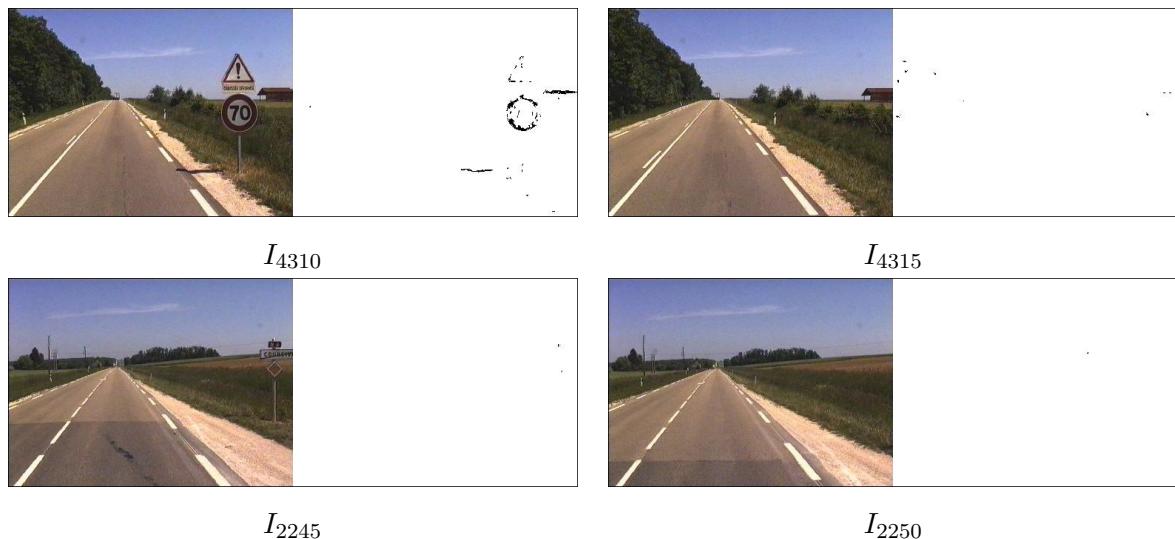


FIG. 4.9: Exemple de résultats de pré-détection couleur ( $\gamma = 2$ ) lorsque l'hypothèse  $H_3$  n'est pas strictement vérifiée.

Dans les images successives  $I_{2245}$  et  $I_{2250}$ , la disparition de la partie jaune du panneau n'est pas détectée car ses statistiques couleurs se confondent avec celles du champ présent en arrière plan. D'autre part, la variation due à la disparition du panneau, trop faible, n'est pas perçue par les détecteurs.

Les mesures couleurs seules ne permettent donc pas de détecter l'ensemble de nos objets d'intérêt. On propose au paragraphe suivant d'utiliser conjointement des mesures de forme.

### 4.3.3 Pré-détection par mesures de forme

La figure 4.10 et la figure 4.11 présentent le résultat de pré-détection sur l'image  $I_{75}$ , réalisé avec des histogrammes classiques et interpolés des mesures de forme. La qualité des résultats est meilleure avec des histogrammes interpolés, moins sensibles au bruit. Nous sommes principalement intéressés par les changements sur la mesure  $\alpha$ , qui indiquent un changement de position de l'alignement. La mesure d'angle  $\theta$  ne varie pas sur nos objets d'intérêt car ils n'y a pas d'effet de rotation entre deux images successives. Les contours, notamment verticaux, du panneau de priorité et de l'arbre qui présentent des mesures  $\alpha$  changeantes (hypothèse  $H_4$ ) sont bien détectés. Sont également détectées des mesures locales caractérisant les contours changeants présents sur la route, les feuillages des arbres ou la texture du bas-côté. Ces résultats de pré-détection par mesures de forme, peuvent être comparés à la détection de contours par la norme du gradient (figure 4.12).

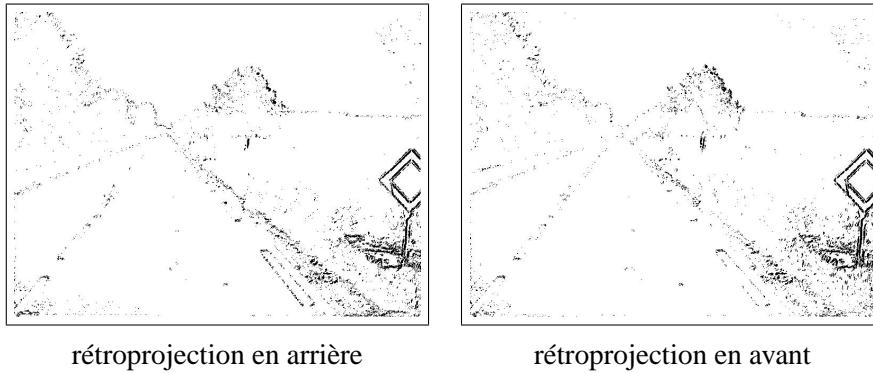


FIG. 4.10: Résultat de rétroposition avec des histogrammes classiques calculés sur des mesures de forme ( $\theta, \alpha, N$ ).

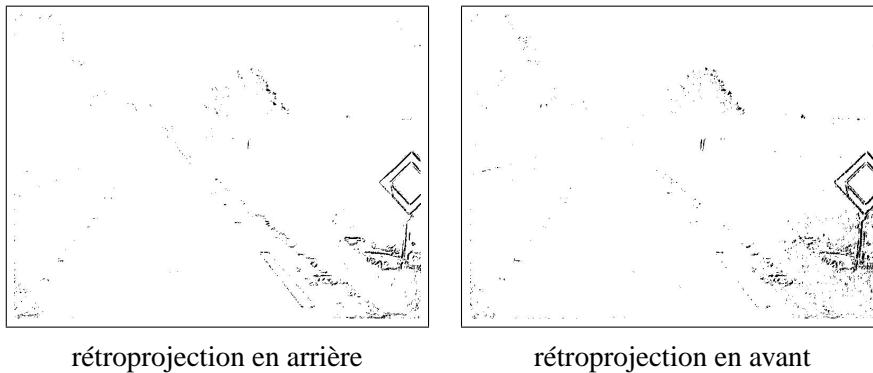
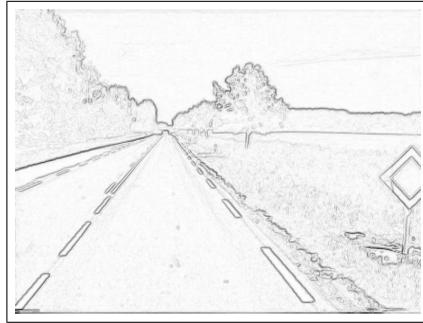
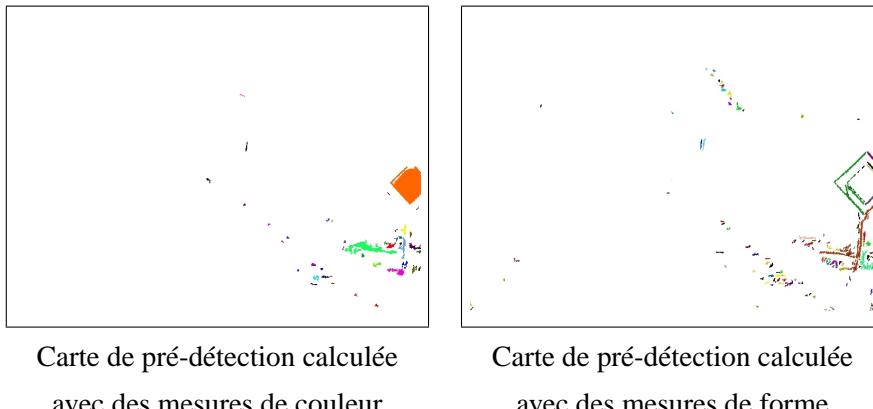


FIG. 4.11: Résultat de rétroposition avec des histogrammes interpolés calculés sur des mesures de forme ( $\theta, \alpha, N$ ).

FIG. 4.12: Norme du gradient,  $N$ , de l'image  $I_{75}$ .

#### 4.3.4 Fusion d'information

La figure 4.13 présente les cartes fusionnées de pré-détection calculées avec des mesures de couleur et de forme : les zones détectées par  $\Delta_+$  et  $\Delta_-$  sont réunies dans une seule carte. Concrètement, les cartes de rétroposition en avant et en arrière sont sommées puis seuillées à 0. Les petites régions (de taille inférieure à 10 pixels) sont éliminés par une analyse en composantes connexes.

FIG. 4.13: Fusion d'information entre la rétroposition en avant et en arrière pour l'image  $I_{75}$ .

La figure 4.14 présente la carte de détection obtenue lorsque que les résultats de rétroposition avec les mesures de couleur et de forme, sont fusionnées. Cette carte permet de mieux localiser les zones intéressantes. Néanmoins, les deux cartes sont conservées séparément pour une analyse ultérieure plus fine, comme par exemple, l'analyse des alignements verticaux dans la carte de pré-détection par mesures de forme.

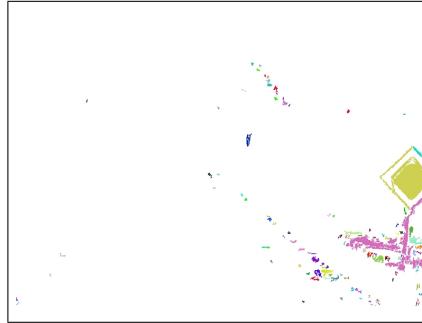


FIG. 4.14:*Fusion des résultats de rétroprojection avec les mesures de couleur et de forme pour l'image  $I_{75}$ .*

### 4.3.5 Résultats quantitatifs sur 5 km

Pour évaluer notre méthode, les troncs d'arbre, les panneaux et les délinéateurs qui ont été détectés, sont comptés. On analyse les cartes de pré-détection fusionnées (rétroréprojection en avant et en arrière) pour les mesures de forme et de couleur.

On considère qu'un objet est détecté par la pré-détection avec les mesures couleurs, si ses régions colorées sont localisées dans au moins une des images de la séquence où l'objet apparaît. De même pour les mesures de forme, un objet est détecté si ses contours sont localisés dans au moins une des images de la séquence où il apparaît.

**Mesures couleurs.** Sur les 5 km de la séquence, apparaissent 21 panneaux. 12 ont leurs parties colorées détectées par notre méthode de pré-détection (soit 57%). Les panneaux non détectés sont ceux qui comportent très peu de couleurs, et ceux situés dans de fortes ombres.

**Mesures de forme.** Lorsqu'on considère les mesures de forme, 17 panneaux sont détectés (81%). Les 4 panneaux non-détectés sont ceux situés dans des zones ombragées où les contours sont peu marqués. Ils ont également été manqués en utilisant les mesures couleurs.

**Eléments verticaux.** La majeure partie des objets d'intérêt présente des éléments verticaux (poteaux, troncs d'arbre, délinéateurs,...). Les mesures ayant un angle vertical sont retenues pour définir des régions d'intérêt. La figure 4.15 présente les mesures verticales retenues de la carte de pré-détection de forme présentée à la figure 4.13. Les régions détectées sont localisées dans l'image originale par des rectangles englobants jaunes. Le poteau du panneau et le tronc d'arbre sont détectés. Des fausses alertes apparaissent sur les régions proches de la caméra dont la texture présente des éléments verticaux comme l'herbe. Celles-ci pourront être écartées par une analyse avec d'autres mesures (par exemple couleur).

Sur les 5 kilomètres de la séquence, 73 délinéateurs sur 78 ont été parfaitement détectés (soit 93%).



FIG. 4.15:*Régions à mesures de forme verticales.*

Les délinéateurs manqués se situent dans l'ombre d'arbres et leurs contours ont une norme de gradient faible. La figure 4.16 présente un exemple de délinéateur non détecté. On remarque sur ces exemples que les troncs d'arbre, les poteaux de panneaux sont parfaitement détectés. La figure 4.17 présente des images de la séquence où les mesures de forme verticales permettent de détecter le délinéateur parfaitement. Sur la séquence, 53 troncs d'arbre ont été détectés sur 54 présents. Les résultats de pré-détection obtenus sur 5 km sont rappelés au tableau 4.2.



FIG. 4.16:*Pré-détection par mesures de forme verticales : l'ombre portée sur le délinéateur atténue ses contours et empêche sa détection.*



FIG. 4.17:*Pré-détection par mesures de forme verticales : le délinéateur est détecté.*

Mesure	couleur	forme
Panneau (21)	57%	81%
Délinéateurs (73)	-	93%
Troncs d'arbres (54)	-	98%

TAB. 4.2:Résultats de pré-détection sur une séquence de 5km.

## 4.4 Conclusion

Ce chapitre a présenté une méthode de détection non supervisée de changements statistiques permettant de localiser dans des séquences d’images avec caméra en mouvement, les objets apparaissants, disparaissants et changeants. Une image est analysée par comparaison de la distribution de ses mesures locales avec celle de l’image précédente (rétrécissement *en avant*), et avec celle de l’image suivante (rétrécissement *en arrière*). Notre méthode utilise peu de connaissance *a priori* : seules quelques hypothèses sur les comportements statistiques des objets d’intérêt et de leur environnement sont utilisées. Ces hypothèses sont raisonnablement bien vérifiées en pratique, sauf parfois l’hypothèse  $H_3$ . Cette dernière, qui suppose que la mesure locale choisie permet de discriminer l’objet d’intérêt du reste de l’image, peut provoquer des oubliers lorsqu’elle n’est pas vérifiée. Si l’utilisation conjointe de plusieurs types de mesures permet de limiter ses oubliers, il y a tout un travail à faire sur le choix ou la détermination de «bonnes» mesures locales, c’est à dire de mesures discriminantes pour tel ou tel objet par rapport à son environnement.

Notre méthode est facile à mettre en œuvre et rapide. Contrairement à la plupart des méthodes d’indexation existantes, elle ne nécessite aucun apprentissage sur les objets recherchés. Les temps de calcul, sur un PC-800MHz, pour cette pré-détection, sont rapides. En considérant des images de dimension  $768 \times 576$  pixels, 15 cartes de pré-détection de forme par minute sont calculées. Cela comprend le calcul des mesures, de leur histogramme classique, et de la rétrécissement. Cette performance tombe à 3 cartes de pré-détection par minute lorsque l’on utilise les histogrammes interpolés.

Les mesures de couleur et de forme ont été utilisées dans nos expériences. Elles permettent de détecter la plupart des objets d’intérêt dans une séquence de scènes routières. Cependant, un objet peut

être détecté dans plusieurs images successives sans que l'on sache s'il s'agit du même. Lorsque qu'un évènement est détecté à un instant  $t$  (objet apparaissant, changeant, disparaissant), on le recherche par rétroposition dans l'image  $I(t)$ . La figure 4.18 présente le résultat obtenu lorsque  $\mathcal{P}(\mathbf{m}, 75|\mathcal{B})$  (calculée sur des mesures couleurs avec  $\Delta_+$ ) est rétroposé dans les images précédentes  $I_{70}$ ,  $I_{65}$  et  $I_{60}$ . La rétroposition de  $\mathcal{P}(\mathbf{m}, 75|\mathcal{B})$  permet de suivre la trajectoire de la partie colorée du panneau sur plusieurs images successives. Cette dernière remarque laisse envisager des perspectives quant au suivi d'objets dans les séquences d'images.

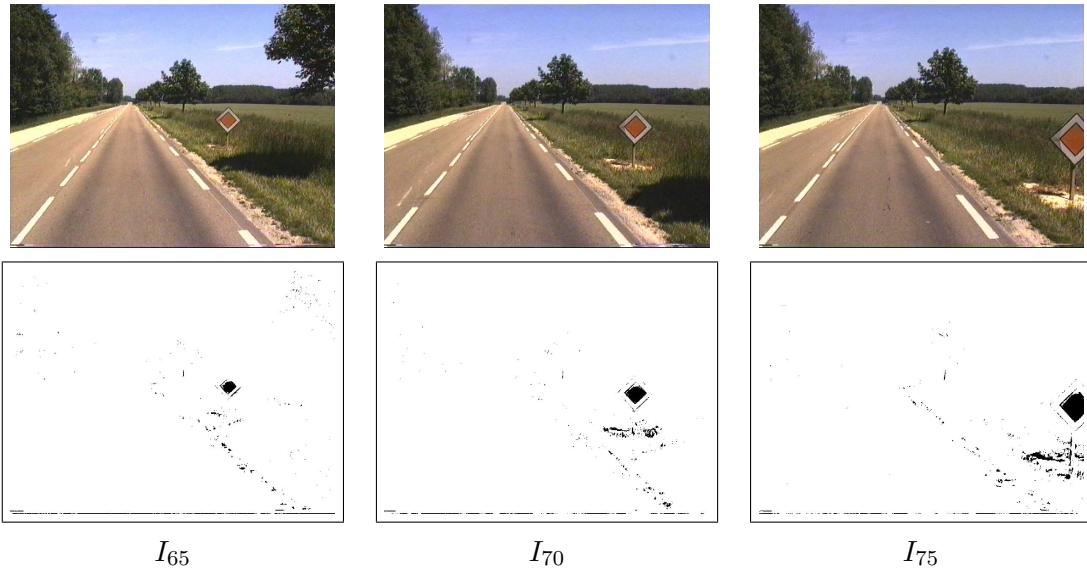


FIG. 4.18:*Suivi : rétroposition de la distribution  $\mathcal{P}(\mathbf{m}, 75|\mathcal{B})$  dans les images précédentes.*

Plusieurs travaux proposent d'apprendre des statistiques propres aux objets,  $\mathcal{P}(\mathbf{m}|\mathcal{B})$ , pour les détecter par rétroposition et les reconnaître [133, 131, 142]. Un apprentissage sur leur comportement statistique,  $\mathcal{P}(\mathbf{m}, t|\mathcal{B})$ , pourrait aussi être envisagé pour améliorer nos résultats de détection.



**Deuxième partie**

**Apparence Globale**



## Introduction

Cette seconde partie s'intéresse à la représentation globale de l'apparence. Les objets d'intérêt sont caractérisés par un ensemble d'images les représentant dans leur globalité, noté  $\mathcal{B} = \{\mathbf{x}^1, \dots, \mathbf{x}^K\}$ . Ces images  $\mathbf{x}$  sont de grande dimension. Plusieurs méthodes de réduction d'information ont été proposées dans l'état de l'art au paragraphe 2.1.3.1, qui permettent d'indexer chaque image efficacement par des coordonnées réduites  $\mathbf{c}$  liées à  $\mathbf{x}$  par une relation  $g$  linéaire ou non, à une erreur de reconstruction près  $\mathbf{w}^r$  :

$$\mathbf{x} = g(\mathbf{c}) + \mathbf{w}^r$$

Les coordonnées  $\mathbf{c}$  permettent de représenter chaque image de la base d'apprentissage, de manière parcimonieuse et, la classe  $\mathcal{B}$  peut être représentée par la distribution des images :  $\mathcal{P}(\mathbf{x}|\mathcal{B})$ . Nous avons choisi de lier  $\mathbf{x}$  à  $\mathbf{c}$  par une relation linéaire calculée par Analyse en Composantes Principales qui a l'avantage d'être simple à mettre en œuvre.

Au chapitre 5, nous rappelons comment calculer l'application linéaire  $g$  par ACP et nous proposons quelques améliorations. A partir de cette représentation, on caractérise la base d'apprentissage  $\mathcal{B}$  par la distribution des images  $\mathcal{P}(\mathbf{x}|\mathcal{B})$ .

La reconnaissance d'une image, ou observation  $\mathbf{y}$ , est réalisée en calculant les composantes réduites  $\mathbf{c}$ , puis en les comparant à celles qui indexent chaque image d'apprentissage de  $\mathcal{B}$ . Cependant, contrairement aux mesures locales, l'observation  $\mathbf{y}$  est sensible à l'apparition de perturbations, comme, par exemple, les occultations partielles. Au chapitre 6, pour prendre en compte l'apparition de telles perturbations, nous proposons de lier l'observation  $\mathbf{y}$  aux images d'apprentissage  $\mathbf{x}$  par la relation :

$$\mathbf{y} = \mathbf{x} + \mathbf{w}^o = g(\mathbf{c}) + \mathbf{w}$$

où  $\mathbf{w}^o$  correspond aux perturbations liées au bruit de capteur, occultations partielles ou dégradations. L'erreur  $\mathbf{w}$  est l'addition entre l'erreur de reconstruction et ces perturbations. En utilisant une hypothèse adaptée sur la distribution de l'erreur  $\mathcal{P}(\mathbf{w}|\mathcal{B})$ , on propose au chapitre 6, des méthodes robustes d'estimation des coordonnées  $\mathbf{c}$  de l'observation  $\mathbf{y}$ , qui permettent de reconnaître des occurrences fortement dégradées de nos images d'apprentissage. Les énergies ou vraisemblances robustes considérées dans cette estimation, sont suffisamment robustes pour qu'on puisse étendre leur usage au problème de la détection d'objets sur des fonds texturés au chapitre 7.



## Chapitre 5

# Mesures globales d'apparence

La caractéristique globale de l'apparence d'une image correspond à un vecteur contenant en ordre lexicographique, les valeurs des pixels de l'image. On dispose d'une base  $\mathcal{B}$ , contenant  $K$  images multi-bandes d'apprentissage représentées par leur vecteur  $\mathbf{x}^k \in \mathcal{B}$  défini par :

$$\mathbf{x}^k = \begin{pmatrix} x_{1,1}^k \\ \vdots \\ x_{1,N}^k \\ \vdots \\ x_{p,n}^k \\ \vdots \\ x_{P,1}^k \\ \vdots \\ x_{P,N}^k \end{pmatrix}$$

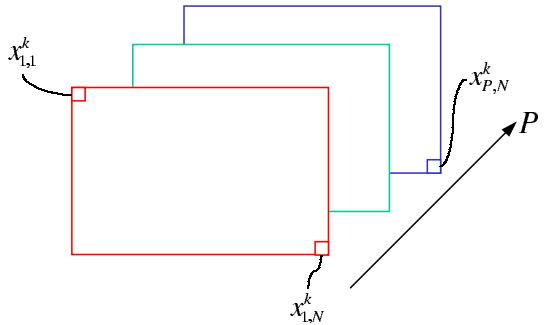


FIG. 5.1:Image multi-bande  $\mathbf{x}^k$ .

où  $P$  est le nombre de bandes et  $N$  le nombre de pixels. Ce chapitre a pour but de modéliser la vraisemblance  $\mathcal{P}(\mathbf{x}|\mathcal{B})$  à l'aide des échantillons de la base  $\mathcal{B}$ , ainsi que d'indexer chaque image d'apprentissage par une signature de plus petite dimension (très inférieure à  $NP$ ).

L'Analyse en Composantes Principales (ACP) est une technique usuelle de réduction de l'information. Elle permet de définir une relation linéaire entre la variable image  $\mathbf{x}$  et celle réduite  $\mathbf{c}$ , à une faible erreur de reconstruction près  $\mathbf{w}^r$  (cf. section 2.1.3.1).

L'ACP nécessite l'estimation sur l'ensemble d'échantillons d'apprentissage, des moments du premier ordre, la moyenne, et du second ordre, la matrice d'autocovariance. Le paragraphe 5.1.1 rappelle les estimateurs classiquement utilisés. Cependant, certains pixels des échantillons  $\mathbf{x}^k$  portent une information inutile concernant l'objet considéré. C'est pourquoi on propose de nouveaux estimateurs robustes de la moyenne et de la matrice de covariance, à la section 5.1.2, qui les écartent des estimations. La sec-

tion 5.1.3 compare les composantes principales, valeurs et vecteurs propres, de la matrice de covariance, extraites suivant ces deux approches.

Si l'on note  $E$ , l'espace des images, de dimension  $NP$ , les  $J$  premiers vecteurs propres, correspondant aux plus grandes valeurs propres, permettent de définir un sous espace  $F$  sur lequel chaque image d'apprentissage peut se projeter pour une erreur de reconstruction, notée  $\mathbf{w}^r$ , de norme euclidienne minimale. Cette projection, ou reconstruction dans  $F$ , d'une image  $\mathbf{x}$ , s'exprime en fonction de  $J$  coordonnées réduites notées  $\{c_1, \dots, c_J\}$ . Dans la section 5.2, sous certaines hypothèses et en utilisant la relation entre  $\mathbf{x}$  et les coordonnées réduites  $\mathbf{c}$ , nous présentons quelques modèles de distribution de la vraisemblance  $\mathcal{P}(\mathbf{x}|\mathcal{B})$ .

Pour illustrer ce chapitre, cette technique est appliquée à la base  $\mathcal{B}$  contenant les 72 images du premier objet de la base COIL [106] dont certaines sont représentées à la figure 5.2.

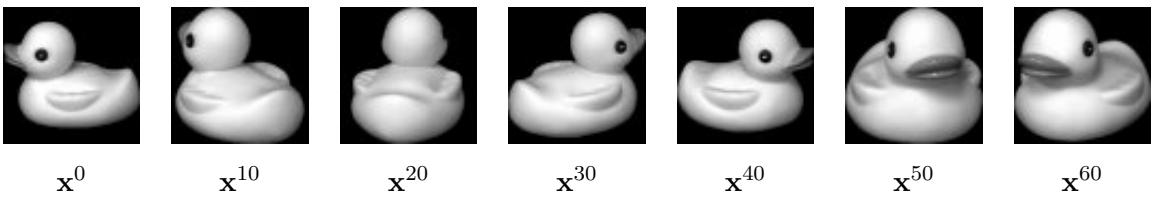


FIG. 5.2: Exemple d'images d'apprentissage de  $\mathcal{B}$  (issues de la base COIL [106]).

## 5.1 Analyse en Composantes Principales

Le paragraphe 5.1.1 rappelle les estimateurs de la moyenne et de la matrice de covariance, classiquement utilisés dans le cadre de l'ACP. A la section 5.1.2, on propose de nouveaux estimateurs. Une fois estimée, la matrice de covariance est diagonalisée ce qui nécessite le calcul de ses valeurs et vecteurs propres. On présente à la section 5.1.3, les composantes principales extraites de la matrice de covariance estimée par la méthode classique, ainsi que par les nouveaux estimateurs.

### 5.1.1 Estimation classique de la moyenne et de la matrice de covariance

#### 5.1.1.1 Estimateur classique de la moyenne : $\hat{\mu}^c$

L'image moyenne  $\hat{\mu}^c$  est classiquement définie comme l'isobarycentre des  $K$  images de  $\mathcal{B}$  soit :

$$\hat{\mu}^c = \frac{1}{K} \sum_{k=1}^K \mathbf{x}^k$$

Le  $n^{\text{ème}}$  pixel de l'image moyenne  $\hat{\mu}^c$  correspond à la moyenne des  $n^{\text{èmes}}$  pixels des images d'apprentissage.

**Calcul des données centrées associées l'estimateur  $\hat{\mu}^c$ .** Chaque image d'apprentissage  $\mathbf{x}_k$  est centrée par :

$$\tilde{\mathbf{x}}^k = \mathbf{x}^k - \hat{\mu}^c$$

La figure 5.4 présente l'image moyenne  $\hat{\mu}^c$  calculée sur la base  $\mathcal{B}$ .

### 5.1.1.2 Estimateur classique de la matrice d'autocovariance : $\hat{\Sigma}^c$

La matrice d'autocovariance est classiquement calculée par :

$$\hat{\Sigma}^c = \frac{1}{K} \sum_{k=1}^K \tilde{\mathbf{x}}^k \tilde{\mathbf{x}}^{k T} = \frac{1}{K} \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T$$

où  $\tilde{\mathbf{X}}$  est la matrice dont chaque colonne  $k$  correspond à l'image centrée  $\tilde{\mathbf{x}}^k$ . La matrice de covariance  $\hat{\Sigma}^c$  est carrée, réelle symétrique et semi-définie positive [68]. Elle est donc diagonalisable et ses valeurs propres sont positives.

La matrice  $\hat{\Sigma}^c$  est de très grande dimension :  $NP \times NP$ . Le nombre  $K$  d'images d'apprentissage dans  $\mathcal{B}$ , est souvent inférieur au nombre  $NP$ , et l'espace vectoriel qu'elles engendrent est de dimension inférieure ou égale à  $K - 1$  (la soustraction de l'image moyenne enlève une dimension). Aussi, pour réduire le temps de calcul nécessaire à l'extraction des valeurs et vecteurs propres de  $\hat{\Sigma}^c$ , on utilise la décomposition en valeurs singulières qui diagonalise la matrice  $\mathbf{Q} = \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}$ , souvent de dimension plus petite  $K \times K$  [120]. Les  $NP$  valeurs propres de  $\hat{\Sigma}^c$  sont les  $K$  valeurs propres de  $\mathbf{Q}$  complétées par des valeurs propres nulles. Si l'on note  $\mathbf{v}_j$  un vecteur propre de  $\mathbf{Q}$  associé à la valeur propre  $\lambda_j$ , alors le vecteur propre  $\mathbf{u}_j$  de  $\hat{\Sigma}^c$  se calcule simplement par :

$$\forall j \in \{1, \dots, K\} : \mathbf{u}_j = \frac{\tilde{\mathbf{X}} \mathbf{v}_j}{\sqrt{\lambda_j}}$$

Les autres  $NP - K$  vecteurs propres sont associés à une valeur propre nulle, et ne sont généralement pas calculés.

## 5.1.2 Nouveaux estimateurs de la moyenne et de la matrice de covariance

L'image d'apprentissage d'un objet représente l'une de ses réalisations sur un fond souvent choisi de couleur uniforme (c'est en particulier le cas des panneaux ou d'autres objets normalisés). Ainsi, une image d'apprentissage est composée d'un ensemble de pixels appartenant à l'objet  $O$  et d'un ensemble de pixels appartenant au fond  $\bar{O}$ . L'estimation classique de la moyenne et de la matrice de covariance ne permet pas de faire une distinction entre ces deux classes de données. Aussi, les pixels de fond ont-ils une influence sur les estimées et donc sur leurs composantes principales. Dans cette section, on propose un estimateur de la moyenne et de la matrice de covariance qui écartent les pixels de  $\bar{O}$ .

### 5.1.2.1 Nouvel estimateur de la moyenne : $\hat{\mu}^n$

Nous proposons un nouvel estimateur de l'image moyenne de  $\mathcal{B}$  où la moyenne sur chaque pixel  $n$  des images d'apprentissage est calculée en ne considérant que les pixels appartenant à l'objet :

$$\forall n \in [1, \dots, N] : \hat{\mu}_n^n = \frac{1}{\sum_{k=1}^K b_n^k} \sum_{k=1}^K b_n^k x_n^k$$

avec  $b_n^k$  la pondération appliquée aux valeurs  $x_n^k$  du pixel  $n$  de l'image  $\mathbf{x}^k$ , définie par :

$$\begin{cases} b_n^k = 1 \text{ si } x_n^k \in O \\ b_n^k = 0 \text{ sinon} \end{cases}$$

La figure 5.3 présente les pondérations appliquées à l'image  $\mathbf{x}^0$  : les pixels noirs indiquent des pondérations nulles et les pixels blancs des pondérations égales à 1.



FIG. 5.3: Image des pondérations  $\mathbf{b}^0$  appliquée à l'image d'apprentissage  $\mathbf{x}^0$ .

**Calcul des données centrées associées à l'estimateur  $\hat{\mu}^n$ .** Chaque pixel  $n$  de chaque modèle  $\mathbf{x}^k$  est centré par  $\tilde{x}_n^k = b_n^k (x_n^k - \hat{\mu}_n^n)$ , ou :

$$\tilde{\mathbf{x}}^k = \mathbf{B}^k (\mathbf{x}^k - \hat{\mu}^n)$$

en posant  $\mathbf{B}^k$  la matrice diagonale contenant  $P$  fois les  $N$  pondérations  $b_n^k$  (la pondération  $b_n^k$  est appliquée aux  $P$  valeurs du pixel  $n$ ). Les pixels de  $\mathbf{x}^k$  appartenant à l'objet  $O$  sont centrés et les pixels de  $\bar{O}$  sont mis à zéro. La figure 5.4 présente l'image moyenne  $\hat{\mu}^n$  calculée sur la base  $\mathcal{B}$ . Les pixels appartenant au fond, de valeur 0, influencent le calcul classique de la moyenne. On remarque que l'image centrée  $\tilde{\mathbf{x}}^0$  sur  $\hat{\mu}^n$ , présente des variations sur le fond ne concernant pas l'objet, ce qui n'est pas le cas en centrant avec  $\hat{\mu}^n$  à l'aide des pondérations.

### 5.1.2.2 Nouvel estimateur de la matrice d'autocovariance : $\hat{\Sigma}^n$

On propose d'estimer de manière indépendante chaque élément de la matrice de covariance, en ne considérant que les pixels appartenant à l'objet  $O$  :

$$\hat{\Sigma}_{(p_1, n_1), (p_2, n_2)}^n = \hat{\Sigma}_{p_1 * N + n_1, p_2 * N + n_2}^n = \frac{1}{\sum_{k=1}^K b_{n_1}^k b_{n_2}^k} \sum_{k=1}^K b_{n_1}^k \tilde{x}_{p_1, n_1}^k b_{n_2}^k \tilde{x}_{p_2, n_2}^k$$

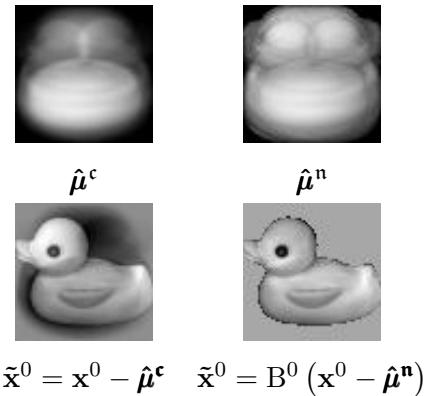


FIG. 5.4: Estimées des moyennes  $\hat{\mu}^c$  et  $\hat{\mu}^n$ , de la base  $\mathcal{B}$ , et image centrée  $\tilde{x}^0$ .

$\hat{\Sigma}^n$  correspond à l'estimateur non biaisé de la matrice de covariance [152]. Elle est carrée, réelle, symétrique et diagonalisable. La définition de  $\hat{\Sigma}^n$  ne permet pas de l'exprimer comme le produit de deux matrices. On ne peut utiliser la décomposition en valeurs singulières. De plus, le caractère semi-définie positive de la matrice d'autocovariance classique n'est pas conservé par  $\hat{\Sigma}^n$  [152]. Les valeurs et vecteurs propres sont alors extraits directement de la matrice de covariance  $\hat{\Sigma}^n$ , de dimension  $NP \times NP$ .

### 5.1.2.3 Remarque

De La Torre et Black ont proposé récemment une méthode d'estimation robuste non supervisée de l'ACP [74], qui permet d'écartier les images ou les pixels faux parmi les échantillons d'apprentissage. Cependant leur approche est coûteuse en temps de calcul.

### 5.1.3 Extraction des Composantes Principales

Un vecteur propre  $\mathbf{u}$  de la matrice  $\Sigma$  vérifie la relation suivante :

$$\Sigma \mathbf{u} = \lambda \mathbf{u}$$

où  $\lambda$  est sa valeur propre associée. La matrice  $\Sigma$  peut alors se décomposer :

$$\Sigma = \mathbf{U}_{NP} \Lambda_{NP} \mathbf{U}_{NP}^T$$

où  $\mathbf{U}_{NP}$  est une matrice collectant tous les  $NP$  vecteurs propres de  $\Sigma$ , et  $\Lambda_{NP}$  est la matrice diagonale des valeurs propres.

Aux paragraphes 5.1.3.1 et 5.1.3.2, on présente les valeurs propres et vecteurs propres de la matrice de covariance, obtenus suivant les différents estimateurs. On considère les composantes principales issues de :

- la diagonalisation de la matrice de covariance  $\hat{\Sigma}^c$  calculée à partir des images de  $\mathcal{B}$  centrées sur  $\hat{\mu}^c : (\hat{\mu}^c, \hat{\Sigma}^c)$ ,
- la diagonalisation de la matrice de covariance  $\hat{\Sigma}^n$  calculée à partir des images de  $\mathcal{B}$  centrées sur  $\hat{\mu}^n : (\hat{\mu}^n, \hat{\Sigma}^n)$ ,
- la diagonalisation de la matrice de covariance  $\hat{\Sigma}^n$  calculée à partir des images de  $\mathcal{B}$  centrées sur  $\hat{\mu}^n : (\hat{\mu}^n, \hat{\Sigma}^c)$ ,

### 5.1.3.1 Valeurs Propres

Le spectre de la matrice de covariance contient  $K - 1$  valeurs propres éventuellement non nulles, et  $NP - K + 1$  valeurs propres nulles. Elles sont rangées en ordre décroissant et notées :  $\{\lambda_j\}_{j \in \{1, \dots, NP\}}$ . On ne représente dans ce paragraphe que les  $K$  premières valeurs propres.

**Valeurs propres de  $\hat{\Sigma}^c$ .** La valeur propre  $\lambda_j$  associée au vecteur propre  $\mathbf{u}_j$  s'interprète comme la variance des coordonnées sur  $\mathbf{u}_j$  des images d'apprentissage centrées, soit :

$$\lambda_j = \frac{1}{K} \sum_{k=1}^K (c_j^k)^2$$

avec  $c_j^k = \mathbf{u}_j^T \tilde{\mathbf{x}}^k = \langle \mathbf{u}_j | \tilde{\mathbf{x}}^k \rangle$ .

La figure 5.5 présente les spectres des matrices de covariance  $\hat{\Sigma}^c$  calculés à partir des images d'apprentissage centrées sur  $\hat{\mu}^c$  et  $\hat{\mu}^n$ . Le rang  $j$  des valeurs propres est reporté en abscisse et leur valeur  $\lambda_j$ , en ordonnée. On remarque que les premières valeurs propres sont beaucoup plus faibles dans le cas  $(\hat{\mu}^n, \hat{\Sigma}^c)$ , que  $(\hat{\mu}^c, \hat{\Sigma}^c)$ . Ainsi, les variances des coordonnées des images centrées d'apprentissage sur les premiers vecteurs propres, sont plus faibles pour l'apprentissage  $(\hat{\mu}^n, \hat{\Sigma}^c)$  que  $(\hat{\mu}^c, \hat{\Sigma}^c)$ .

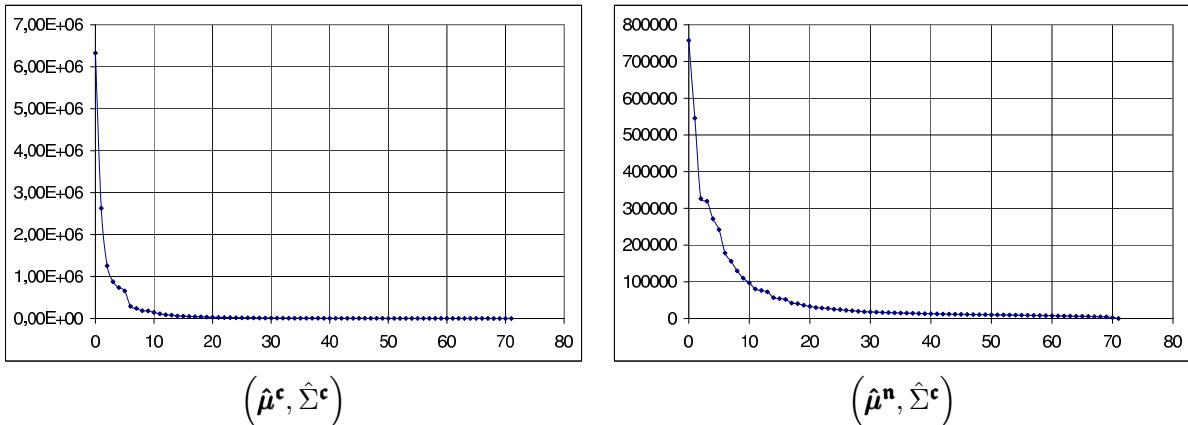


FIG. 5.5:Spectres de  $(\hat{\mu}^c, \hat{\Sigma}^c)$  et  $(\hat{\mu}^n, \hat{\Sigma}^c)$ .

**Valeurs propres de  $\hat{\Sigma}^n$ .** La figure 5.6 présente le spectre de  $\hat{\Sigma}^n$  calculée sur la base  $\mathcal{B}$ . On remarque l'apparition de valeurs négatives (liées à la perte du caractère semi-défini positif de la matrice). Les valeurs propres de  $\hat{\Sigma}^n$  ne peuvent plus s'interpréter comme des variances.

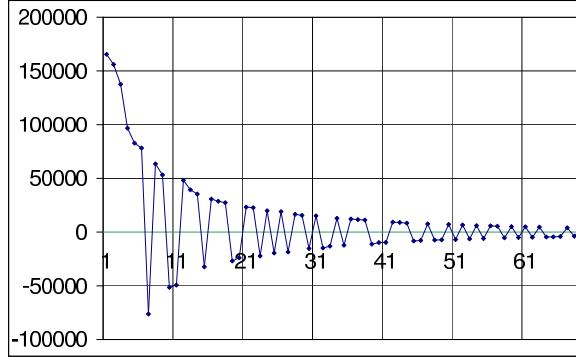


FIG. 5.6: Spectre de  $(\hat{\mu}^n, \hat{\Sigma}^n)$ .

**Part d'inertie expliquée.** L'information contenue dans les  $J$  premiers vecteurs propres ou la qualité globale de la représentation est évaluée par la *part d'inertie expliquée* :

$$\mathcal{Q}_J = \frac{\sum_{j=1}^J \lambda_j}{\sum_{j=1}^{NP} \lambda_j} \quad (5.1)$$

Le nombre  $J$  de vecteurs propres retenus est choisi tel que la part d'inertie expliquée soit supérieure à un seuil fixé *a priori* [3]. La figure 5.7 présente la part d'inertie expliquée calculée pour  $(\hat{\mu}^c, \hat{\Sigma}^c)$  et  $(\hat{\mu}^n, \hat{\Sigma}^c)$ . L'abscisse correspond au nombre  $J$  de vecteurs propres retenus pour décrire l'espace propre  $F$ , et la part d'inertie expliquée  $\mathcal{Q}_J$  est reportée en ordonnée.

La part d'inertie expliquée de  $(\hat{\mu}^c, \hat{\Sigma}^c)$  a une croissance beaucoup plus rapide que  $(\hat{\mu}^n, \hat{\Sigma}^c)$ . Cela est dû aux premières valeurs propres qui sont beaucoup plus grandes pour  $(\hat{\mu}^c, \hat{\Sigma}^c)$  (cf. figure 5.5).

La part d'inertie expliquée n'a pas été calculée pour la matrice de covariance  $\hat{\Sigma}^n$  car la présence de valeurs propres négatives rend difficile son interprétation.

### 5.1.3.2 Vecteurs Propres

Cette section présente les vecteurs propres issus des deux estimateurs possibles de la matrice de covariance. Les images d'apprentissage peuvent être reconstruites efficacement suivant les  $J$  premiers vecteurs propres pour une erreur de reconstruction minimale telle que :

$$\tilde{\mathbf{x}} = \sum_{j=1}^J c_j \mathbf{u}_j + \mathbf{w}^r = \tilde{\mathbf{x}}^F + \mathbf{w}^r$$

où

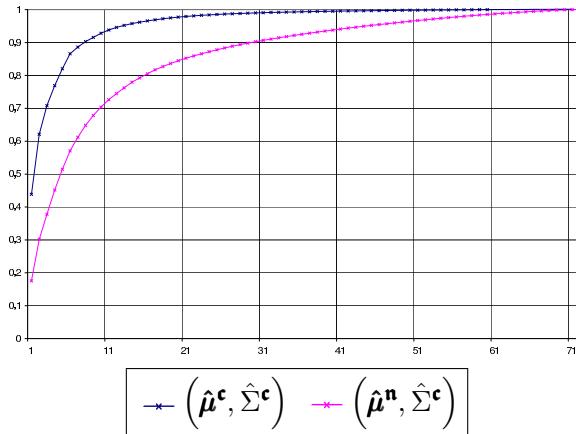


FIG. 5.7: Part d'inertie expliquée  $Q_J$  en fonction du nombre  $J$  de vecteurs propres retenus, pour les apprentissages  $(\hat{\mu}^c, \hat{\Sigma}^c)$  et  $(\hat{\mu}^n, \hat{\Sigma}^c)$ .

- $\tilde{\mathbf{x}}^F$  est la composante ou projection orthogonale, de  $\tilde{\mathbf{x}}$  sur l'espace propre  $F$ ,
- $\mathbf{w}^r$  est l'erreur de reconstruction de  $\tilde{\mathbf{x}}$  sur  $F^\perp$ ,
- $c_j$  est la coordonnée de  $\tilde{\mathbf{x}}$  sur le vecteur propre  $\mathbf{u}_j$  associé à la valeur propre  $\lambda_j$ .

On rappelle que, lorsque l'on considère la moyenne  $\hat{\mu}^n$ , la relation entre l'image  $\mathbf{x}^k$  et son image centrée  $\tilde{\mathbf{x}}^k$ , dépend d'une matrice de pondération  $\mathbf{B}^k : \tilde{\mathbf{x}}^k = \mathbf{B}^k (\mathbf{x}^k - \hat{\mu}^n)$ . Dans le cas de la moyenne classique  $\hat{\mu}^c$ ,  $\tilde{\mathbf{x}}^k$  est lié à  $\mathbf{x}^k$  par :  $\tilde{\mathbf{x}}^k = \mathbf{x}^k - \hat{\mu}^c$ . Cependant, la relation entre l'image centrée et ses coordonnées  $\mathbf{c}$  sur les vecteurs propres est, dans les deux cas, :

$$\mathbf{c}^k = \mathbf{U}^T \tilde{\mathbf{x}}^k$$

où  $\mathbf{U}$  est la matrice  $NP \times J$  contenant les vecteurs propres  $\{\mathbf{u}_j\}_{j \in \{1, \dots, J\}}$ .

**Vecteurs propres de  $(\hat{\mu}^c, \hat{\Sigma}^c)$ .** La figure 5.8 présente les cinq premiers vecteurs propres,  $\{\mathbf{u}_j\}_{j=\{1, \dots, 5\}}$ , de la matrice de covariance  $\hat{\Sigma}^c$  calculée en centrant les images d'apprentissage sur  $\hat{\mu}^c$ .

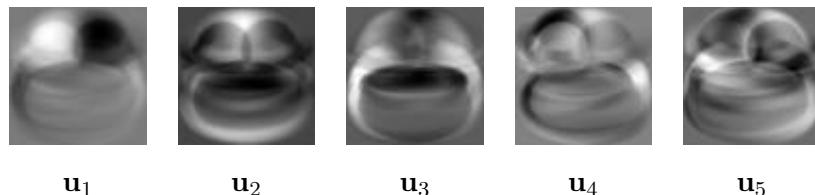


FIG. 5.8: Cinq premiers vecteurs propres issus des estimateurs  $(\hat{\mu}^c, \hat{\Sigma}^c)$ .

La figure 5.9 présente la reconstruction sur ces cinq premiers vecteurs propres, obtenue pour l'image d'apprentissage  $\mathbf{x}^0$ . On rappelle que dans ce cas,  $\tilde{\mathbf{x}}^0$  est obtenue en soustrayant  $\hat{\mu}^c$  (cf. paragraphe

5.1.1.1). On remarque que l'œil du canard n'est pas reconstruit par les cinq premiers vecteurs propres et par conséquent, apparaît dans l'erreur de reconstruction.

$$\tilde{x}^0 = \tilde{x}^{0,F} + w^{0,r}$$

FIG. 5.9: Reconstruction de  $\tilde{x}^0$  sur les cinq premiers vecteurs propres issus de  $(\hat{\mu}^c, \hat{\Sigma}^c)$ .

**Vecteurs propres de  $(\hat{\mu}^n, \hat{\Sigma}^c)$ .** La figure 5.10 présente les cinq premiers vecteurs propres de la matrice de covariance  $\hat{\Sigma}^c$  calculée en centrant les images d'apprentissage sur  $\hat{\mu}^n$ . On peut remarquer la différence entre ceux obtenus par  $(\hat{\mu}^c, \hat{\Sigma}^c)$  qui codent des variations liées au fond noir des images d'apprentissage (cf. figure 5.8).

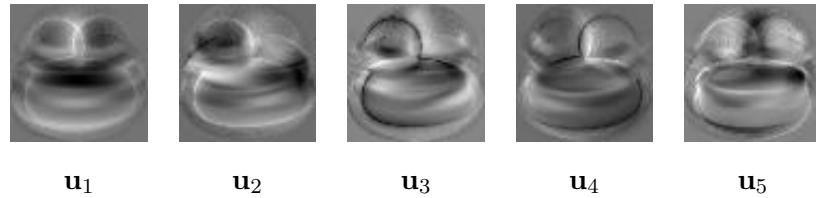


FIG. 5.10: Cinq premiers vecteurs propres issus des estimateurs  $(\hat{\mu}^n, \hat{\Sigma}^c)$ .

La figure 5.11 présente la reconstruction obtenue sur ces cinq vecteurs propres.  $x^0$  est centré sur  $\hat{\mu}^n$  par la méthode présentée au paragraphe 5.1.2.1.

$$\tilde{x}^0 = \tilde{x}^{0,F} + w^{0,r}$$

FIG. 5.11: Reconstruction de  $\tilde{x}^0$  sur les cinq premiers vecteurs propres issus de  $(\hat{\mu}^n, \hat{\Sigma}^c)$ .

**Vecteurs propres de  $(\hat{\mu}^n, \hat{\Sigma}^n)$ .** La figure 5.12 présente les cinq premiers vecteurs propres de  $\hat{\Sigma}^n$  calculée en centrant les images d'apprentissage sur  $\hat{\mu}^n$ .

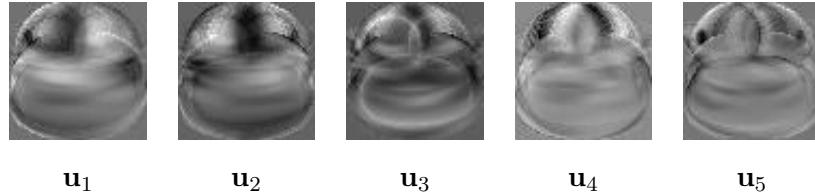


FIG. 5.12: Cinq premiers vecteurs propres issus des estimateurs  $(\hat{\mu}^n, \hat{\Sigma}^n)$ .

La figure 5.13 présente la reconstruction  $\tilde{x}^{0,F}$ , obtenue sur les cinq vecteurs propres extraits de  $(\hat{\mu}^n, \hat{\Sigma}^n)$ . Cette reconstruction est, visuellement, moins proche de l'image centrée  $\tilde{x}^0$ , que celle obtenue par l'apprentissage  $(\hat{\mu}^c, \hat{\Sigma}^c)$ . Cette impression est confirmée lorsque l'on compare la variance des erreurs de reconstruction  $w^{0,r}$  sur chaque pixel : celle-ci est plus grande pour l'apprentissage  $(\hat{\mu}^n, \hat{\Sigma}^n)$  que  $(\hat{\mu}^c, \hat{\Sigma}^c)$ . Ces variances de l'erreur de reconstruction sont reportées dans le tableau 5.1 pour les trois apprentissages proposés.

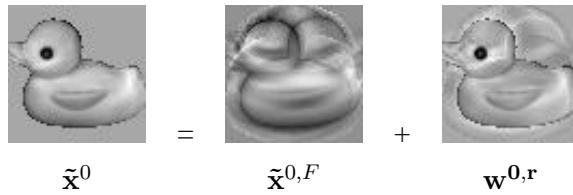


FIG. 5.13: Reconstruction de  $\tilde{x}^0$  sur les cinq premiers vecteurs propres issus de  $(\hat{\mu}^n, \hat{\Sigma}^n)$ .

	$(\hat{\mu}^c, \hat{\Sigma}^c)$	$(\hat{\mu}^n, \hat{\Sigma}^c)$	$(\hat{\mu}^n, \hat{\Sigma}^n)$
Variance	359	348	642

TAB. 5.1: Variance de l'erreur de reconstruction  $w^{0,r}$  pour les différents apprentissages, pour un espace propre  $F$  de dimension  $J = 5$ .

### 5.1.3.3 Conclusion

Dans cet exemple, on peut remarquer visuellement que les premiers vecteurs propres issus des estimateurs  $(\hat{\mu}^c, \hat{\Sigma}^c)$  portent l'information de variation entre le fond et l'objet. Ces variations sont associées à de très grandes valeurs propres en comparaison des premières valeurs propres issues des estimateurs  $(\hat{\mu}^n, \hat{\Sigma}^c)$ . Dans ce cas, les variations majeures entre les images d'apprentissage sont liées à la différence de niveaux de gris entre le fond et l'objet dans les images d'apprentissage. Cette variation dépasse celle liée au changement d'apparence de l'objet.

Les premières valeurs propres issues de  $(\hat{\mu}^n, \hat{\Sigma}^n)$  sont inférieures à celles extraites de  $(\hat{\mu}^c, \hat{\Sigma}^c)$ . Cela implique que la distribution des images centrées sur  $\hat{\mu}^n$  est plus compacte que celle des images centrées sur  $\hat{\mu}^c$ .

L'apparition de valeurs propres négatives, pour l'apprentissage  $(\hat{\mu}^n, \hat{\Sigma}^n)$ , rend impossible leur interprétation en terme de variance. De plus l'erreur de reconstruction entre l'image et son représentant dans l'espace propre est plus importante que pour les apprentissages  $(\hat{\mu}^c, \hat{\Sigma}^c)$  et  $(\hat{\mu}^n, \hat{\Sigma}^n)$ . Aussi, le nouvel estimateur de la matrice de covariance,  $\hat{\Sigma}^n$ , n'a pas été retenu pour extraire les composantes principales et n'est pas utilisé par la suite.

## 5.2 Distribution des images d'apprentissage $\mathcal{P}(\mathbf{x}|\mathcal{B})$

Les images d'apprentissage de  $\mathcal{B}$  sont  $K$  échantillons de la variable aléatoire  $\mathbf{x}$ . On cherche à modéliser la densité de probabilité  $\mathcal{P}(\mathbf{x}|\mathcal{B})$  pour caractériser la base  $\mathcal{B}$ .

On note  $\mathbf{c}_{NP} = \{c_1, \dots, c_{NP}\}$ , les coordonnées de l'image centrée  $\tilde{\mathbf{x}}$  sur les  $NP$  vecteurs propres rangés dans les colonnes de la matrice  $\mathbf{U}_{NP}$ . La variable  $\mathbf{c}_{NP}$  est liée à l'image  $\mathbf{x}$  par les relations linéaires et inversibles :

$$\begin{cases} \tilde{\mathbf{x}} = \mathbf{U}_{NP} \mathbf{c}_{NP} \\ \mathbf{c}_{NP} = \mathbf{U}_{NP}^T \tilde{\mathbf{x}} \end{cases}$$

On rappelle que le vecteur  $\mathbf{c} = \{c_1, \dots, c_J\}$  est lié à l'image centrée  $\tilde{\mathbf{x}}$  par :

$$\begin{cases} \tilde{\mathbf{x}} = \mathbf{U} \mathbf{c} + \mathbf{w}^r \\ \mathbf{c} = \mathbf{U}^T \tilde{\mathbf{x}} \end{cases}$$

$\mathbf{U}$  est la matrice de dimension  $NP \times J$  collectant les  $J$  vecteurs propres principaux.

Ces définitions sont illustrées par la figure 5.14, pour  $NP = 3$  et  $J = 2$ .

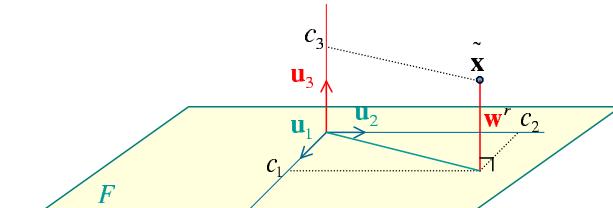


FIG. 5.14: Représentation dans un espace de dimension  $NP = 3$  :  $\mathbf{c}_{NP} = \{c_1, c_2, c_3\}$ ,  $\mathbf{c} = \{c_1, c_2\}$  et l'espace propre  $F$  est défini par les vecteurs  $(\mathbf{u}_1, \mathbf{u}_2)$ .

**Approche de Moghaddam et Pentland.** Moghaddam et Pentland supposent que les images d'apprentissage de  $\mathcal{B}$  ont une distribution gaussienne dans l'espace des images  $E$  de dimension  $NP$  [97]. Dans ce cas, les paramètres de la distribution  $\mathcal{P}(\mathbf{x}|\mathcal{B})$  sont entièrement déterminés par l'apprentissage ACP :

$$\mathcal{P}(\mathbf{x}|\mathcal{B}) = \frac{\exp \left[ -\frac{1}{2} \tilde{\mathbf{x}}^T \hat{\Sigma}^{-1} \tilde{\mathbf{x}} \right]}{(2\pi)^{\frac{NP}{2}} |\det(\hat{\Sigma})|^{\frac{1}{2}}}$$

En changeant de variable  $\mathbf{c}_{NP} = \mathbf{U}_{NP}^T \tilde{\mathbf{x}}$ , on peut écrire :

$$\mathcal{P}(\mathbf{x}|\mathcal{B}) = \frac{\exp\left[-\frac{1}{2} \mathbf{c}_{NP}^T \Lambda_{NP}^{-1} \mathbf{c}_{NP}\right]}{(2\pi)^{\frac{NP}{2}} \prod_{j=1}^{NP} \sqrt{\lambda_j}} = \frac{\exp\left[-\sum_{j=1}^{NP} \frac{c_j^2}{2\lambda_j}\right]}{(2\pi)^{\frac{NP}{2}} \prod_{j=1}^{NP} \sqrt{\lambda_j}}$$

Moghaddam et al. approchent la distribution dans  $F^\perp$  par une distribution gaussienne indépendante identiquement distribuée [97], alors :

$$\sum_{j=J+1}^{NP} \frac{c_j^2}{\lambda_j} \simeq \frac{\sum_{j=J+1}^{NP} c_j^2}{\sigma^2} = \frac{\|\mathbf{w}^r\|^2}{\sigma^2} \quad (5.2)$$

et la distribution des images d'apprentissage s'écrit :

$$\mathcal{P}(\mathbf{x}|\mathcal{B}) \simeq \hat{\mathcal{P}}(\mathbf{x}|\mathcal{B}) = \frac{\exp\left[-\sum_{j=1}^J \frac{c_j^2}{2\lambda_j}\right]}{\prod_{j=1}^J \sqrt{2\pi\lambda_j}} \cdot \frac{\exp\left[-\frac{\|\mathbf{w}^r\|^2}{2\sigma^2}\right]}{(\sqrt{2\pi}\sigma)^{NP-J}} \quad (5.3)$$

avec  $\mathbf{c} = \mathbf{U}^T \tilde{\mathbf{x}}$  et  $\mathbf{w}^r = \tilde{\mathbf{x}} - \mathbf{U}\mathbf{U}^T \tilde{\mathbf{x}}$ . La densité de probabilité  $\hat{\mathcal{P}}(\mathbf{x}|\mathcal{B})$  correspond au produit de deux distributions marginales, l'une dans  $F$  et l'autre dans  $F^\perp$ . L'approximation de la distribution dans l'espace propre orthogonale  $F^\perp$  par une distribution gaussienne indépendante identiquement distribuée, permet de ne pas calculer explicitement les composantes  $\{c_j\}_{j \in \{J+1, \dots, NP\}}$  de l'image centrée  $\tilde{\mathbf{x}}$  sur les vecteurs propres de  $F^\perp$  qui ne sont généralement pas connus.

La variance  $\sigma^2$  est estimée en minimisant la divergence de Kullbach Leibler entre la vraie densité de probabilité  $\mathcal{P}(\mathbf{x}|\mathcal{B})$  et son approximation  $\hat{\mathcal{P}}(\mathbf{x}|\mathcal{B})$ , définie par :

$$\mathcal{J}_{KL} = E \left[ \log \left( \frac{\mathcal{P}(\mathbf{x}|\mathcal{B})}{\hat{\mathcal{P}}(\mathbf{x}|\mathcal{B})} \right) \right]$$

La variance  $\sigma^2$  qui minimise  $\mathcal{J}_{KL}$ , correspond à la valeur moyenne des valeurs propres de  $F^\perp$  [97, 153] :

$$\sigma^2 = \frac{1}{PN - J} \sum_{j=J+1}^{PN} \lambda_j \quad (5.4)$$

**Notre approche.** L'image centrée  $\tilde{\mathbf{x}}$  est liée à ses coordonnées  $\mathbf{c}_{NP}$  par la relation linéaire inversible :  $\mathbf{c}_{NP} = \mathbf{U}_{NP}^T \tilde{\mathbf{x}}$ . Alors, la densité de probabilité  $\mathcal{P}(\mathbf{x}|\mathcal{B})$  est liée à  $\mathcal{P}(\mathbf{c}_{NP}|\mathcal{B})$  par [152] :

$$\mathcal{P}(\mathbf{x}|\mathcal{B}) = \frac{1}{|\det(\mathbf{U}_{NP})|} \mathcal{P}(\mathbf{c}_{NP}|\mathcal{B})$$

La matrice  $\mathbf{U}_{NP}$  est orthogonale aussi la valeur absolue de son déterminant  $|\det(\mathbf{U}_{NP})|$ , est égale à 1, et  $\mathcal{P}(\mathbf{x}|\mathcal{B}) = \mathcal{P}(\mathbf{c}_{NP}|\mathcal{B})$ . On suppose que les coordonnées  $\mathbf{c} = (c_1, \dots, c_J)$  dans l'espace propre  $F$  et  $\mathbf{c} = (c_{J+1}, \dots, c_{NP})$  dans l'espace orthogonal  $F^\perp$ , sont indépendantes alors :

$$\mathcal{P}(\mathbf{x}|\mathcal{B}) = \mathcal{P}(c_1, \dots, c_J|\mathcal{B}) \mathcal{P}(c_{J+1}, \dots, c_{NP}|\mathcal{B})$$

Cependant on ne connaît que les  $J$  premiers vecteurs propres, et on ne peut pas calculer explicitement les coordonnées  $(c_{J+1}, \dots, c_{NP})$ . La section 5.2.1 présente un modèle de la distribution  $\mathcal{P}(c_{J+1}, \dots, c_{NP}|\mathcal{B})$ . Le paragraphe 5.2.2 propose plusieurs densités de probabilité pour les coordonnées réduites  $\mathbf{c}$  dans le plan propre. Ces résultats sur la distribution de  $(c_{J+1}, \dots, c_{NP})$  dans  $F^\perp$ , et la distribution de  $\{c_1, \dots, c_J\}$  dans  $F$ , sont reliés à la vraisemblance *a priori* de la variable  $\mathbf{x}$ ,  $\mathcal{P}(\mathbf{x}|\mathcal{B})$ , à la section 5.2.3.

### 5.2.1 Distribution dans l'espace orthogonal $F^\perp$

**Hypothèse.** On suppose, comme Moghaddam et Pentland, que les coordonnées  $(c_{J+1}, \dots, c_{NP})$  ont une distribution gaussienne indépendante uniformément distribuée, de moyenne nulle et de variance  $\sigma^2$  :

$$\mathcal{P}(c_{J+1}, \dots, c_{NP}|\mathcal{B}) = \frac{1}{(\sqrt{2\pi}\sigma)^{NP-J}} \exp\left[-\frac{\sum_{j=J+1}^{NP} c_j^2}{2\sigma^2}\right]$$

On rappelle que les composantes  $(c_{J+1}, \dots, c_{NP})$  sont les coordonnées de l'image centrée  $\tilde{\mathbf{x}}$  dans  $F^\perp$ . Or, l'erreur de reconstruction  $\mathbf{w}^r$  est la projection orthogonale de  $\tilde{\mathbf{x}}$  sur l'espace propre  $F^\perp$ , et permet de calculer  $\sum_{j=J+1}^{NP} c_j^2$  (cf. équation (5.2)). Ainsi, on peut réécrire la densité de probabilité :

$$\mathcal{P}(c_{J+1}, \dots, c_{NP}|\mathcal{B}) = \frac{1}{(\sqrt{2\pi}\sigma)^{NP-J}} \exp\left[-\frac{\|\mathbf{w}^r\|^2}{2\sigma^2}\right] \quad (5.5)$$

avec  $\mathbf{w}^r = \tilde{\mathbf{x}} - \mathbf{U}\mathbf{U}^T\tilde{\mathbf{x}}$ .

**Estimation de la variance  $\sigma^2$ .** Plusieurs articles [97, 153], proposent d'estimer la variance  $\sigma^2$ , comme la valeur moyenne des valeurs propres de  $F^\perp$  (cf. équation (5.4)). Une difficulté de cette expression, est qu'elle nécessite le calcul de toutes les valeurs propres de la matrice de covariance, ou, dans le cas usuel où le nombre d'échantillons d'apprentissage  $K$  est inférieur à la dimension  $NP$  de l'espace des images, au calcul des  $K - 1$  valeurs propres éventuellement non-nulles. Or, dans le cas de la matrice de covariance classique  $\hat{\Sigma}^c$ , la valeur propre  $\lambda_j$  correspond à la variance des composantes  $c_j$  des modèles d'apprentissage soit :

$$\lambda_j = \frac{1}{K} \sum_{k=1}^K (c_j^k)^2$$

$c_j^k$  est la composante de l'image d'apprentissage  $\tilde{\mathbf{x}}^k$  sur le vecteur propre  $\mathbf{u}_j$  tel que  $c_j^k = \mathbf{u}_j^T \tilde{\mathbf{x}}^k$  ou,  $c_j^k = \langle \mathbf{u}_j | \tilde{\mathbf{x}}^k \rangle$ . Alors, on peut écrire :

$$\begin{aligned} \sum_{j=J+1}^{NP} \lambda_j &= \sum_{j=J+1}^{NP} \frac{1}{K} \sum_{k=1}^K (c_j^k)^2 \\ &= \frac{1}{K} \sum_{k=1}^K \sum_{j=J+1}^{NP} (c_j^k)^2 \\ &= \frac{1}{K} \sum_{k=1}^K \|\mathbf{w}^{k,r}\|^2 \end{aligned}$$

Ainsi, la variance de l'erreur,  $\sigma^2$ , peut être calculée de façon efficace en considérant les erreurs de reconstruction des images d'apprentissage :

$$\sigma^2 = \frac{1}{K(PN - J)} \sum_{k=1}^K \|\mathbf{w}^{k,r}\|^2$$

ou

$$\sigma^2 = \frac{1}{K(PN - J)} \sum_{k=1}^K \|\tilde{\mathbf{x}}^k - \mathbf{U}\mathbf{U}^T\tilde{\mathbf{x}}^k\|^2 \quad (5.6)$$

Cette remarque permet de calculer  $\sigma$  facilement lorsque les valeurs propres  $\{\lambda_j\}_{j>J}$  n'ont pas été calculées.

## 5.2.2 Distribution dans l'espace propre $F$

On propose plusieurs modèles de distribution de probabilité  $\mathcal{P}(c_1, \dots, c_J | \mathcal{B})$ . Le paragraphe 5.2.2.1 propose quelques modèles simples, distribution uniforme et gaussienne des coordonnées  $\mathbf{c}$  dans le plan propre  $F$ . Le paragraphe 5.2.2.2 présente les distributions particulières de  $\mathbf{c}$  pour les bases d'apprentissage que l'on considère dans cette thèse.

### 5.2.2.1 Hypothèses simples : distributions uniforme et gaussienne

**Distribution uniforme.** Si les projections des images d'apprentissage sont réparties de manière uniforme dans l'espace propre  $F$ , alors :

$$\mathcal{P}(c_1, \dots, c_J | \mathcal{B}) = cste$$

**Distribution gaussienne.** Si  $\{c_1, \dots, c_J\}$  a une distribution gaussienne dans  $F$  alors les valeurs propres  $\{\lambda_j\}_{j \in \{1, \dots, J\}}$  en sont les variances sur les axes propres associés  $\mathbf{u}_j$  telles que :

$$\mathcal{P}(c_1, \dots, c_J | \mathcal{B}) = \frac{1}{\prod_{j=1}^J \sqrt{2\pi\lambda_j}} \exp \left[ - \sum_{j=1}^J \frac{c_j^2}{2\lambda_j} \right] \quad (5.7)$$

### 5.2.2.2 Distributions particulières

Dans les expériences menées aux chapitres 6 et 7, on considère trois bases d'images d'apprentissage :

- la base COIL qui contient 72 images en niveau de gris, d'un même objet variant suivant l'angle de prise de vue,
- la base A1 qui contient 180 images en couleur, d'un panneau variant suivant une rotation dans le plan de l'image,

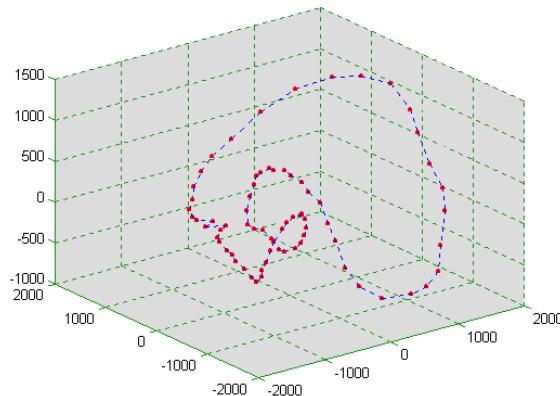
- la base A43 qui contient 1548 images en couleur, de 43 panneaux différents variant suivant leur rotation dans le plan de l'image.

On présente dans ce paragraphe, les distributions particulières des projections des modèles d'apprentissage dans le plan propre  $F$  pour ces trois bases.

---

### Base COIL.

La figure 5.15 représente la répartition de l'objet «canard» de la base COIL sur son espace propre réduit à ses trois premières composantes. Murase et Nayar remarquent que les projections des modèles d'apprentissage dans l'espace propre se répartissent sur une courbe paramétrée [103].



$$F = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$$

FIG. 5.15: Répartition des coordonnées  $(c_1, c_2, c_3)$  des images de la base COIL .

---

### Base A1.

Uenohara et Kanade montrent que les vecteurs propres issus de l'ACP appliquée à un ensemble d'images d'un objet variant suivant la rotation dans le plan de l'image ont une forme particulière [155], et sont liés deux par deux à une valeur propre double. Nous montrons dans l'annexe D que les images d'apprentissage sont réparties sur une courbe dépendant d'un seul paramètre [69]. De plus, leurs projections dans chaque plan propre caractérisé par une valeur propre double  $\lambda_j$ , sont distribuées suivant un cercle de rayon  $2\lambda_j$ .

Dans cet exemple, on considère une bibliothèque constituée de 180 images d'un même panneau, contenant une image tous les deux degrés de rotation dont quelques occurrences apparaissent à la figure 5.16.

La figure 5.17 présente les 10 premiers vecteurs propres extraits par ACP. Chaque couple de vecteur propre  $\{\mathbf{u}_{2j-1}, \mathbf{u}_{2j}\}$  est associé à une même valeur propre.

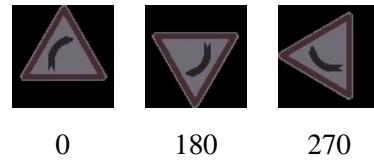


FIG. 5.16: Quelques images d'apprentissage de la base A1 avec indication de leur angle de rotation

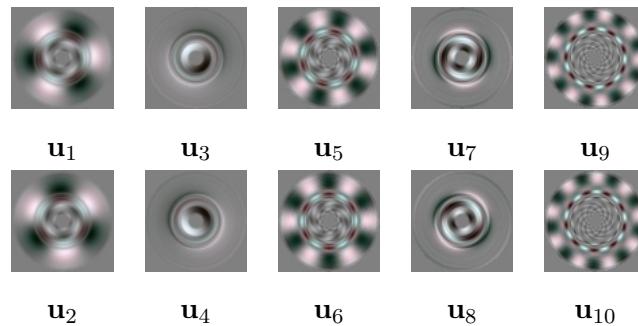


FIG. 5.17: Couples de vecteurs propres associés à une valeur propre double

La figure 5.18 présente la répartition des projections des images d'apprentissage, sur un espace propre défini par  $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$  et  $\{\mathbf{u}_1, \mathbf{u}_2\}$ . Dans le plan propre défini par une valeur propre double, par exemple  $F = \{\mathbf{u}_1, \mathbf{u}_2\}$ , les projections des images d'apprentissage se répartissent suivant un cercle.

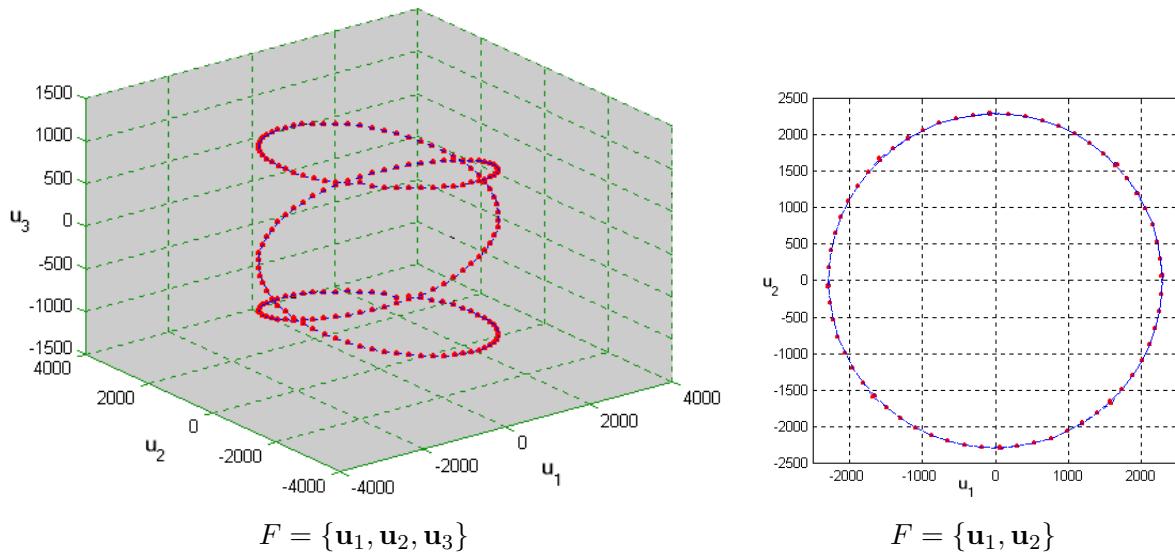


FIG. 5.18: Distribution des images d'apprentissage de la base A1 dans  $F$

Si l'on considère le plan propre  $F = \{\mathbf{u}_1, \mathbf{u}_2\}$ , les coordonnées  $c_1$  et  $c_2$  des modèles d'apprentissage sur les deux premiers vecteurs propres, appartiennent à un cercle. On peut représenter leur distribution

par :

$$\mathcal{P}(c_1, c_2 | \mathcal{B}) = \frac{1}{Z} \delta(c_1^2 + c_2^2 - \lambda_1 - \lambda_2)$$

où  $\delta$  est la distribution de Dirac et  $Z$  le facteur de normalisation. On rappelle que l'espace propre défini par  $F = \{\mathbf{u}_1, \mathbf{u}_2\}$ , a ses valeurs propres telles que  $\lambda_1 = \lambda_2$ .

#### Base A43.

---

La base d'images d'apprentissage contient 43 panneaux triangulaires appris avec leur rotation dans le plan de l'image (une image tous les 10 degrés), soit 1548 images d'apprentissage. Ces panneaux ont une couleur dominante blanche ou jaune. Quelques uns sont représentés à la figure 5.19. Les 43 panneaux différents sont présentés à l'annexe H.



FIG. 5.19: *Quelques images d'apprentissage de la base A43*

Pour limiter le temps de calcul nécessaire à la diagonalisation de la matrice de covariance, seuls les 200 premiers vecteurs et valeurs propres ont été calculés.

**Valeurs propres.** La figure 5.20 présente les valeurs propres obtenues pour  $(\hat{\mu}^c, \hat{\Sigma}^c)$  et  $(\hat{\mu}^n, \hat{\Sigma}^c)$ . Les premières valeurs propres de  $(\hat{\mu}^n, \hat{\Sigma}^c)$  sont très inférieures à celles de  $(\hat{\mu}^c, \hat{\Sigma}^c)$ .

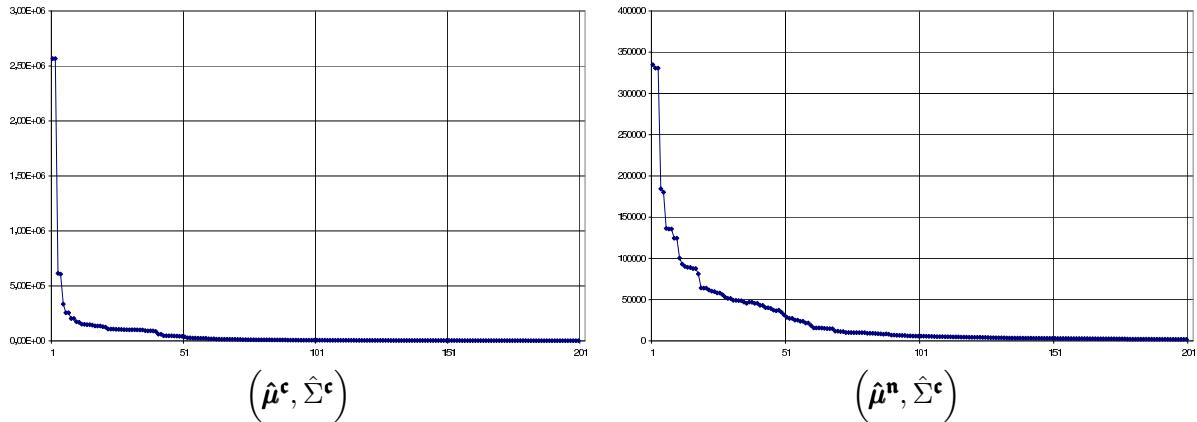


FIG. 5.20: *Valeurs propres de la base A43.*

**Vecteurs propres.** La figure 5.21 présente les cinq premiers vecteurs propres calculés pour les deux méthodes d'apprentissage. On remarque des similitudes entre  $\mathbf{u}_5$  extrait de  $(\hat{\mu}^c, \hat{\Sigma}^c)$  et  $\mathbf{u}_1$  extrait de  $(\hat{\mu}^n, \hat{\Sigma}^c)$  qui code l'information de couleur dominante (jaune ou blanche) du panneau.

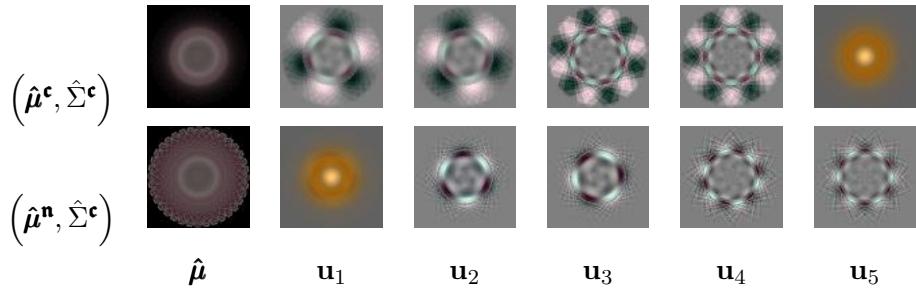


FIG. 5.21: Vecteurs propres de la base A43, suivant les deux apprentissages  $(\hat{\mu}^c, \hat{\Sigma}^c)$  et  $(\hat{\mu}^n, \hat{\Sigma}^c)$ .

**Distribution.** La figure 5.22 présente la distribution des projections des images d'apprentissage dans l'espace propre  $F$  défini par les trois premières composantes principales de  $(\hat{\mu}^n, \hat{\Sigma}^c)$ . Deux classes distinctes apparaissent dans le plan  $(u_1, u_2)$  : la plus importante correspond aux panneaux triangulaires à fond blanc, et la seconde aux panneaux triangulaires à fond jaune.

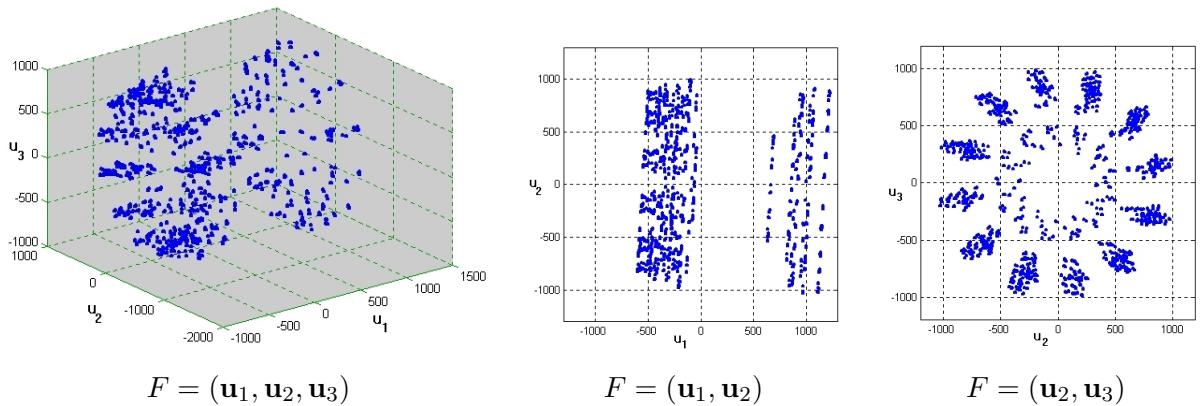


FIG. 5.22: Distribution des coordonnées  $\mathbf{c}$  des images d'apprentissage de A43 (apprentissage  $\hat{\mu}^n, \hat{\Sigma}^c$ ).

### 5.2.3 Distribution $\mathcal{P}(\mathbf{x}|\mathcal{B})$ des images d'apprentissage de $\mathcal{B}$

En supposant que les coordonnées  $\{c_j\}_{j \in \{J+1, \dots, NP\}}$  ont une distribution gaussienne indépendante uniformément distribuée de variance  $\sigma^2$  dans l'espace  $F^\perp$ , et qu'elles sont indépendantes de  $\{c_j\}_{j \in \{1, \dots, J\}}$ , alors la distribution de probabilité  $\mathcal{P}(\mathbf{x}|\mathcal{B})$  s'écrit :

$$\mathcal{P}(\mathbf{x}|\mathcal{B}) = \mathcal{P}(c_1, \dots, c_J|\mathcal{B}) \cdot \frac{\exp\left[\frac{-\|\mathbf{w}^r\|^2}{2\sigma^2}\right]}{(\sqrt{2\pi}\sigma)^{NP-J}} \quad (5.8)$$

avec

$$\left| \begin{array}{l} \mathbf{c} = \mathbf{U}^T \tilde{\mathbf{x}} \\ \mathbf{w}^r = \tilde{\mathbf{x}} - \mathbf{U} \mathbf{U}^T \tilde{\mathbf{x}} \end{array} \right.$$

Si l'on suppose que les composantes  $\mathbf{c}$  ont une distribution gaussienne dans le plan propre, alors l'équation (5.8) permet de retrouver la vraisemblance proposée par Moghaddam et Pentland [97].

## 5.3 Conclusion

L'analyse en composantes principales est une technique usuelle de réduction d'information qui nécessite l'estimation de la moyenne et de la matrice d'autocovariance sur un ensemble d'échantillons. En écartant les pixels des images d'apprentissage qui ne caractérisent pas l'objet, on a proposé un nouvel estimateur de la moyenne  $\hat{\mu}^n$  qui permet d'obtenir une répartition plus compacte des images centrées. Le nouvel estimateur de la matrice d'autocovariance proposé a des valeurs propres négatives. C'est pourquoi il n'a pas été utilisé.

Les premiers vecteurs propres de la matrice d'autocovariance classique définissent un espace propre sur lequel les images de la base d'apprentissage se reconstruisent efficacement pour une erreur de reconstruction de norme euclidienne minimale. Ils permettent de définir une relation linéaire entre les images d'apprentissage  $\mathbf{x}$  de grande dimension, et une signature plus petite  $\mathbf{c}$ . Cette dernière est utilisée pour indexer efficacement chaque image.

Pour caractériser la base d'apprentissage  $\mathcal{B}$ , on utilise la densité de probabilité de  $\mathbf{x}$  :  $\mathcal{P}(\mathbf{x}|\mathcal{B})$ . Suivant quelques hypothèses, on a montré comment exprimer cette densité de probabilité en fonction des coordonnées  $\mathbf{c}$  et de l'erreur de reconstruction.

Les deux méthodes d'apprentissage, classique  $(\hat{\mu}^c, \hat{\Sigma}^c)$  et nouvelle  $(\hat{\mu}^n, \hat{\Sigma}^n)$ , seront comparées expérimentalement au chapitre 6 pour la reconnaissance d'objets, et au chapitre 7 pour la détection.



# Chapitre 6

## Reconnaissance robuste

Ce chapitre présente plusieurs méthodes de reconnaissance utilisant la représentation globale de l'apparence. Une image ou observation  $y$  doit être identifiée parmi les  $K$  images de la base  $\mathcal{B}$  d'apprentissage. La stratégie utilisée consiste à estimer ses coordonnées réduites  $\mathbf{c} = \{c_j\}_{j \in \{1, \dots, J\}}$  dans l'espace propre  $F$ , puis à l'identifier au modèle d'apprentissage dont la projection dans  $F$  est la plus proche. Ceci est illustré par la figure 6.1 : l'estimée  $\hat{\mathbf{c}}$  du représentant de  $y$  dans l'espace propre  $F$  est comparée en terme de distance euclidienne aux coordonnées  $\mathbf{c}^k$  qui indexent chaque image d'apprentissage  $\mathbf{x}^k$ . L'observation  $y$  est identifiée par le modèle  $\mathbf{x}^k$  dont l'index  $\mathbf{c}^k$  est le plus proche.

**Modèle classique.** Les méthodes classiques de reconnaissance sur espaces propres, supposent que l'observation  $y$  est une réalisation de la variable apprise  $\mathbf{x}$ , ce qui implique :  $\tilde{\mathbf{y}} = \mathbf{U}\mathbf{c} + \mathbf{w}^r$ . Les coordonnées réduites  $\mathbf{c}$  de  $y$  sont donc simplement calculées en utilisant le projecteur défini dans l'apprentissage soit :

$$\mathbf{c} = \mathbf{U}^T \tilde{\mathbf{y}}$$

Cependant, on souhaite aussi pouvoir reconnaître des occurrences  $y$  dégradées ou occultées de  $\mathbf{x}$ . De tels effets impliquent l'apparition de données erronées qui peuvent perturber le calcul des coordonnées réduites de l'observation. Ainsi, les composantes réduites de  $\mathbf{x}$ ,  $\mathbf{c}_x = \mathbf{U}^T \tilde{\mathbf{x}}$ , et d'une de ses occurrences dégradées  $\mathbf{c}_y = \mathbf{U}^T \tilde{\mathbf{y}}$ , peuvent être très éloignées l'une de l'autre, ce qui compromet la reconnaissance de l'observation.

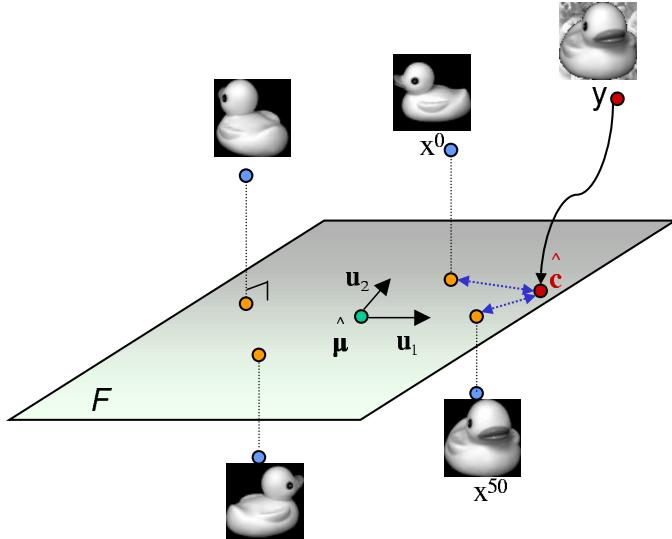


FIG. 6.1: Reconnaissance par modèle d'apparence globale. Les images d'apprentissage  $\mathbf{x}^0, \mathbf{x}^1, \dots, \mathbf{x}^{50}$  sont représentées comme points dans un espace propre  $F$ . Chaque point  $\mathbf{x}^k$  a une coordonnée réduite  $\mathbf{c}^k$  associée. L'observation  $y$  est estimée par son vecteur  $\hat{\mathbf{c}}$  dans  $F$ . La distance entre  $\hat{\mathbf{c}}$  et les  $\mathbf{c}^k$  indique la proximité de l'observation avec les images d'apprentissage. Cependant, si  $y$  est dégradée ou occultée, l'estimation  $\hat{\mathbf{c}}$  peut être éloignée des  $\mathbf{c}^k$ , ce qui peut compromettre la reconnaissance.

**Nouveau modèle.** Le modèle classique qui suppose que l'observation  $\mathbf{y}$  est une réalisation de la variable apprise  $\mathbf{x}$ , n'est pas satisfaisant pour reconnaître des occurrences fortement dégradées. Aussi, on suppose que  $\mathbf{y}$  est une occurrence bruitée d'une image d'apprentissage  $\mathbf{x}$  telle que  $\mathbf{y} = \mathbf{x} + \mathbf{w}^o$  ou :

$$\tilde{\mathbf{y}} = \sum_{j=1}^J c_j \mathbf{u}_j + \mathbf{w}$$

Le bruit noté  $\mathbf{w}$  est la somme de l'erreur de reconstruction  $\mathbf{w}^r$  et d'un bruit d'observation  $\mathbf{w}^o$ . Ce dernier correspond à toutes les perturbations qui peuvent détériorer  $\mathbf{y}$  : occultations partielles, fond texturé, bruit d'acquisition ...

**Bruit  $\mathbf{w}$ .** Deux hypothèses sur la distribution de l'erreur  $\mathbf{w}$  sont envisagées dans ce chapitre. La première suppose une distribution gaussienne de l'erreur  $\mathbf{w}$ . Cette hypothèse permet de modéliser un bruit dû au système d'acquisition mais ne permet pas la prise en compte des erreurs grossières induites par les occultations. Aussi, nous proposons une seconde hypothèse pour la distribution de  $\mathbf{w}$ , plus robuste aux données aberrantes.

**Hypothèse de bruit gaussien.** On suppose que l'erreur a une distribution gaussienne de moyenne nulle, indépendante et identiquement distribuée :

$$\mathcal{P}(\mathbf{w}|\mathcal{B}) \propto \exp\left[-\frac{\|\mathbf{w}\|^2}{2\sigma_g^2}\right]$$

Dans le cadre de cette hypothèse, on note  $\mathcal{N}(\mathbf{0}, \sigma_g^2 I_{NP})$  cette loi normale de la variable aléatoire  $\mathbf{w}$  de dimension  $NP$ .  $I_{NP}$  est la matrice identité de dimension  $NP \times NP$ .

**Hypothèse robuste.** L'hypothèse d'un bruit  $\mathbf{w}$  gaussien n'est pas satisfaisante lorsque l'on souhaite reconnaître des observations représentant une réalisation d'un objet de la base sur un fond texturé, ou lorsque des occultations apparaissent. Pour prendre en compte l'apparition éventuelle d'erreurs grossières (*outliers*), nous modélisons la distribution de l'erreur par :

$$\begin{aligned} \mathcal{P}(\mathbf{w}|\mathcal{B}) &\propto \prod_{n=1}^N \exp\left[-\frac{1}{2} \rho\left(\frac{\|w_n\|}{\sigma_\rho}\right)\right] \\ &\propto \prod_{n=1}^N \exp\left[-\frac{1}{2} \rho\left(\frac{\epsilon_n}{\sigma_\rho}\right)\right] \end{aligned}$$

Les fonctions  $\rho$  utilisées sont présentées à la section 6.3.1. Le résidu  $\epsilon_n$  défini sur chaque pixel  $n$ , est calculé par  $\epsilon_n = \|w_n\|$  pour  $P > 1$  soit :

$$\epsilon_n = \sqrt{\sum_{p=1}^P (w_{p,n})^2}$$

Dans le cas  $P = 1$ , on a  $\epsilon_n = w_n$ . La distribution robuste peut être interprétée comme un mélange de deux lois, l'une gaussienne correspondant à la distribution des données «bonnes» (*inliers*), l'autre correspondant à la distribution des données erronées (*outliers*) [62, 56]. Cette remarque permet de proposer à la section 6.3.2, une nouvelle méthode pour fixer le paramètre d'échelle  $\sigma_\rho$  à partir des images d'apprentissage.

**Estimations des composantes  $\mathbf{c}$ .** Deux approches sont utilisées pour l'estimation des  $J$  coordonnées de  $\mathbf{c}$ , l'approche classique où  $\mathbf{c}$  est un vecteur de paramètres cachés, et l'approche bayésienne où  $\mathbf{c}$  est un vecteur aléatoire caché [68]. Ces deux approches sont développées en section 6.1.

Pour calculer les composantes  $\mathbf{c}$ , l'hypothèse robuste sur la distribution du bruit implique la minimisation d'une énergie non quadratique. Deux algorithmes de minimisation sont proposés à la section 6.2. Deux approches différentes permettent de les obtenir : la théorie semi-quadratique [25], selon le terme introduit par Geman et al. [50, 49], et l'utilisation d'une famille de polynômes proposée par Huber [62]. En plus de leur caractère non quadratique, les énergies peuvent être non convexes, ce qui implique la présence éventuelle de minima locaux. Afin de les éviter, la section 6.3 présente une stratégie originale de minimisations successives, en continuation, d'une énergie convexe vers une énergie non convexe.

La section 6.4 présente des résultats de reconnaissance sur deux bases d'apprentissage, l'une d'images en niveaux de gris, et l'autre sur des images en couleurs.

## 6.1 Approches statistiques

De manière générale, la vraisemblance conjointe de l'observation  $\mathbf{y}$  et de ses composantes  $\mathbf{c}$  s'écrit :

$$\begin{aligned} \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B}) &= \mathcal{P}(\mathbf{y}|\mathbf{c}, \mathcal{B}) \mathcal{P}(\mathbf{c}|\mathcal{B}) \\ &= \mathcal{P}(\mathbf{w}|\mathcal{B}) \mathcal{P}(\mathbf{c}|\mathcal{B}) \end{aligned}$$

Les deux approches statistiques proposées dans cette section diffèrent sur la nature supposée de  $\mathbf{c}$ .

### 6.1.1 Approche classique : estimateur du maximum de vraisemblance

Dans cette approche, les composantes  $\mathbf{c}$  sont des paramètres cachés, déterministes, donc :

$$\mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B}) = \mathcal{P}(\mathbf{w}|\mathcal{B})$$

Dans ce cas, l'estimation au sens du maximum de vraisemblance classique se calcule par :

$$\hat{\mathbf{c}} = \arg \max_{\mathbf{c}} \{\mathcal{P}(\mathbf{w}|\mathcal{B})\}$$

On pose  $\mathcal{J}(\mathbf{c}) = -\ln \mathcal{P}(\mathbf{w}|\mathcal{B})$ , alors :

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c}} \{\mathcal{J}(\mathbf{c})\}$$

### 6.1.1.1 Bruit gaussien

Si le bruit  $\mathbf{w}$  a une distribution gaussienne, alors l'énergie à minimiser, notée  $\mathcal{J}_{MV}$ , s'écrit à une constante additive près :

$$\mathcal{J}_{MV}(\mathbf{c}) = \|\mathbf{w}\|^2 = \|\tilde{\mathbf{y}} - \mathbf{U}\mathbf{c}\|^2 \quad (6.1)$$

L'estimateur  $\hat{\mathbf{c}}_{MV}$  qui minimise l'énergie  $\mathcal{J}_{MV}$ , est simplement la projection orthogonale de  $\tilde{\mathbf{y}}$  sur l'espace propre  $F$ .

### 6.1.1.2 Hypothèse robuste de la distribution du bruit

Dans le cas de l'hypothèse robuste de distribution du bruit, l'estimation des composantes  $\mathbf{c}$  est calculée par :

$$\hat{\mathbf{c}}_{RMV} = \arg \min_{\mathbf{c}} \left\{ \mathcal{J}_{RMV}(\mathbf{c}) = \sum_{n=1}^N \rho \left( \frac{\epsilon_n}{\sigma_\rho} \right) \right\} \quad (6.2)$$

Ces estimateurs, appelés estimateurs généralisés du maximum de vraisemblance ou *M-estimateurs*, ont été développés par Huber dans le cadre des statistiques robustes [62]. Leur utilisation pour réaliser la reconnaissance d'images sur les espaces propres a été proposée par Black [15].

## 6.1.2 Approche Bayésienne

On propose d'utiliser l'approche bayésienne où  $\mathbf{c}$  est un vecteur aléatoire :

$$\mathcal{P}(\mathbf{y}, \mathbf{c} | \mathcal{B}) = \mathcal{P}(\mathbf{w} | \mathcal{B}) \mathcal{P}(\mathbf{c} | \mathcal{B})$$

L'estimation de  $\mathbf{c}$  est réalisée en maximisant  $\mathcal{P}(\mathbf{y}, \mathbf{c} | \mathcal{B})$  ou en minimisant son énergie associée :

$$\mathcal{J}(\mathbf{c}) = -\ln \mathcal{P}(\mathbf{y}, \mathbf{c} | \mathcal{B})$$

Cet estimateur est équivalent à l'estimateur du Maximum a Posteriori [68] car, par le théorème de Bayes, on a :

$$\mathcal{P}(\mathbf{c} | \mathbf{y}, \mathcal{B}) = \frac{\mathcal{P}(\mathbf{w} | \mathcal{B}) \mathcal{P}(\mathbf{c} | \mathcal{B})}{\mathcal{P}(\mathbf{y} | \mathcal{B})}$$

La probabilité  $\mathcal{P}(\mathbf{y} | \mathcal{B})$  ne dépend pas de  $\mathbf{c}$  alors  $\arg \max_c \mathcal{P}(\mathbf{c} | \mathbf{y}, \mathcal{B}) = \arg \max_c \mathcal{P}(\mathbf{y}, \mathbf{c} | \mathcal{B})$ . L'estimation de  $\mathbf{c}$  dépend :

- de l'hypothèse faite *a priori* sur la distribution de  $\mathbf{c}$  dans le plan propre  $F$ . A titre d'exemple, deux hypothèses simples de distribution sont abordées dans les paragraphes suivants : la distribution uniforme et la distribution gaussienne.
- et de l'hypothèse faite *a priori* sur la distribution du bruit  $\mathbf{w}$ . Deux hypothèses ont été considérées : la distribution gaussienne centrée indépendante identiquement distribuée, et la distribution robuste.

Ces quatre cas sont étudiés dans les paragraphes suivants.

### 6.1.2.1 Hypothèses : distribution uniforme des coordonnées c et distribution gaussienne du bruit w

On suppose que les coordonnées  $\mathbf{c}$  sont distribuées uniformément dans le plan propre  $F$ , soit  $\mathcal{P}(\mathbf{c}|\mathcal{B}) = cste$ , et que l'erreur  $\mathbf{w}$  a une distribution gaussienne. Alors, l'estimateur qui maximise la vraisemblance  $\mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$  se calcule par :

$$\hat{\mathbf{c}}_{MAPu} = \arg \min_{\mathbf{c}} \left\{ \mathcal{J}_{MAPu}(\mathbf{c}) = \|\mathbf{w}\|^2 \right\}$$

On remarque que l'estimateur bayésien  $\hat{\mathbf{c}}_{MAPu}$  est équivalent à l'estimateur du maximum de vraisemblance  $\hat{\mathbf{c}}_{MV}$ .

### 6.1.2.2 Hypothèses : distribution uniforme des coordonnées c et distribution robuste du bruit w

On suppose que le vecteur aléatoire  $\mathbf{c}$  est distribué uniformément dans l'espace propre  $F$ . La probabilité conjointe  $\mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$  est proportionnelle à  $\mathcal{P}(\mathbf{w}|\mathcal{B})$ . En négligeant les constantes, l'estimation s'écrit :

$$\hat{\mathbf{c}}_{RMAPu} = \arg \min_{\mathbf{c}} \left\{ \mathcal{J}_{RMAPu}(\mathbf{c}) = \sum_{n=1}^N \rho \left( \frac{\epsilon_n}{\sigma_\rho} \right) \right\} \quad (6.3)$$

On remarque que l'estimateur  $\hat{\mathbf{c}}_{RMAPu}$  est équivalent à l'estimateur robuste du maximum de vraisemblance  $\hat{\mathbf{c}}_{RMV}$ .

### 6.1.2.3 Hypothèses : distribution gaussienne des coordonnées c et distribution gaussienne du bruit w

On suppose une distribution gaussienne de  $\mathbf{c}$  dans l'espace propre  $F$ . Les valeurs propres  $\{\lambda_j\}_{j \in \{1, \dots, J\}}$  calculées par ACP (cf. chapitre 5), sont les variances associées aux  $\{c_j\}_{j \in \{1, \dots, J\}}$  et :

$$\mathcal{P}(c_1, \dots, c_J | \mathcal{B}) = \prod_{j=1}^J \frac{1}{\sqrt{2\pi\lambda_j}} \exp \left[ -\frac{c_j^2}{2\lambda_j} \right]$$

En supposant un bruit gaussien, l'estimateur s'écrit :

$$\hat{\mathbf{c}}_{MAP} = \arg \min_{\mathbf{c}} \left\{ \mathcal{J}_{MAP}(\mathbf{c}) = \frac{\|\mathbf{w}\|^2}{\sigma_g^2} + \sum_{j=1}^J \frac{c_j^2}{\lambda_j} \right\}$$

### 6.1.2.4 Hypothèses : distribution gaussienne des coordonnées $\mathbf{c}$ et distribution robuste du bruit $\mathbf{w}$

On suppose une distribution gaussienne de  $\mathbf{c}$  dans l'espace propre  $F$  et la distribution robuste du bruit  $\mathbf{w}$ . Alors, l'estimateur bayésien se calcule par :

$$\hat{\mathbf{c}}_{RMAP} = \arg \min_{\mathbf{c}} \left\{ \mathcal{J}_{RMAP}(\mathbf{c}) = \sum_{n=1}^N \rho \left( \frac{\epsilon_n}{\sigma_\rho} \right) + \sum_{j=1}^J \frac{c_j^2}{\lambda_j} \right\} \quad (6.4)$$

### 6.1.3 Remarques

L'approche bayésienne permet d'introduire un *a priori* sur la distribution des composantes  $\mathbf{c}$  dans l'espace propre  $F$ . Deux cas particuliers ont été abordés dans le cadre de l'approche bayésienne : la distribution uniforme et la distribution gaussienne de  $\mathbf{c}$  dans  $F$ . D'autres hypothèses relatives à la distribution de  $\mathbf{c}$  dans  $F$  pourraient être envisagées (cf. section 5.2.2). Cependant, la complexité algorithmique nécessaire à l'estimation peut augmenter.

Les hypothèses,  $\mathbf{c}$  paramètre et  $\mathbf{c}$  variable aléatoire uniformément distribuée, ne sont pas équivalentes [118]. Cependant, l'estimateur bayésien  $\hat{\mathbf{c}}_{MAPu}$  (respectivement  $\hat{\mathbf{c}}_{RMAPu}$ ) est équivalent à l'estimateur du maximum de vraisemblance classique  $\hat{\mathbf{c}}_{MV}$  (resp.  $\hat{\mathbf{c}}_{RMV}$ ). Aussi, les notations des différents estimateurs finalement utilisées, sont rappelées au tableau 6.1.

	Bruit gaussien	Hypothèse robuste
Distribution uniforme de $\mathbf{c}$ ou $\mathbf{c}$ paramètre	$\hat{\mathbf{c}}_{MV} = \arg \min_{\mathbf{c}} \mathcal{J}_{MV}(\mathbf{c})$	$\hat{\mathbf{c}}_{RMV} = \arg \min_{\mathbf{c}} \mathcal{J}_{RMV}(\mathbf{c})$
Distribution gaussienne de $\mathbf{c}$	$\hat{\mathbf{c}}_{MAP} = \arg \min_{\mathbf{c}} \mathcal{J}_{MAP}(\mathbf{c})$	$\hat{\mathbf{c}}_{RMAP} = \arg \min_{\mathbf{c}} \mathcal{J}_{RMAP}(\mathbf{c})$

TAB. 6.1:Notations des estimateurs suivant les hypothèses *a priori* sur les distributions de probabilité de  $\mathbf{c}$  et  $\mathbf{w}$ .

Dans le cas de la distribution robuste de  $\mathbf{w}$ , les énergies à minimiser sont non quadratiques. Le paragraphe suivant rappelle deux méthodes de minimisation.

## 6.2 Minimisation d'énergies non quadratiques

A la section 6.2.1, on propose deux approches pour minimiser l'énergie  $\mathcal{J}_{RMV}$ , non quadratique lorsque l'on considère l'hypothèse de la distribution robuste de l'erreur. Pour simplifier, on considère le cas des images en niveaux de gris. L'extension de ces méthodes aux images multibandes est proposée dans la section suivante 6.2.2. L'ensemble des algorithmes utilisés pour minimiser les énergies définies à la section 6.1, est résumé au paragraphe 6.2.3.

### 6.2.1 Cas des images en niveau de gris

La première méthode de minimisation de l'énergie non quadratique  $\mathcal{J}_{RMV}$ , utilise la théorie semi-quadratique [25]. Cette méthode introduit une nouvelle fonction d'énergie définie sur un domaine plus étendu, mais qui a le même minimum global que l'énergie originale, et qui peut être réduite par des techniques linéaires [25]. La seconde, proposée par Huber dans le cadre des statistiques robustes, utilise une famille de polynômes pour approcher l'énergie non-quadratique à minimiser [62]. On considère des images en niveaux de gris, soit  $P = 1$ , alors on a l'égalité  $\epsilon_n = w_n$ . On rappelle que l'erreur  $\mathbf{w}$  est liée à la variable  $\mathbf{c}$  par la relation linéaire  $\mathbf{w} = \tilde{\mathbf{y}} - \mathbf{U} \mathbf{c}$ .

#### 6.2.1.1 Théorie semi-quadratique

**Principe.** En introduisant une variable auxiliaire, la théorie semi-quadratique propose de minimiser une énergie augmentée  $\mathcal{J}_{RMV}^*$  définie par [50, 49, 25] :

$$\mathcal{J}_{RMV}(\mathbf{c}) = \min_{\mathbf{b}} \left\{ \mathcal{J}_{RMV}^*(\mathbf{c}, \mathbf{b}) = \sum_{n=1}^N \rho^* \left( \frac{w_n}{\sigma_\rho}, b_n \right) \right\} \quad (6.5)$$

Deux formes de fonctions augmentées  $\rho^*$  ont été proposées dans la littérature [25], et conduisent à deux algorithmes de minimisation.

**Propriétés 1** *La fonction  $\rho$  considérée vérifie les conditions suivantes :*

- $\rho$  symétrique, croissante sur  $\mathbb{R}^+$ .
- $b^A(t) = \frac{\rho'(t)}{2t}$ , strictement décroissante sur  $\mathbb{R}^+$  et bornée :

$$\begin{cases} \lim_{t \rightarrow \infty} b^A(t) = 0 \\ \lim_{t \rightarrow 0} b^A(t) = 1 \end{cases}$$

Les fonctions  $\rho$ , utilisées dans cette thèse et présentées au paragraphe 6.3.1, vérifient ces conditions.

**Principe des minimisations alternées.** L'énergie  $\mathcal{J}_{RMV}^*(\mathbf{c}, \mathbf{b})$  est diminuée en résolvant itérativement jusqu'à convergence :

$$\left| \begin{array}{l} \frac{\partial \mathcal{J}_{RMV}^*(\mathbf{c}, \mathbf{b})}{\partial \mathbf{c}} = 0 \\ \frac{\partial \mathcal{J}_{RMV}^*(\mathbf{c}, \mathbf{b})}{\partial \mathbf{b}} = 0 \end{array} \right.$$

Ce schéma déterministe converge vers un minimum local de  $\mathcal{J}_{RMV}$ .

---

### Algorithme ARTUR.

Ce premier algorithme, appelé ARTUR, découle du « théorème de Geman et Reynold étendu » [50, 25] :

**Théorème 1** Pour une fonction  $\rho$  qui vérifie les propriétés 1,

1) il existe une fonction  $\Psi : [0, 1] \rightarrow [0, 1]$  strictement convexe et décroissante telle que :

$$\rho(t) = \inf_{b^A} \{b^A t^2 + \Psi(b^A)\}$$

2) Pour tout  $t \leq 0$ , la valeur de  $b^A$  pour laquelle l'inf est atteint, est unique et est donnée par :

$$b^A = \frac{\rho'(t)}{2t}$$

L'énergie augmentée, notée  $\mathcal{J}_{RMV}^A$ , s'écrit :

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

On note  $\mathbf{b}^A$  le vecteur de dimension  $N$  collectant les valeurs auxiliaires  $b_n^A$ . A l'étape  $(m+1)$ , connaissant l'estimée  $\mathbf{c}^{(m)}$ , l'algorithme itératif ARTUR minimisant  $\mathcal{J}_{RMV}^A$  alternativement en  $\mathbf{c}$  et  $\mathbf{b}^A$ , s'écrit :

$$\left| \begin{array}{l} \mathbf{w}^{(m)} = \tilde{\mathbf{y}} - \mathbf{U}\mathbf{c}^{(m)} \\ \\ \forall n \in \{1 \dots N\}, b_n^{A(m+1)} = \frac{\rho'\left(\frac{w_n^{(m)}}{\sigma_\rho}\right)}{2\frac{w_n^{(m)}}{\sigma_\rho}} \\ \\ (\mathbf{U}^T \cdot \mathbf{B}^{A(m+1)} \cdot \mathbf{U}) \mathbf{c}^{(m+1)} = \mathbf{U}^T \cdot \mathbf{B}^{A(m+1)} \cdot \tilde{\mathbf{y}} \end{array} \right. \quad (6.6)$$

avec  $\mathbf{B}^A$  une matrice diagonale de dimension  $N \times N$  contenant sur sa diagonale les pondérations  $\{b_1^A, \dots, b_N^A\}$ . Cet algorithme ARTUR est un algorithme des moindres carrés pondérés itératifs.

**Algorithme LEGEND.**

Le « deuxième théorème de Geman et Reynold étendu » a été démontré en utilisant les paires de Legendre [49, 25] :

**Théorème 2** *Pour une fonction  $\rho$  qui vérifie les propriétés 1,*

1) *il existe une fonction  $\xi : [0, +\infty] \rightarrow [0, +\infty]$  telle que :*

$$\rho(t) = \inf_{b^L} \{(b^L - t)^2 + \xi(b^L)\}$$

2) *Pour tout  $t \leq 0$ , la valeur de  $b^L$  pour laquelle l'inf est atteint, est unique et est donnée par :*

$$b^L = t \left( 1 - \frac{\rho'(t)}{2t} \right)$$

L'énergie augmentée à minimiser, notée  $\mathcal{J}_{RMV}^L$ , s'écrit :

$$\mathcal{J}_{RMV}^L(\mathbf{c}, \mathbf{b}^L) = \sum_{n=1}^N \left( \frac{w_n}{\sigma_\rho} - b_n^L \right)^2 + \xi(b_n^L) \quad (6.7)$$

$\mathbf{b}^L$  est le vecteur de dimension  $N$  collectant les valeurs des variables auxiliaires  $b_n^L$ . A l'étape  $(m+1)$ , connaissant l'estimée  $\mathbf{c}^{(m)}$ , l'algorithme itératif LEGEND minimisant  $\mathcal{J}_{RMV}^L$  s'écrit :

$$\left| \begin{array}{l} \mathbf{w}^{(m)} = \tilde{\mathbf{y}} - \mathbf{U}\mathbf{c}^{(m)} \\ \\ \forall n \in \{1 \dots N\}, \quad b_n^{L(m+1)} = \frac{w_n^{(m)}}{\sigma_\rho} \left( 1 - \frac{\rho'\left(\frac{w_n^{(m)}}{\sigma_\rho}\right)}{2 \frac{w_n^{(m)}}{\sigma_\rho}} \right) \\ \\ \mathbf{c}^{(m+1)} = \mathbf{U}^T \cdot (\tilde{\mathbf{y}} - \sigma_\rho \mathbf{b}^{L(m+1)}) \end{array} \right. \quad (6.8)$$

**6.2.1.2 Approche de Huber**

**Propriétés 2** *La fonction  $\rho$  considérée vérifie les conditions suivantes :*

- $\rho$  symétrique, convexe sur  $\mathbb{R}^+$ .
- $\beta^w(t) = \frac{\rho'(t)}{2t}$  décroissante sur  $\mathbb{R}^+$  et bornée :

$$\left| \begin{array}{l} \lim_{t \rightarrow \infty} \beta^w(t) = 0 \\ \lim_{t \rightarrow 0} \beta^w(t) = 1 \end{array} \right.$$

La méthode de minimisation proposée par Huber, consiste à approcher localement en  $t_0$  la fonction  $\rho(t)$  par un polynôme de degré 2,  $q^*$ , vérifiant en  $t_0$  les conditions [62] :

$$\left| \begin{array}{l} q^*(t_0) = \rho(t_0) \\ q^*(t_0) = \rho'(t_0) \end{array} \right. \quad (6.9)$$

En  $\mathbf{c}^{(m)}$  ( $\mathbf{w}^{(m)} = \tilde{\mathbf{y}} - \mathbf{U}\mathbf{c}^{(m)}$ ), l'énergie  $\mathcal{J}_{RMV}$  est approchée par l'énergie quadratique :

$$\mathcal{J}_{RMV}^*(\mathbf{c}) = \sum_{n=1}^N q^* \left( \frac{w_n^{(m)}}{\sigma_\rho} \right) \quad (6.10)$$

Si  $\mathbf{c}^{(m+1)}$  est le minimum de  $\mathcal{J}_{RMV}^*$ , Huber montre que l'énergie  $\mathcal{J}_{RMV}$  diminue :

$$\mathcal{J}_{RMV}(\mathbf{c}^{(m+1)}) \leq \mathcal{J}_{RMV}(\mathbf{c}^{(m)})$$

On redéfinit en  $\mathbf{c}^{(m+1)}$  un nouveau polynôme  $q^*$ , pour diminuer à nouveau l'énergie  $\mathcal{J}_{RMV}$ . Ce processus est répété jusqu'à atteindre le minimum de  $\mathcal{J}_{RMV}$ . Les algorithmes proposés sont des algorithmes de descente.

Deux formes de polynômes ont été proposées et permettent d'aboutir à deux algorithmes de minimisation. Celles-ci sont représentées à la figure 6.2, pour la fonction  $\rho_{GM}$  de Geman et McClure présentée au paragraphe 6.3.1.

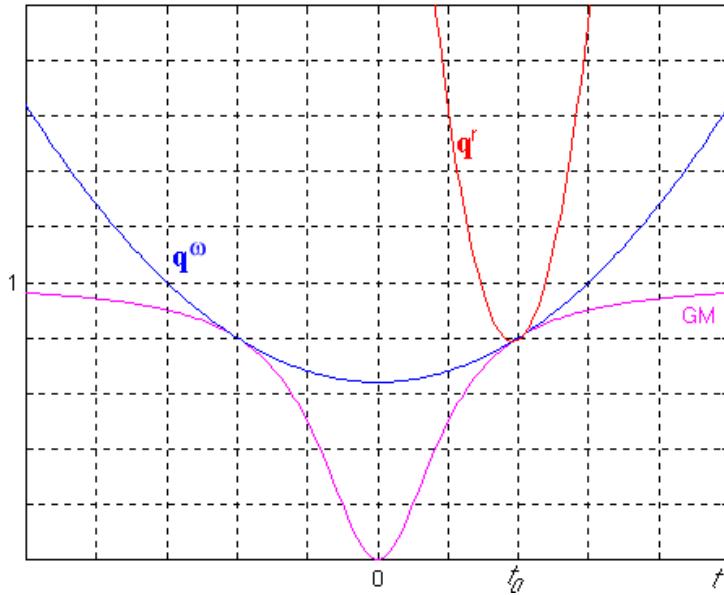


FIG. 6.2: Polynômes  $q^r$  et  $q^\omega$  approchant, en un point  $t_0$ , la fonction  $\rho_{GM}$ .

---

### Minimisation avec des poids modifiés.

Au voisinage de  $t_0$ , la fonction  $\rho(t)$  est approchée par le polynôme (cf. figure 6.2) :

$$q^\omega(t) = \alpha^\omega + \beta^\omega t^2 \text{ avec } \begin{cases} \alpha^\omega = \rho(t_0) - \frac{\rho'(t_0)}{2} t_0 \\ \beta^\omega = \frac{\rho'(t_0)}{2 t_0} \end{cases} \quad (6.11)$$

Au voisinage de  $\mathbf{c}^{(m)}$  ( $\mathbf{w}^{(m)} = \tilde{\mathbf{y}} - \mathbf{U}\mathbf{c}^{(m)}$ ), l'énergie  $\mathcal{J}_{RMV}$  est approchée par l'énergie quadratique  $\mathcal{J}_{RMV}^w$  :

$$\mathcal{J}_{RMV}^w = \sum_{n=1}^N \beta_n^\omega \left( \frac{w_n^{(m)}}{\sigma_\rho} \right)^2 + \alpha_n^\omega \quad \text{avec} \quad \begin{cases} \beta_n^\omega = \frac{\rho' \left( \frac{w_n^{(m)}}{\sigma_\rho} \right)}{2 \frac{w_n^{(m)}}{\sigma_\rho}} \\ \alpha_n^\omega = \rho \left( \frac{w_n^{(m)}}{\sigma_\rho} \right) - \rho' \left( \frac{w_n^{(m)}}{\sigma_\rho} \right) \frac{w_n^{(m)}}{2 \sigma_\rho} \end{cases}$$

L'estimation de  $\mathbf{c}^{(m+1)}$  est réalisée par l'algorithme itératif proposé par Huber [62] :

$$\left| \begin{array}{l} \mathbf{w}^{(m)} = \tilde{\mathbf{y}} - \mathbf{U}\mathbf{c}^{(m)} \\ \\ \forall n \in \{1 \dots N\}, \beta_n^{\omega(m+1)} = \frac{\rho' \left( \frac{w_n^{(m)}}{\sigma_\rho} \right)}{2 \frac{w_n^{(m)}}{\sigma_\rho}} \\ \\ \mathbf{U}^T \cdot \underline{\beta}^{\omega(m+1)} \cdot \mathbf{U} \tau^{(m+1)} = \mathbf{U}^T \cdot \underline{\beta}^{\omega(m+1)} \cdot \mathbf{w}^{(m)} \\ \\ \mathbf{c}^{(m+1)} = \mathbf{c}^{(m)} + \tau \end{array} \right. \quad (6.12)$$

avec  $\underline{\beta}^\omega$  la matrice diagonale de dimension  $N \times N$  contenant les valeurs  $\{\beta_n^w\}_{n \in \{1, \dots, N\}}$ .

### Minimisation avec des résidus modifiés.

---

Au voisinage de  $t_0$ , la fonction  $\rho(t)$  est approchée par le polynôme (cf. figure 6.2) :

$$q^r(t) = (t - \beta^r)^2 + \alpha^r \quad \text{avec les constantes} \quad \begin{cases} \alpha^r = \rho(t_0) - \left( \frac{\rho'(t_0)}{2} \right)^2 \\ \beta^r = t_0 \left( 1 - \frac{\rho'(t_0)}{2 t_0} \right) \end{cases} \quad (6.13)$$

Au voisinage de  $\mathbf{c}^{(m)}$ , l'énergie quadratique de Huber a la forme :

$$\mathcal{J}_{RMV}^r(\mathbf{c}) = \sum_{n=1}^N \left( \frac{w_n^{(m)}}{\sigma_\rho} - \beta_n^r \right)^2 + \alpha_n^r \quad \text{avec les constantes} \quad \begin{cases} \beta_n^r = \frac{w_n^{(m)}}{\sigma_\rho} \left( 1 - \frac{\rho' \left( \frac{w_n^{(m)}}{\sigma_\rho} \right)}{2 \frac{w_n^{(m)}}{\sigma_\rho}} \right) \\ \alpha_n^r = \rho \left( \frac{w_n^{(m)}}{\sigma_\rho} \right) - \frac{1}{4} \rho' \left( \frac{w_n^{(m)}}{\sigma_\rho} \right)^2 \end{cases}$$

On introduit le vecteur  $\mathbf{w}^*$  des résidus modifiés défini par :

$$\forall n \in \{1, \dots, N\}, \mathbf{w}_n^* = \frac{\sigma_\rho}{2} \rho' \left( \frac{w_n}{\sigma_\rho} \right)$$

Le minimum  $\mathbf{c}^{(m+1)}$  de l'énergie quadratique  $\mathcal{J}_{RMV}^r$ , qui permet la décroissance de  $\mathcal{J}_{RMV}$ , est calculé

par :

$$\left| \begin{array}{l} \forall n \in \{1, \dots, N\}, \mathbf{w}_n^{*(m)} = \frac{\sigma_\rho}{2} \rho' \left( \frac{w_n^{(m)}}{\sigma_\rho} \right) \\ \tau^{(m+1)} = \mathbf{U}^T \mathbf{w}^{*(m)} \\ \mathbf{c}^{(m+1)} = \mathbf{c}^{(m)} + \tau^{(m+1)} \end{array} \right. \quad (6.14)$$

**Remarques.** Huber remarque que si l'on omet la propriété de convexité de la fonction  $\rho$ , en supposant qu'elle est simplement croissante, les deux algorithmes proposés convergent vers un minimum local [62]. On peut donc redéfinir les propriétés de la fonction  $\rho$  :

**Propriétés 3** La fonction  $\rho$  considérée vérifie les conditions suivantes :

- $\rho$  croissante symétrique sur  $\mathbb{R}^+$ .
- $\beta^w(t) = \frac{\rho'(t)}{2t}$  décroissante sur  $\mathbb{R}^+$  et bornée :

$$\left| \begin{array}{l} \lim_{t \rightarrow \infty} \beta^w(t) = 0 \\ \lim_{t \rightarrow 0} \beta^w(t) = 1 \end{array} \right.$$

### 6.2.1.3 Lien entre l'approche de Huber et la théorie semi-quadratique

**ARTUR ou minimisation avec des poids modifiés.** On remarque les définitions identiques des variables  $\beta^w$  et  $b^A$ . En remplaçant  $\tau^{(m+1)} = \mathbf{c}^{(m+1)} - \mathbf{c}^{(m)}$  dans l'algorithme de *Minimisation avec des poids modifiés*, l'équation  $\mathbf{U}^T \cdot \underline{\beta}^{\omega(m+1)} \cdot \mathbf{U} \tau = \mathbf{U}^T \cdot \underline{\beta}^{\omega(m+1)} \cdot \mathbf{w}^{(m)}$  devient :

$$\begin{aligned} \mathbf{U}^T \cdot \underline{\beta}^{\omega(m+1)} \cdot \mathbf{U} \mathbf{c}^{(m+1)} &= \mathbf{U}^T \cdot \underline{\beta}^{\omega(m+1)} \cdot (\mathbf{w}^{(m)} + \mathbf{U} \mathbf{c}^{(m)}) \\ &= \mathbf{U}^T \cdot \underline{\beta}^{\omega(m+1)} \cdot \tilde{\mathbf{y}} \end{aligned}$$

On retrouve exactement l'algorithme ARTUR. C'est un algorithme itératif de moindres carrés pondérés, souvent proposé pour résoudre les estimations robustes avec les M-estimateurs [16, 59, 107, 146].

**LEGEND ou minimisation avec des résidus modifiés.** On remarque les définitions identiques des variables  $b^L$  et  $\beta^r$ . Les résidus modifiés  $\mathbf{w}^*$  peuvent se réécrire :

$$\mathbf{w}^* = \mathbf{w} - \sigma_\rho \mathbf{b}^L$$

Alors l'étape  $\tau^{(m+1)} = \mathbf{U}^T \mathbf{w}^{*(m)}$  de l'algorithme de *Minimisation avec des résidus modifiés*, est équivalente à :

$$\begin{aligned} \mathbf{c}^{(m+1)} &= \mathbf{U}^T (\mathbf{w}^{*(m)} + \mathbf{U} \mathbf{c}^{(m)}) \\ &= \mathbf{U}^T (\tilde{\mathbf{y}} - \sigma_\rho \mathbf{b}^{L(m+1)}) \end{aligned}$$

L'algorithme LEGEND est équivalent à l'algorithme de minimisation avec des résidus modifiés.

L'annexe F met en évidence le lien entre la théorie semi-quadratique et la méthode de Huber. Cette comparaison permet de préciser les définitions des fonctions  $\Psi(b^A)$  et  $\xi(b^L)$ .

### 6.2.1.4 Vitesse des deux algorithmes de minimisation

**Nombre d'étapes nécessaire pour atteindre le minimum.** Lorsque les polynômes  $q^r(t)$  et  $q^\omega(t)$  sont définis localement en  $t_0$ , Huber remarque que  $q^\omega(t) \leq q^r(t)$  (cf. annexe F). Cette remarque est illustrée par la figure 6.2. La fonction  $q^\omega$  est la meilleure fonction de comparaison quadratique [62]. Cela implique que l'algorithme de minimisation avec des poids modifiés (ou ARTUR) converge plus rapidement que l'algorithme de minimisation par résidus modifiés (ou LEGEND). La figure 6.3 illustre la convergence plus lente de LEGEND : l'énergie  $J_{RMV}$  est reportée en ordonnée et l'abscisse indique le numéro de l'étape ( $m$ ) des algorithmes. Ces courbes sont calculées pour l'estimation d'un paramètre  $c$  de dimension 60.

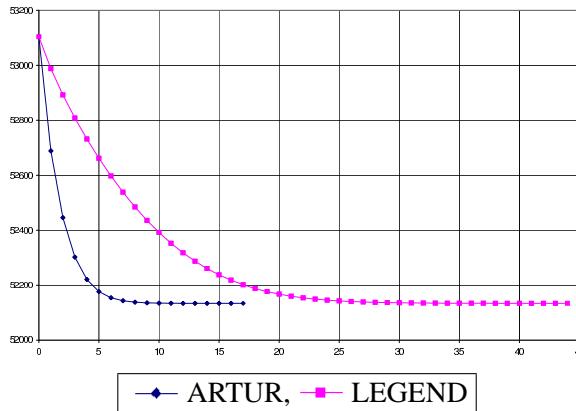


FIG. 6.3: Courbes de convergence des algorithmes ARTUR et LEGEND.

**Coût calculatoire.** Le coût calculatoire d'une étape de LEGEND (ou Minimisation avec des résidus modifiés) est moins important qu'une étape de l'algorithme ARTUR (Minimisation avec des poids modifiés) car elle ne nécessite pas de multiplication de matrice (terme  $U^T B^A U$ ) ni son inversion [62]. En pratique, l'algorithme LEGEND est la plupart du temps plus rapide que ARTUR.

### 6.2.2 Généralisation à des images couleurs ou multi-bandes

Dans le cas des images multi-bandes  $P > 1$ , on a le résidu  $\epsilon = \|w\|$ . Le résidu  $\epsilon$  n'est pas linéaire en  $c$ . Cette section présente les adaptations des algorithmes ARTUR et LEGEND nécessaires pour leur

utilisation sur les images couleurs. On rappelle l'énergie à minimiser :

$$\mathcal{J}_{RMV}(\mathbf{c}) = \sum_{n=1}^N \rho\left(\frac{\epsilon_n}{\sigma_\rho}\right)$$

$$= \sum_{n=1}^N \rho\left(\frac{\|w_n\|}{\sigma_\rho}\right)$$

$w_n$  correspond au vecteur contenant les  $P$  valeurs du pixel  $n$  de l'erreur  $\mathbf{w}$ .

### 6.2.2.1 ARTUR

A l'annexe F, on montre que :

$$\rho(\|w\|) = \inf_{b^A \in \mathbb{R}} \{\rho^A(w, b^A) = b^A \|w\|^2 + \Psi(b^A)\}$$

et l'inf est atteint pour  $b^A = \frac{\rho'(\|w\|)}{2\|w\|}$ .

Sur chaque pixel  $n$ , on peut donc écrire l'énergie augmentée  $\mathcal{J}_{RMV}^A$  :

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

$$= b^A \frac{\|w_n\|^2}{\sigma_\rho^2} + \Psi(b^A)$$

L'algorithme ARTUR pour les images multi-bandes est identique à celui défini pour les images en niveau de gris (équation 6.6), à ceci près que la matrice diagonale  $B^{A(m+1)}$  définie à chaque étape  $(m+1)$  est de dimension  $NP \times NP$  et contient sur sa diagonale  $P$  fois les  $N$  pondérations  $\{b_n^{A(m+1)}\}_{n \in \{1, \dots, N\}}$  :

$$B^{A(m+1)} = \begin{pmatrix} b_1^{A(m+1)} & & & (0) \\ \ddots & & & \\ & b_N^{A(m+1)} & & \\ & & \ddots & \\ & & & b_1^{A(m+1)} \\ & & & \ddots \\ (0) & & & b_N^{A(m+1)} \end{pmatrix}$$

La pondération  $b_n^{A(m+1)} = \frac{\rho\left(\frac{\|w_n^{(m)}\|}{\sigma_\rho}\right)}{2\frac{\|w_n^{(m)}\|}{\sigma_\rho}}$  est appliquée aux  $P$  valeurs de l'erreur  $w_n$ .

### 6.2.2.2 LEGEND

Pour les images multi-bandes, la fonction  $\rho^L$  est définie de  $\mathbb{R}^P \times \mathbb{R}^P$  dans  $\mathbb{R}$  (cf. annexe F) :

$$\rho(\|w\|) = \inf_{b^L \in \mathbb{R}^P} \{\rho^L(w, b^L) = \|w - b^L\|^2 + \xi(b^L)\}$$

et l'inf est atteint pour  $b^L = w \left(1 - \frac{\rho'(\|w\|)}{2\|w\|}\right)$ . L'énergie augmentée  $\mathcal{J}_{RMV}^L$  s'écrit :

$$\mathcal{J}_{RMV}^L(\mathbf{c}, \mathbf{b}^L) = \sum_{n=1}^N \left\| \frac{w_n}{\sigma_\rho} - b_n^L \right\|^2 + \xi(b_n^L) \quad (6.15)$$

L'algorithme LEGEND pour les images multi-bandes est identique à celui défini pour les images en niveau de gris (équation 6.8), avec le vecteur  $\mathbf{b}^{L(m+1)}$  de dimension  $NP$ , calculé par :

$$\mathbf{b}^{L(m+1)} = \left( \mathbf{I}_{NP} - \mathbf{B}^{A(m+1)} \right) \frac{\mathbf{w}^{(m)}}{\sigma_\rho} \quad (6.16)$$

avec  $\mathbf{I}_{NP}$  la matrice identité de dimension  $NP \times NP$  et  $\mathbf{B}^{A(m+1)}$  la matrice diagonale des pondérations.

### 6.2.3 Algorithmes d'estimation

Cette section résume les algorithmes utilisés pour calculer les estimateurs  $\hat{\mathbf{c}}$  définis à la section 6.1.

#### 6.2.3.1 Estimation de $\hat{\mathbf{c}}_{MV}$

L'estimation au sens des moindres carrés se ramène à une simple projection sur l'espace propre  $F$  :

$$\hat{\mathbf{c}}_{MV} = \mathbf{U}^T \cdot \tilde{\mathbf{y}} \quad (6.17)$$

#### 6.2.3.2 Estimation de $\hat{\mathbf{c}}_{RMV}$

L'estimation se calcule par les algorithmes itératifs ARTUR ou LEGEND rappelés aux équations 6.18 et 6.19.

$$ARTUR \left| \begin{array}{l} \mathbf{w}^{(m)} = \tilde{\mathbf{y}} - \mathbf{U} \mathbf{c}^{(m)} \\ \forall n \in \{1 \dots N\}, b_n^{A(m+1)} = \frac{\rho' \left( \frac{\|w_n^{(m)}\|}{\sigma_\rho} \right)}{2 \frac{\|w_n^{(m)}\|}{\sigma_\rho}} \\ \mathbf{c}^{(m+1)} = (\mathbf{U}^T \cdot \mathbf{B}^{A(m+1)} \cdot \mathbf{U})^{-1} \cdot \mathbf{U}^T \cdot \mathbf{B}^{A(m+1)} \cdot \tilde{\mathbf{y}} \end{array} \right. \quad (6.18)$$

et

$$LEGEND \left| \begin{array}{l} \mathbf{w}^{(m)} = \tilde{\mathbf{y}} - \mathbf{U} \mathbf{c}^{(m)} \\ \forall n \in \{1 \dots N\}, b_n^{L(m+1)} = \frac{w_n^{(m)}}{\sigma_\rho} \left( 1 - \frac{\rho' \left( \frac{\|w_n^{(m)}\|}{\sigma_\rho} \right)}{2 \frac{\|w_n^{(m)}\|}{\sigma_\rho}} \right) \\ \mathbf{c}^{(m+1)} = \mathbf{U}^T \cdot (\tilde{\mathbf{y}} - \sigma_\rho \mathbf{b}^{L(m+1)}) \end{array} \right. \quad (6.19)$$

A convergence, on a l'estimée  $\hat{\mathbf{c}}_{RMV}$

### 6.2.3.3 Estimation de $\hat{\mathbf{c}}_{MAP}$

L'estimateur  $\hat{\mathbf{c}}_{MAP}$  se calcule par :

$$\hat{\mathbf{c}}_{MAP} = (\mathbf{I}_J + \sigma_g^2 \Lambda^{-1})^{-1} \mathbf{U}^T \cdot \tilde{\mathbf{y}} \quad (6.20)$$

où  $\Lambda$  est la matrice diagonale  $J \times J$  des valeurs propres calculées lors de l'apprentissage de  $\mathcal{B}$  par ACP :

$$\Lambda = \begin{pmatrix} \lambda_1 & & & (0) \\ & \ddots & & \\ & & \ddots & \\ (0) & & & \lambda_J \end{pmatrix}$$

### 6.2.3.4 Estimation de $\hat{\mathbf{c}}_{RMAP}$

Les énergies augmentées s'écrivent :

$$J_{RMAP}^A(\mathbf{c}, \mathbf{b}^A) = \sum_{n=1}^N \left( b_n^A \frac{\epsilon_n^2}{\sigma_\rho^2} + \Psi(b_n^A) \right) + \sum_{j=1}^J \frac{c_j^2}{\lambda_j}$$

et

$$J_{RMAP}^L(\mathbf{c}, \mathbf{b}^L) = \sum_{n=1}^N \left( \left\| \frac{w_n}{\sigma_\rho} - b_n^L \right\|^2 + \xi(b_n^L) \right) + \sum_{j=1}^J \frac{c_j^2}{\lambda_j}$$

Les équations 6.21 et 6.22 présentent les deux algorithmes itératifs, ARTUR et LEGEND, pour l'estimation de  $\hat{\mathbf{c}}_{RMAP}$ .

$$ARTUR \left| \begin{array}{l} \mathbf{w}^{(m)} = \tilde{\mathbf{y}} - \mathbf{U} \mathbf{c}^{(m)} \\ \forall n \in \{1 \dots N\}, b_n^{A(m+1)} = \frac{\rho' \left( \frac{\|w_n^{(m)}\|}{\sigma_\rho} \right)}{2 \frac{\|w_n^{(m)}\|}{\sigma_\rho}} \\ \mathbf{c}^{(m+1)} = (\mathbf{U}^T \cdot \mathbf{B}^{A(m+1)} \cdot \mathbf{U} + \sigma_\rho^2 \Lambda^{-1})^{-1} \cdot \mathbf{U}^T \cdot \mathbf{B}^{A(m+1)} \cdot \tilde{\mathbf{y}} \end{array} \right. \quad (6.21)$$

$$LEGEND \left| \begin{array}{l} \mathbf{w}^{(m)} = \tilde{\mathbf{y}} - \mathbf{U} \mathbf{c}^{(m)} \\ \forall n \in \{1 \dots N\}, b_n^{L(m+1)} = \frac{w_n^{(m)}}{\sigma_\rho} \left( 1 - \frac{\rho' \left( \frac{\|w_n^{(m)}\|}{\sigma_\rho} \right)}{2 \frac{\|w_n^{(m)}\|}{\sigma_\rho}} \right) \\ \mathbf{c}^{(m+1)} = (\mathbf{I}_J + \sigma_\rho^2 \Lambda^{-1})^{-1} \mathbf{U}^T \cdot (\tilde{\mathbf{y}} - \sigma_\rho \mathbf{b}^{L(m+1)}) \end{array} \right. \quad (6.22)$$

$\mathbf{I}_J$  est la matrice identité de dimension  $J \times J$ .

### 6.2.4 Conclusion

Dans cette section, on a rappelé deux approches utilisées pour minimiser des énergies non-quadratiques de la forme de  $\mathcal{J}_{RMV}$  : la théorie semi-quadratique et l'approche de Huber. Ces deux approches sont similaires et leur comparaison permet de préciser les fonctions des variables auxiliaires introduites dans la théorie semi-quadratique.

Deux algorithmes sont proposés : ARTUR (ou la minimisation avec des poids modifiés), et LEGEND (ou la minimisation avec des résidus modifiés). Ce dernier nécessite plus d'étapes pour converger vers le minimum recherché, mais il requiert beaucoup moins d'opérations que l'algorithme ARTUR. Il est finalement beaucoup plus rapide dans le cadre de notre problème. D'autres algorithmes [81, 143, 88], ont été proposés dans la littérature, spécifiquement adaptés à la fonction convexe  $\rho$  définie par Huber [62].

Le terme additif, relatif à l'hypothèse de distribution gaussienne de  $\mathcal{P}(\mathbf{c}|\mathcal{B})$ , est un terme quadratique. La minimisation de  $\mathcal{J}_{RMAP}$  ne pose alors pas de difficulté, et de manière similaire à  $\mathcal{J}_{RMV}$ , deux algorithmes itératifs sont proposés pour calculer  $\hat{\mathbf{c}}_{RMAP}$ .

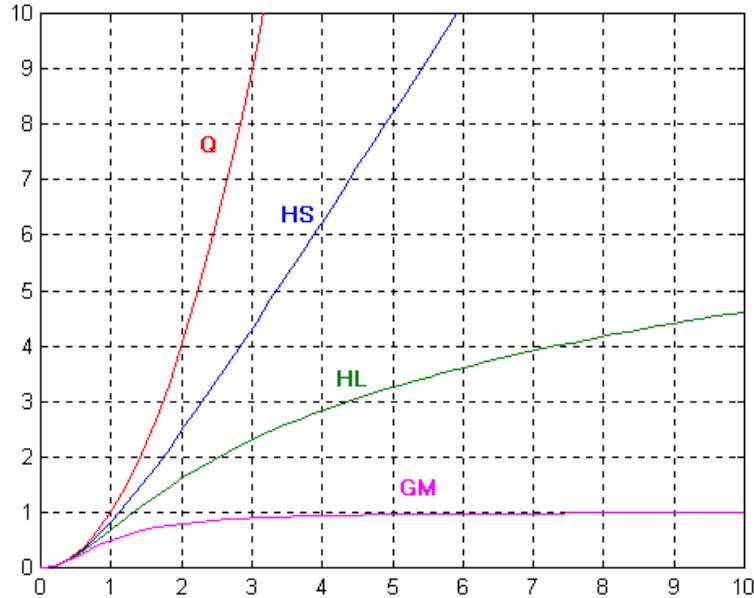
## 6.3 Paramètre d'échelle $\sigma_\rho$ et estimation en continuation

Le paragraphe 6.3.1 présente les fonctions robustes  $\rho$  utilisées dans cette thèse. Le paragraphe 6.3.2 présente les méthodes utilisées pour fixer le paramètre d'échelle  $\sigma_\rho$ . Afin d'éviter les possibles minima locaux des fonctions non convexes, nous proposons à la section 6.3.3 d'utiliser ces fonctions en continuation. Cette stratégie est similaire au GNC («graduated non convexity» ou non convexité graduelle), déjà utilisé pour la reconnaissance d'objets avec la fonction GM par Black et al. [13, 16]. Cependant, dans le GNC, la non-convexité est progressivement introduite en ajustant le paramètre d'échelle. A la première étape, l'énergie est rendue convexe sur le domaine de variation des résidus en choisissant une grande valeur pour  $\sigma$ . Cette valeur est liée au plus grand résidu attendu, qui doit être connu à l'avance. Dans notre approche, la première fonction utilisée est convexe indépendamment de la valeur des résidus.

### 6.3.1 Fonctions $\rho$

Trois fonctions  $\rho$  ont été utilisées pour l'estimation des coordonnées  $\mathbf{c}$ . La fonction HS est convexe et ne présente qu'un seul minimum. La fonction HL proposée par Hebert et Leahy, et la fonction GM proposée par Geman et McClure, sont non convexes mais atténuent plus fortement les outliers. Les expressions de ces différentes fonctions sont explicitées dans le tableau 6.2 et représentées à la figure 6.4 avec la fonction quadratique  $Q$ , définie par  $Q(t) = t^2$ . L'expression de la variable auxiliaire  $b^A$  est également précisée. Notons que d'autres fonctions  $\rho$  ont été proposées dans la littérature [62, 16, 25, 107, 59].

	$\rho(t)$	$b^A(t)$	convexité
HS	$2\sqrt{1+t^2} - 2$	$\frac{1}{\sqrt{1+t^2}}$	convexe
HL	$\ln(1+t^2)$	$\frac{1}{1+t^2}$	non convexe
GM	$\frac{t^2}{1+t^2}$	$\frac{1}{(1+t^2)^2}$	non convexe

TAB. 6.2: Expressions des fonctions robustes  $\rho$  utilisées.FIG. 6.4:Fonctions quadratique  $Q$  et robustes  $\rho$  : HS «surfaces minimales», HL (d'Hebert et Leahy) et GM (de Geman et Mc Clure).

### 6.3.2 Paramètre d'échelle $\sigma_\rho$

Plusieurs méthodes ont été proposées pour estimer  $\sigma_\rho$  : l'estimation jointe de  $(\mathbf{c}, \sigma_\rho)$  ou l'estimation préalable de  $\sigma_\rho$  qui ensuite reste fixe lors de l'estimation des composantes  $\mathbf{c}$ .

### 6.3.2.1 Estimation de $\sigma_\rho$ à partir de l'observation $\mathbf{y}$

**Estimation conjointe de  $(\mathbf{c}, \sigma_\rho)$ .** On considère que les résidus ont une densité de probabilité dépendant du paramètre  $\sigma_\rho$  [62] :

$$\mathcal{P}(\mathbf{w}, \sigma_\rho | \mathcal{B}) \propto \prod_{n=1}^N \frac{1}{\sigma_\rho} \exp \left[ -\frac{1}{2} \rho \left( \frac{\epsilon_n}{\sigma_\rho} \right) \right]$$

L'énergie à minimiser définie par  $\mathcal{J}_{BO}(\mathbf{c}, \sigma_\rho) = -\ln \mathcal{P}(\mathbf{w}, \sigma_\rho | \mathcal{B})$ , s'écrit [18] :

$$\mathcal{J}_{BO}(\mathbf{c}, \sigma_\rho) = \sum_{n=1}^N \left\{ \ln(\sigma_\rho) + \frac{1}{2} \rho \left( \frac{\epsilon_n}{\sigma_\rho} \right) \right\}$$

Cette énergie est approchée par l'énergie augmentée  $\mathcal{J}_{BO}^A$  (méthode ARTUR) définie par :

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

Cette énergie peut être minimisée itérativement en  $\mathbf{b}^A$ ,  $\mathbf{c}$  et  $\sigma_\rho$ . A  $\sigma_\rho$  fixé, minimiser  $\mathcal{J}_{BO}$  en  $\mathbf{c}$  est équivalent à diminuer  $\mathcal{J}_{RMV}$ . A  $\mathbf{c}$  fixé, la minimisation en  $\sigma_\rho$  de l'énergie augmentée  $\mathcal{J}_{BO}^A$  conduit à :

$$\sigma_\rho^2 = \frac{1}{N} \sum_{n=1}^N b_n^A \epsilon_n^2$$

Cette expression est proposée par Boyer et al. dans [18]. Elle est attractive car elle est directement liée à la densité de probabilité des résidus. Malheureusement,  $\mathcal{J}_{BO}$  n'est pas convexe en  $\sigma_\rho$ . De plus, Huber remarque que cet estimateur n'est généralement pas robuste. Aussi, propose-t-il une autre expression de l'énergie :

$$\mathcal{J}_{HU}(\mathbf{c}, \sigma_\rho) = \frac{1}{n} \sum_{n=1}^N \left[ \rho \left( \frac{\epsilon_n}{\sigma_\rho} \right) + a \right] \cdot \sigma_\rho \quad (6.23)$$

où  $\rho$  doit être convexe et  $a$  est un paramètre d'ajustement. Minimiser  $\mathcal{J}_{HU}$  en  $\mathbf{c}$  à  $\sigma_\rho$  fixé, est équivalent à minimiser  $\mathcal{J}_{RMV}$ . L'énergie  $\mathcal{J}_{HU}$  est convexe en  $(\mathbf{c}, \sigma)$  et a un seul minimum global (cf. annexe E). Huber propose de minimiser  $\mathcal{J}_{HU}$  de manière alternée en  $\mathbf{c}$  et en  $\sigma_\rho$ . A  $\mathbf{c}^{(m)}$  fixé,  $\sigma_\rho$  est estimé à l'étape  $(m+1)$  par :

$$\left( \sigma^{(m+1)} \right)^2 = \frac{1}{n \cdot a} \sum_{n=1}^N \chi(\epsilon_n^{(m)} / \sigma_\rho^{(m)}) \left( \sigma_\rho^{(m)} \right)^2 \quad (6.24)$$

où  $\chi(t) = t \cdot \rho'(t) - \rho(t)$ . La valeur de  $a$  est choisie pour que l'équation (6.24) corresponde à l'estimation classique de la variance dans le cas quadratique [62]. Cette étape est ajoutée aux algorithmes ARTUR [27, 28] ou LEGEND.

A  $\mathbf{c}^{(m)}$  fixé, Huber montre que l'étape d'estimation de  $\sigma_\rho$  fait décroître  $\mathcal{J}_{HU}$ . Lorsque  $\sigma_\rho^{(m)}$  est fixé, calculer  $\mathbf{c}^{(m)}$  fait aussi décroître  $\mathcal{J}_{HU}$ . Cette minimisation alternée n'est définie que pour les fonctions convexes. En pratique, les fonctions non convexes présentent un plus grand intérêt car elles rejettent mieux les *outliers*.

**Estimation de  $\sigma_\rho$  après une première estimation de  $\mathbf{c}$ .** Après une première étape d'estimation de  $\mathbf{c}$ , le paramètre d'échelle  $\sigma_\rho$  peut être estimé par le MAD (pour « median absolute deviation ») [56, 62] :

$$\sigma_\rho = 1,483 \operatorname{med}_n \{ |\epsilon_n - \operatorname{med}_{n'}(\epsilon_{n'})| \}$$

L'estimation de  $\mathbf{c}$  est poursuivie à paramètre d'échelle fixe. Hampel et al.[56] précisent que les algorithmes à paramètre d'échelle fixe sont plus performants.

### 6.3.2.2 Utilisation de l'apprentissage pour fixer le paramètre d'échelle *a priori*

Pour réduire le coût calculatoire, nous proposons d'estimer le paramètre d'échelle hors ligne en utilisant les images d'apprentissage [29]. Le paramètre d'échelle contrôle le point où l'influence des outliers commence à décroître. On rappelle que la distribution des erreurs dans le cadre de l'hypothèse robuste, peut être interprétée comme le mélange de lois [62, 56] :

$$\exp \left[ -\frac{1}{2} \rho(\epsilon) \right] \propto \alpha \mathcal{P}(\epsilon | \text{outliers}, \mathcal{B}) + (1 - \alpha) \mathcal{P}(\epsilon | \text{inliers}, \mathcal{B})$$

où  $\mathcal{P}(\epsilon | \text{inliers}, \mathcal{B})$  est la distribution des données «bonnes», supposée gaussienne centrée sur 0 [62, 56], et  $\mathcal{P}(\epsilon | \text{outliers}, \mathcal{B})$  est une distribution inconnue des erreurs grossières. Les résidus calculés sur les images d'apprentissage sont considérés comme des *inliers*. On propose donc de fixer le paramètre d'échelle en fonction de leur variance estimée grâce à nos images d'apprentissage.

**Apprentissage  $(\hat{\mu}^{\mathbf{c}}, \hat{\Sigma}^{\mathbf{c}})$ .** Pour chaque image  $\mathbf{x}^k$  de la base d'apprentissage, on calcule la variance :

$$(\sigma^k)^2 = \frac{1}{N} \sum_{n=1}^N (\epsilon_n^k)^2$$

On définit :

$$\sigma_{max}^{\mathbf{c}} = \max_{k \in [1 \dots K]} \{ \sigma^k \}$$

**Apprentissage  $(\hat{\mu}^{\mathbf{n}}, \hat{\Sigma}^{\mathbf{n}})$ .** De la même façon que pour les estimateurs  $\hat{\mu}^{\mathbf{n}}$  et  $\hat{\Sigma}^{\mathbf{n}}$ , on peut estimer la variance de l'erreur de reconstruction sur l'image  $\mathbf{x}^k$  en ne considérant que les pixels appartenant à l'objet :

$$(\sigma^k)^2 = \frac{1}{\sum_{n=1}^N b_n^k} \sum_{n=1}^N b_n^k (\epsilon_n^k)^2$$

On définit :

$$\sigma_{max}^{\mathbf{n}} = \max_{k \in [1 \dots K]} \{ \sigma^k \}$$

**Réglage du paramètre d'échelle.** Le paramètre d'échelle doit être choisi :

- suffisamment petit pour rejeter efficacement les outliers,
- suffisamment grand pour ne pas rejeter les inliers nécessaires à la reconnaissance des images de la base d'apprentissage.

Il est donc nécessaire de trouver le paramètre d'échelle qui réalise le meilleur compromis entre ces deux objectifs. Pour cela, on propose d'effectuer une étude pour chaque base considérée. Dans nos expériences, on l'a fixé en fonction de  $\sigma_{max}$  ou  $2\sigma_{max}$  suivant la base d'apprentissage. Cette remarque sera illustrée par les expériences présentées à la section 6.4.

En considérant que l'influence des outliers sur la régression commence à décroître au point d'inflexion des fonctions  $\rho$  non-convexe [15], les paramètres d'échelle sont choisis tels que  $\rho''(\frac{\sigma_{max}}{\sigma_\rho}) = 0$  ou  $\rho''(\frac{2\sigma_{max}}{\sigma_\rho}) = 0$ . La fonction convexe HS n'a pas de points d'inflexion et son paramètre d'échelle est fixé arbitrairement :  $\sigma_{HS} = \sigma_{max}$  ou  $\sigma_{HS} = 2\sigma_{max}$ . Les paramètres d'échelle utilisés sont précisés au tableau 6.3.

	$\sigma_\rho$	$\sigma_\rho$
$\sigma_{\rho_{HS}}$	$\sigma_{max}$	$2\sigma_{max}$
$\sigma_{\rho_{HL}}$	$\sigma_{max}$	$2\sigma_{max}$
$\sigma_{\rho_{GM}}$	$\sqrt{3}\sigma_{max}$	$2\sqrt{3}\sigma_{max}$

TAB. 6.3:Paramètres d'échelle des différentes fonctions  $\rho$  utilisées.

**Remarques.** Dans le cas des images multibandes, le résidu  $\epsilon_n$  sur le pixel est défini par  $\epsilon_n = \|w_n\|$ . A cause de cette définition, l'hypothèse de distribution gaussienne des résidus «inliers»,  $\mathcal{P}(\epsilon|\text{inliers}, \mathcal{B})$ , n'est pas vérifiée. En supposant que les  $P$  composantes du pixel  $w_n$  sont indépendantes et ont la même distribution gaussienne, alors le résidu  $\epsilon$  a une densité de probabilité de *chi* de degré  $P$  [110]. Pour les images en couleur ( $P = 3$ ), cette densité correspond à la distribution de Maxwell. Cependant, l'hypothèse simple de distribution gaussienne est conservée pour modéliser  $\mathcal{P}(\epsilon|\text{inliers}, \mathcal{B})$ , car on suppose que les erreurs grossières ou outliers, ont un résidu  $\epsilon$  élevé. La figure 6.5 présente la distribution de Maxwell des résidus «inliers»,  $\mathcal{P}(\epsilon|\text{inliers}, \mathcal{B})$ , approchée par une distribution gaussienne pour les résidus importants.

### 6.3.3 Estimation en continuation

Les fonctions convexes comme HS assurent l'unicité de la solution, mais leurs fonctions d'influence  $\rho'$  sont monotones. L'influence des *outliers* est alors limitée mais non nulle. De ce point de vue, les fonctions «hard redescender» [145] comme GM, sont plus intéressantes. Malheureusement, elles ne sont

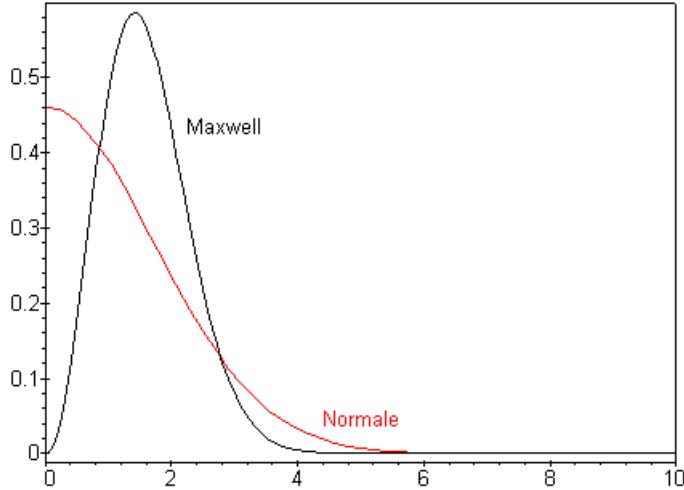


FIG. 6.5: Approximation de la distribution  $\mathcal{P}(\epsilon|\text{inliers}, \mathcal{B})$  par une gaussienne (cas des images en couleur).

pas convexes. Des algorithmes déterministes efficaces peuvent pourtant être définis par une approche graduelle de la non convexité. Nous proposons d'utiliser en continuation la fonction HS, puis la fonction non convexe HL à échelle fixée et enfin la fonction GM à échelle fixée.

**Système de reconnaissance avec le paramètre d'échelle fixé *a priori*.** La figure 6.6 présente l'estimation en continuation utilisant un paramètre d'échelle fixé *a priori* par l'apprentissage. L'estimée initiale  $c_0$  correspond aux estimateurs classiques définis aux équations 6.17 et 6.20. Puis l'énergie robuste  $\mathcal{J}_{RMV}$  (respectivement  $\mathcal{J}_{RMAP}$ ) est minimisée grâce aux algorithmes ARTUR ou LEGEND définis aux équations 6.18 et 6.19 (resp. 6.21 et 6.22).

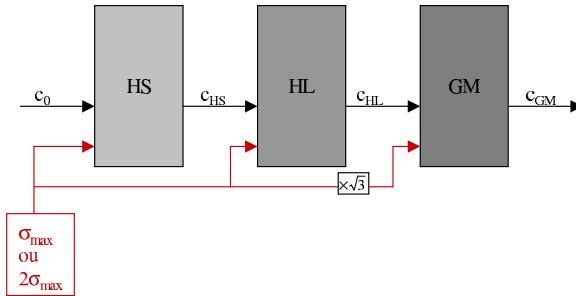


FIG. 6.6: Estimation en continuation avec le paramètre d'échelle fixé *a priori* grâce aux images l'apprentissage.

**Système de reconnaissance avec le paramètre d'échelle estimée conjointement à  $\mathbf{c}$ .** La figure 6.7 présente l'estimation en continuation utilisant un paramètre d'échelle estimé conjointement à  $\mathbf{c}$  pour la fonction convexe HS (cf. équation 6.24). Puis  $\sigma_{HS}$  reste fixe pour les fonctions non convexes car l'estimation conjointe de  $(\mathbf{c}, \sigma_\rho)$ , proposée par Huber, n'est définie que pour les fonctions convexes.

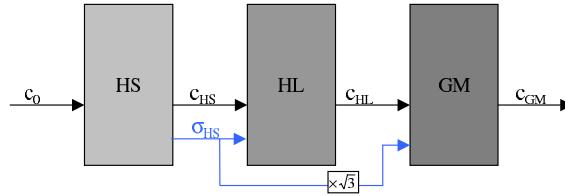


FIG. 6.7: Estimation en continuation avec le paramètre d'échelle estimé conjointement aux paramètres  $\mathbf{c}$  avec la fonction convexe HS.

## 6.4 Résultats de reconnaissance

La reconnaissance est réalisée en estimant les coordonnées  $\mathbf{c}$  sur l'espace propre  $F$  d'une observation  $\mathbf{y}$ . L'image d'apprentissage dont la projection dans  $F$  est la plus proche de cette estimée, identifie l'observation.

### 6.4.1 Images en niveaux de gris

La base d'images d'apprentissage est composée des 72 images du premier objet de la base COIL. Les images test utilisées présentent des occurrences des objets sur des fonds texturés ou uniformes différents de ceux présents dans les images d'apprentissage. La reconnaissance a été réalisée sur un espace propre  $F$  de dimension 10 et 30. La part d'inertie expliquée correspondante est reportée au tableau 6.4.

	$(\hat{\mu}^c, \hat{\Sigma}^c)$	$(\hat{\mu}^n, \hat{\Sigma}^c)$
J=10	$\mathcal{Q}_{10}^c = 93\%$	$\mathcal{Q}_{10}^n = 70\%$
J=30	$\mathcal{Q}_{30}^c = 99\%$	$\mathcal{Q}_{30}^n = 90\%$

TAB. 6.4: Part d'inertie expliquée pour la base d'apprentissage COIL, suivant les deux méthodes d'apprentissage.

### 6.4.1.1 Paramètre d'échelle fixé a priori

Dans cette expérience, le paramètre d'échelle est fixé *a priori* grâce à l'apprentissage (cf. figure 6.6). La première expérience de reconnaissance est réalisée sur les images d'apprentissage avec des paramètres d'échelle  $\sigma_\rho$  calculés à partir de  $\sigma_{max}$ . La reconnaissance est correcte si l'image reconnue correspond exactement à celle utilisée pour créer l'image test.

**Reconnaissance des images d'apprentissage.** Le tableau 6.5 présente les résultats de reconnaissance obtenus sur les 72 images de la base d'apprentissage COIL, qui représente un même objet variant suivant un paramètre de rotation. Une seule image n'est pas exactement reconnue par le système par les estimateurs  $\hat{\mathbf{c}}_{RMV}$  et  $\hat{\mathbf{c}}_{RMAP}$  dans le cas de l'apprentissage  $(\hat{\mu}^n, \hat{\Sigma}^c)$ , sur un espace  $F$  de dimension  $J = 10$ . Dans ce cas, l'image identifiée par le système ne diffère que d'un pas de rotation par rapport au bon résultat. A cette exception, le choix du paramètre d'échelle fixé à partir de  $\sigma_{max}$  paraît donc justifié.

Il pourrait sembler anormal que le choix du nouvel estimateur de la moyenne dans l'apprentissage entraîne des mauvais résultats de reconnaissance sur les images d'apprentissage sous l'hypothèse d'un bruit gaussien. En fait, les images observées sont centrées par  $\tilde{\mathbf{y}} = \mathbf{y} - \hat{\mathbf{\mu}}$  quel que soit l'estimateur  $\hat{\mathbf{\mu}}$  utilisé en apprentissage. Dans le cas de l'apprentissage  $(\hat{\mu}^n, \hat{\Sigma}^c)$ , si l'observation  $\mathbf{y}$  est l'échantillon d'apprentissage  $\mathbf{x}^k$ , alors l'image centrée  $\mathbf{y} - \hat{\mathbf{\mu}}^n$  ne correspond pas à l'image centrée d'apprentissage  $\tilde{\mathbf{x}}^k = \mathbf{B}^k (\mathbf{x}^k - \hat{\mathbf{\mu}}^n)$ . Dans ce cas, centrer les données sur  $\hat{\mathbf{\mu}}^n$ , introduit des outliers qui ne sont pas bien pris en compte par l'estimation classique. Ceci explique la médiocre qualité des résultats dans ce cas (cf. tableau 6.5). Les méthodes robustes sont elles, naturellement adaptées à cette situation et corrigeant parfaitement ce qui pourrait apparaître comme un défaut de la nouvelle méthode d'apprentissage. Les expériences suivantes montrent la supériorité du nouvel apprentissage  $(\hat{\mu}^n, \hat{\Sigma}^c)$  associé aux méthodes d'estimation robuste, en présence de fonds texturés et d'occultations partielles.

	$(\hat{\mu}^c, \hat{\Sigma}^c)$		$(\hat{\mu}^n, \hat{\Sigma}^c)$	
bruit gaussien	$\hat{\mathbf{c}}_{MV}$	$\hat{\mathbf{c}}_{MAP}$	$\hat{\mathbf{c}}_{MV}$	$\hat{\mathbf{c}}_{MAP}$
$J = 10$	100%	100%	16,7% (12/72)	16,7% (12/72)
$J = 30$	100%	100%	44% (32/72)	44% (32/72)
hypothèse robuste	$\hat{\mathbf{c}}_{RMV}$	$\hat{\mathbf{c}}_{RMAP}$	$\hat{\mathbf{c}}_{RMV}$	$\hat{\mathbf{c}}_{RMAP}$
$J = 10$	100%	100%	98,6% (71/72)	98,6% (71/72)
$J = 30$	100%	100%	100%	100%

TAB. 6.5: Résultats de reconnaissance sur les images d'apprentissage.

**Fonds texturés ou de niveaux de gris différents des images d'apprentissage.** Ce test est réalisé sur 288 images présentant des occurrences de nos objets d'intérêt sur des fonds uniformes ou texturés

différents de ceux des images d'apprentissage. La figure 6.8 propose un exemple d'observation  $\mathbf{y}$ , ainsi que les variables auxiliaires obtenues lors de l'estimation robuste de  $\hat{\mathbf{c}}_{RMV}$  dans le cas de l'apprentissage par  $(\hat{\mu}^{\mathbf{c}}, \hat{\Sigma}^{\mathbf{c}})$ . La variable auxiliaire  $b^A$  définie sur chaque pixel a une valeur proche de 1 quand le résidu est faible (pixel blanc), et tend vers 0 pour les forts résidus (pixels noirs). La variable  $b^L$  est proche du pixel de l'observation lorsque celui est classé «outlier» (ce qui est le cas de la majeure partie du fond), et proche de 0 lorsqu'il est classé «inlier» (ce qui est le cas de la majeure partie de l'objet).

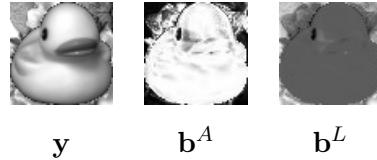


FIG. 6.8:*Exemple des variables auxiliaires utilisées dans les algorithmes de minimisation :  $J = 10$ , apprentissage  $(\hat{\mu}^{\mathbf{c}}, \hat{\Sigma}^{\mathbf{c}})$ .*

Le tableau 6.6 présente les résultats de reconnaissance obtenus. Quel que soit l'estimateur considéré, la reconnaissance avec un apprentissage  $(\hat{\mu}^{\mathbf{n}}, \hat{\Sigma}^{\mathbf{c}})$  obtient de meilleurs résultats que l'apprentissage classique avec  $(\hat{\mu}^{\mathbf{c}}, \hat{\Sigma}^{\mathbf{c}})$ . De plus, les estimateurs qui supposent une distribution gaussienne du bruit  $\mathbf{w}$ , ont des résultats de reconnaissance très médiocres. Au contraire, les estimateurs robustes atteignent 99% d'images correctement reconnues dans ce test pour l'apprentissage  $(\hat{\mu}^{\mathbf{n}}, \hat{\Sigma}^{\mathbf{c}})$ , pour un espace propre de dimension  $J = 30$ .

	$(\hat{\mu}^{\mathbf{c}}, \hat{\Sigma}^{\mathbf{c}})$		$(\hat{\mu}^{\mathbf{n}}, \hat{\Sigma}^{\mathbf{c}})$	
bruit gaussien	$\hat{\mathbf{c}}_{MV}$	$\hat{\mathbf{c}}_{MAP}$	$\hat{\mathbf{c}}_{MV}$	$\hat{\mathbf{c}}_{MAP}$
	20% (58/288)	20% (58/288)	53,4% (154/288)	53,4% (154/288)
$J = 30$	39,6% (114/288)	39,6% (114/288)	86,8% (250/288)	86,8% (250/288)
	$\hat{\mathbf{c}}_{RMV}$	$\hat{\mathbf{c}}_{RMAP}$	$\hat{\mathbf{c}}_{RMV}$	$\hat{\mathbf{c}}_{RMAP}$
hypothèse robuste	55% (159/288)	56% (162/288)	93,7% (270/288)	94% (271/288)
	70,1% (202/288)	79,6% (228/288)	99,6% (287/288)	99,6% (287/288)

TAB. 6.6:*Résultats de reconnaissance sur des images présentant des fonds variés.*

**Occultations partielles.** Cette expérience est réalisée sur 72 images présentant des occurrences des objets d'apprentissage partiellement occultés, sur des fonds texturés. Ces images sont présentées à l'annexe G. Le tableau 6.7 présente les résultats de reconnaissance obtenus. On remarque que dans l'hypothèse d'un bruit gaussien, l'apprentissage  $(\hat{\mu}^{\mathbf{n}}, \hat{\Sigma}^{\mathbf{c}})$  permet d'obtenir de meilleurs résultats de reconnaissance que l'apprentissage classique  $(\hat{\mu}^{\mathbf{c}}, \hat{\Sigma}^{\mathbf{c}})$ , aussi bien sur un espace de dimension 10 que 30. L'estimation robuste améliore systématiquement les résultats des estimateurs classiques.

	$(\hat{\mu}^c, \hat{\Sigma}^c)$		$(\hat{\mu}^n, \hat{\Sigma}^c)$	
bruit gaussien	$\hat{c}_{MV}$	$\hat{c}_{MAP}$	$\hat{c}_{MV}$	$\hat{c}_{MAP}$
	25% (18/72)	25% (18/72)	41,7% (30/72)	41,7% (30/72)
	54,2% (39/72)	54,2% (39/72)	70,08% (51/72)	72,2% (52/72)
hypothèse robuste	$\hat{c}_{RMV}$	$\hat{c}_{RMAP}$	$\hat{c}_{RMV}$	$\hat{c}_{RMAP}$
	52,7% (38/72)	56,9% (41/72)	86,1% (62/72)	84,7% (61/72)
	81,9% (59/72)	88,9% (64/72)	100% (72/72)	100% (72/72)

TAB. 6.7: Résultats de reconnaissance sur des occurrences partiellement occultées.

Les 10 mauvaises reconnaissances de  $(\hat{\mu}^n, \hat{\Sigma}^c)$ , obtenues dans l'hypothèse robuste sur la distribution du bruit, sur un espace propre de dimension 10, sont des erreurs à un pas de rotation près : les solutions proposées par notre système de reconnaissance ont une apparence très proche de la réponse correcte. Par contre plusieurs erreurs de reconnaissance obtenues avec l'apprentissage  $(\hat{\mu}^c, \hat{\Sigma}^c)$ , sont loin de la bonne solution. Ceci est illustré par la figure 6.9 : les observations  $y^k$  présentent des occurrences dégradées des objets des images d'apprentissage  $x^k$ . Les deux premiers exemples sont des occurrences dégradées de  $x^1$  et  $x^{14}$ , qui ne sont pas reconnues avec un apprentissage  $(\hat{\mu}^c, \hat{\Sigma}^c)$ , et dont les solutions proposées par le système,  $x^{48}$  et  $x^{56}$ , sont loin de la bonne réponse. Le dernier exemple montre une occurrence de  $x^8$  pas exactement reconnue avec un apprentissage  $(\hat{\mu}^n, \hat{\Sigma}^c)$ , mais la solution proposée,  $x^9$ , est proche de la bonne réponse.

**Conclusion sur l'expérience de reconnaissance sur des images en niveaux de gris.** Le tableau 6.8 présente les résultats globaux de reconnaissance obtenus sur l'ensemble des 432 images de ce test. L'estimation classique, avec une hypothèse de bruit gaussien, donne de meilleurs résultats lorsque l'apprentissage par ACP est réalisé avec le nouvel estimateur de la moyenne  $\hat{\mu}^n$ . Quel que soit l'apprentissage utilisé, l'estimation robuste améliore les résultats de reconnaissance.

Les 432 images de ce test représentent des objets de la base d'apprentissage dégradés par un fond texturé, et des occultations. Cependant, les parties non occultées sont identiques à celle des images d'apprentissage. Un dernier test a été réalisé sur les mêmes 432 images avec un bruit additif gaussien, de moyenne nulle et d'écart type  $\sigma = 20$ . Cela correspond à un rapport signal à bruit de 22 dB. Les résultats de reconnaissance obtenus restent identiques à ceux du tableau 6.8.

L'utilisation de l'estimateur  $\hat{c}_{RMAP}$  n'améliore que peu les résultats de reconnaissance obtenus avec  $\hat{c}_{RMV}$ . L'estimateur  $\hat{c}_{RMAP}$  utilise un *a priori* de distribution gaussienne sur le paramètre  $c$  à estimer. Rappelons cependant que, comme on l'a vu à la figure 5.15, cette hypothèse n'est pas vérifiée dans le cas de la base COIL.

L'apprentissage classique de l'ACP,  $(\hat{\mu}^c, \hat{\Sigma}^c)$ , obtient de moins bons résultats de reconnaissance que

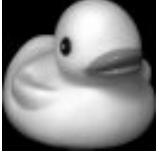
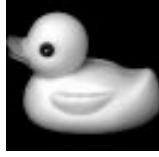
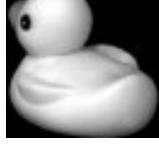
observations $\mathbf{y}$	$(\hat{\mu}^c, \hat{\Sigma}^c)$	$(\hat{\mu}^n, \hat{\Sigma}^c)$
		
$\mathbf{y}^1$	$\mathbf{x}^{48}$	$\mathbf{x}^1$
		
$\mathbf{y}^{14}$	$\mathbf{x}^{56}$	$\mathbf{x}^{14}$
		
$\mathbf{y}^8$	$\mathbf{x}^8$	$\mathbf{x}^9$

FIG. 6.9: Exemple de résultats de reconnaissance : l'observation  $\mathbf{y}$  est identifiée par l'image d'apprentissage  $\mathbf{x}$ , suivant les deux méthodes d'apprentissage avec les estimateurs robustes ( $J = 10$ ).

	$(\hat{\mu}^c, \hat{\Sigma}^c)$		$(\hat{\mu}^n, \hat{\Sigma}^c)$	
bruit gaussien $J = 10$	$\hat{\mathbf{c}}_{MV}$	$\hat{\mathbf{c}}_{MAP}$	$\hat{\mathbf{c}}_{MV}$	$\hat{\mathbf{c}}_{MAP}$
	34,3%	34,3%	45,4%	45,4%
	52,1%	52,1%	77,1%	77,3%
hypothèse robuste $J = 10$	$\hat{\mathbf{c}}_{RMV}$	$\hat{\mathbf{c}}_{RMAP}$	$\hat{\mathbf{c}}_{RMV}$	$\hat{\mathbf{c}}_{RMAP}$
	62,3%	63,7%	93,3%	93,3%
$J = 30$	77,1%	84,3%	99,8%	99,8%

TAB. 6.8: Résultats de reconnaissance sur l'ensemble des 432 images testées.

l'apprentissage proposé au chapitre 5, utilisant le nouvel estimateur de la moyenne,  $(\hat{\mu}^n, \hat{\Sigma}^c)$ . Cette remarque se vérifie quel que soit l'estimateur utilisé, et quelle que soit l'hypothèse sur la distribution du bruit, gaussien ou robuste. Par ailleurs, les résultats de reconnaissance sont naturellement améliorés en augmentant la dimension de l'espace propre.

#### 6.4.1.2 Estimation conjointe ( $\mathbf{c}, \sigma_\rho$ ).

L'estimation conjointe (cf. figure 6.7), a été testée avec un espace propre  $F$  de dimension 10, sur les images présentant des occultations, avec l'apprentissage  $(\hat{\mu}^n, \hat{\Sigma}^c)$ . Les résultats de reconnaissance sont identiques à ceux obtenus avec le paramètre d'échelle fixé grâce à l'apprentissage. Cependant, si l'on compare les estimées après la fonction convexe HS<sup>1</sup> (cf. tableau 6.9), on remarque que l'estimée  $\hat{\mathbf{c}}_{HS}$  calculées conjointement à  $\sigma_\rho$ , est plus éloignée dans le plan propre  $F$  de la bonne réponse que son estimée calculée avec un paramètre d'échelle fixé *a priori*. On remarque également que le paramètre d'échelle  $\sigma_{HS}$  estimé conjointement à  $\mathbf{c}$  est plus grand que  $\sigma_{max}$  fixé *a priori* : l'estimation du paramètre d'échelle semble donc perturbée par les outliers.

estimée $\sigma_{HS}$ (estimation conjointe de $\sigma_{HS}$ et de $\hat{\mathbf{c}}_{HS}$ )	52	65	58
distance de l'estimée $\hat{\mathbf{c}}_{HS}$ à la solution (estimation conjointe de $\sigma_{HS}$ et de $\hat{\mathbf{c}}_{HS}$ )	404 544	462 092	1 100 252
distance de l'estimée $\hat{\mathbf{c}}_{HS}$ à la solution ( $\sigma_{HS}$ fixé <i>a priori</i> $\sigma_{HS} = \sigma_{max} = 25$ )	299 452	270 777	690 334

TAB. 6.9: Comparaison des paramètres  $\hat{\mathbf{c}}$  estimés conjointement avec le paramètre d'échelle et avec ceux obtenus avec un paramètre d'échelle fixé *a priori*.

D'autre part, l'estimation conjointe de  $(\hat{\mathbf{c}}, \sigma_{HS})$ , a un coût calculatoire plus important. Aussi cette méthode d'estimation n'a pas été utilisée par la suite.

#### 6.4.2 Images en couleurs

On considère la base d'apprentissage A43 d'images en couleur, présentée à l'annexe H.1. On rappelle que cette base contient 43 panneaux triangulaires appris avec leurs rotations dans le plan de l'image.

<sup>1</sup>On rappelle que les deux systèmes de reconnaissance proposés à la section 6.3.3, diffèrent essentiellement sur la première étape d'estimation avec la fonction convexe HS.

### 6.4.2.1 Résultats de reconnaissance

**Choix du paramètre d'échelle.** La reconnaissance est effectuée sur un espace propre  $F$  à 60 dimensions. Une première expérience est réalisée sur les images d'apprentissage à paramètre d'échelle fixé à la valeur de  $\sigma_{max}$ . Ce paramètre d'échelle n'est pas très satisfaisant car les pixels du pictogramme sont parfois interprétés comme des outliers. L'estimation robuste détériore alors l'estimation classique jusqu'à induire de mauvaises reconnaissances. La figure 6.10 propose un résultat de mauvaise reconnaissance sur une image d'apprentissage : les pondérations calculées dans l'estimation robuste éliminent les pixels relatifs au pictogramme du panneau. Ces pixels ont une erreur de reconstruction telle que  $\|w_n\| > \sigma_{max}$  et sont alors considérés comme des outliers. Contrairement à l'expérience précédente sur les images à niveaux de gris, les pixels dont l'erreur de reconstruction est supérieure à  $\sigma_{max}$  sont nécessaires pour distinguer les panneaux entre eux et permettre une bonne reconnaissance. Par contre en fixant le paramètre d'échelle à partir de  $2\sigma_{max}$ , on obtient bien 100% de reconnaissance sur la base d'apprentissage.



image d'apprentissage  $x$    pondération  $b^A$    panneau reconnu

FIG. 6.10: Reconnaissance robuste sur une image d'apprentissage à paramètre d'échelle fixé  $\sigma_{max}$ .

**Effets de fonds texturés.** Le test a été réalisé sur 387 images de panneaux triangulaires présentant des fonds différents de ceux des images d'apprentissage, uniformes ou texturées. Un groupe de 43 images présente également des déteriorations relatives au triangle rouge entourant le pictogramme (cf. figure 6.11(a)). Quelques images sont représentées à la figure 6.11.

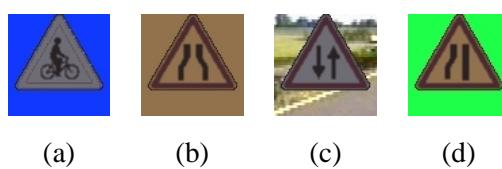


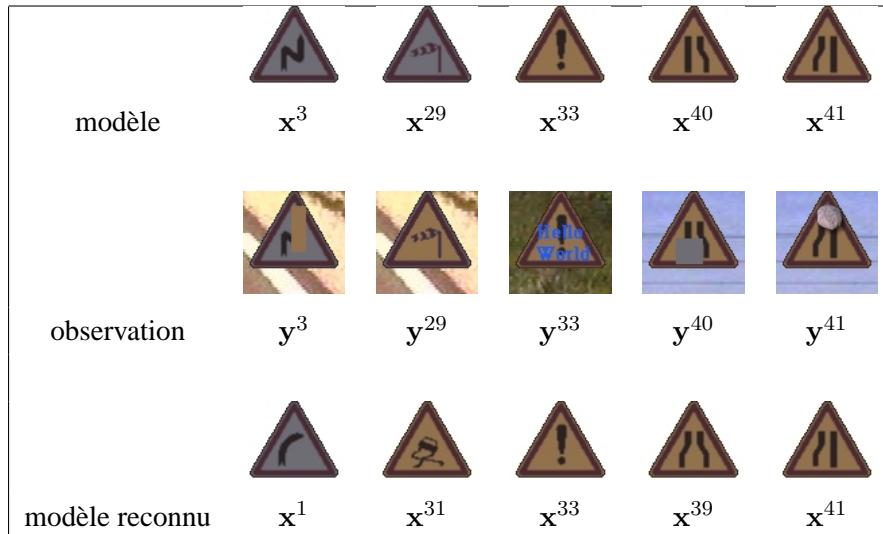
FIG. 6.11: Quelques images testées.

Le tableau 6.10 présente les résultats de reconnaissance obtenus pour nos différents estimateurs suivant les deux apprentissages  $(\hat{\mu}^c, \hat{\Sigma}^c)$  et  $(\hat{\mu}^n, \hat{\Sigma}^c)$ . On peut remarquer les excellents résultats obtenus par l'apprentissage réalisé avec le nouvel estimateur de la moyenne  $\hat{\mu}^n$  pour les estimateurs non robustes  $\hat{c}_{MAP}$  et  $\hat{c}_{MV}$  : 98% des observations sont parfaitement reconnues avec l'apprentissage  $(\hat{\mu}^n, \hat{\Sigma}^c)$ , contre seulement 27% avec l'apprentissage classique  $(\hat{\mu}^c, \hat{\Sigma}^c)$ .

	$(\hat{\mu}^c, \hat{\Sigma}^c)$		$(\hat{\mu}^n, \hat{\Sigma}^c)$	
bruit gaussien $J = 60$	$\hat{c}_{MV}$	$\hat{c}_{MAP}$	$\hat{c}_{MV}$	$\hat{c}_{MAP}$
	27,4%	27,1%	98,4%	99,4%
hypothèse robuste $J = 60$	$\hat{c}_{RMV}$	$\hat{c}_{RMAP}$	$\hat{c}_{RMV}$	$\hat{c}_{RMAP}$
	95%	95,9%	100%	100%

TAB. 6.10: Résultats de reconnaissance sur des images présentant des fonds texturés.

**Occultations.** Les 43 images présentant des occurrences des images d'apprentissage occultées, sont présentées à l'annexe H.2. Les perturbations considérées affectent la partie informative du panneau : son pictogramme est partiellement caché ou effacé. Quelques-unes des observations sont représentées à la figure 6.12, avec le modèle d'apprentissage auquel elles correspondent.

FIG. 6.12: Quelques observations présentant des occurrences fortement dégradées de nos objets d'intérêt et les modèles reconnus par l'estimateur  $\hat{c}_{RMV}$  avec l'apprentissage  $(\hat{\mu}^n, \hat{\Sigma}^c)$ .

L'observation  $y^{29}$  est un piège pour le système car si le pictogramme n'est pas changé par rapport à  $x^{29}$ , la couleur dominante blanche du panneau a par contre été remplacée par la couleur jaune, qui est aussi présente dans la base d'apprentissage. L'observation  $y^{29}$  ne diffère du panneau jaune  $x^{31}$  que par le pictogramme, défini par beaucoup moins de pixels que le fond du panneau. C'est pourquoi les pixels du pictogramme ont été éliminés dans l'estimation et les pixels jaunes ont été conservés. Le même résultat a été obtenu par les estimateurs  $\hat{c}_{RMV}$  et  $\hat{c}_{RMAP}$  pour les deux méthodes d'apprentissage  $(\hat{\mu}^c, \hat{\Sigma}^c)$  et  $(\hat{\mu}^n, \hat{\Sigma}^c)$ . La figure 6.13 présente pour chaque estimateur, le masque de pondération obtenu : les pixels sombres (resp. clairs) correspondent à des pondérations faibles (resp. fortes).

On remarque que l'observation  $y_{40}$  n'est pas reconnue car la zone occultée est nécessaire à sa parfaite

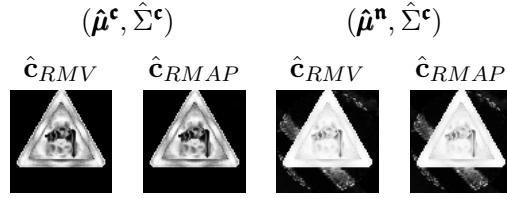


FIG. 6.13: *Masques des pondérations  $\mathbf{b}^A$  estimés pour l'observation  $\mathbf{y}^{29}$  ( $J = 60$ ).*

identification. Dans cet exemple, un opérateur humain est également incapable de l'identifier entre les modèles d'apprentissage  $\mathbf{x}^{40}$  et  $\mathbf{x}^{39}$ .

Quelle que soit l'observation  $\mathbf{y}^k$  de notre test, l'angle de rotation des panneaux occultés est bien estimé, c'est-à-dire que le panneau reconnu, a la même orientation que le modèle considéré. Le tableau 6.11 présente les résultats de reconnaissance obtenus sur les 43 panneaux fortement dégradés, pour des espaces propres  $F$  de dimension  $J = 60$  et  $J = 100$ . L'augmentation de la dimension de l'espace propre permet d'améliorer les résultats.

	$(\hat{\mu}^c, \hat{\Sigma}^c)$		$(\hat{\mu}^n, \hat{\Sigma}^c)$	
bruit gaussien	$\hat{c}_{MV}$	$\hat{c}_{MAP}$	$\hat{c}_{MV}$	$\hat{c}_{MAP}$
$J = 60$	6/43	6/43	23/43	23/43
$J = 100$	8/43	9/43	25/43	26/43
hypothèse robuste	$\hat{c}_{RMV}$	$\hat{c}_{RMAP}$	$\hat{c}_{RMV}$	$\hat{c}_{RMAP}$
$J = 60$	25/43	26/43	26/43	28/43
$J = 100$	30/43	30/43	30/43	29/43

TAB. 6.11: *Résultats de reconnaissance sur les images fortement dégradés de l'annexe H avec un espace propre  $F$  de dimension  $J = 60$  et  $J = 100$ .*

**Importance de la couleur.** La figure 6.14 présente un résultat de reconnaissance robuste obtenu sur deux images d'un même panneau partiellement occulté. La surface d'occultation est la même dans les deux images mais sa couleur est différente. On remarque que la région occultante noire, au contraire de la bleue, induit le système de reconnaissance en erreur car elle crée une ambiguïté avec un autre modèle de la base d'apprentissage. Cet exemple montre que l'information couleur est prise en considération avec profit dans la reconnaissance.

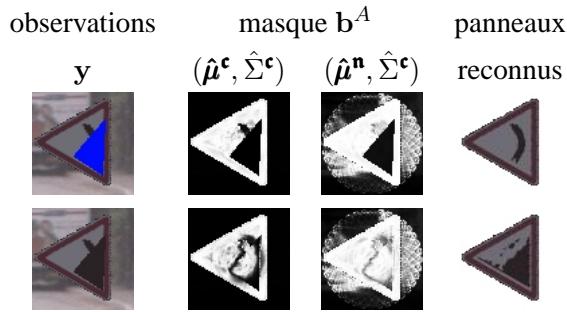


FIG. 6.14:Importance de la couleur pour la reconnaissance.

#### 6.4.2.2 Conclusion sur l'expérience de reconnaissance en couleur

Sur une base d'images d'apprentissage en couleur, on a montré que les méthodes de reconnaissance robustes proposées, obtiennent des résultats de reconnaissance nettement supérieurs aux méthodes classiques. L'utilisation d'un *a priori* de distribution gaussienne sur les coordonnées  $c$  dans le cadre de l'estimation bayésienne, ne permet pas d'améliorer les résultats de reconnaissance obtenus par l'estimateur du maximum de vraisemblance. En effet, comme on l'a remarqué à la section 5.2.2.2, la base d'apprentissage A43, n'a pas de distribution gaussienne dans le plan propre  $F$ .

## 6.5 Conclusion

La représentation d'un ensemble d'images par espace propre permet d'indexer chacune d'elles par un vecteur de dimension plus réduite que les images originales. Ce représentant est classiquement calculé par une projection orthogonale sur l'espace propre. Cependant, cette projection n'est pas robuste lorsqu'il s'agit de l'estimer à partir d'une observation fortement perturbée par du bruit structuré. L'utilisation d'une hypothèse sur la distribution du bruit sur l'observation, mieux adaptée à cette situation, permet d'estimer la projection de manière robuste, c'est-à-dire peu sensible à la présence de données erronées.

L'hypothèse robuste a été développée dans le cadre de deux approches statistiques d'estimation, l'estimation classique par maximum de vraisemblance, et l'estimation bayésienne par le maximum a posteriori. La première approche considère que la projection à estimer est un paramètre déterministe. L'estimateur de maximum de vraisemblance ne maximise que la probabilité de l'erreur entre l'observation et sa projection sur l'espace propre. L'approche bayésienne, considère au contraire, que le représentant de l'observation dans l'espace propre, est une variable aléatoire. Cette formulation permet d'introduire dans l'estimation un *a priori* sur la distribution du représentant dans l'espace propre. Cette dernière peut être modélisée grâce au images d'apprentissage. On a considéré dans cette thèse, deux cas simples : la distribution uniforme et gaussienne des projections des images d'apprentissage dans l'espace propre.

Quelle que soit l'approche adoptée, l'utilisation d'une hypothèse robuste sur la distribution de l'er-

reur, conduit à la minimisation d'une énergie non quadratique. Deux algorithmes itératifs, ARTUR et LEGEND, ont été introduits. Ils sont issus de deux approches différentes, la théorie semi-quadratique, et les statistiques robustes. La comparaison entre ces deux approches a permis de préciser, dans le cas de l'algorithme LEGEND, la formulation de la théorie semi-quadratique. Cette minimisation dépend d'un paramètre d'échelle, qui permet de contrôler la sensibilité de l'estimateur à la présence de données erronées. Nous avons proposé d'estimer ce paramètre d'échelle grâce aux images d'apprentissage, ce qui conduit à des méthodes de reconnaissance complètement non supervisée. L'ensemble de ces techniques robustes de reconnaissance sur un espace propre, a été étendu aux images multibandes.

Les méthodes de reconnaissance robuste proposées, ont été testées sur deux ensembles d'apprentissage, le premier comportant 72 images en niveaux de gris représentant un objet variant suivant l'angle de vue, le second contenant 1548 images couleurs des panneaux triangulaires variant suivant leur rotation dans le plan de l'image. Les méthodes robustes de reconnaissance ont clairement montré leur supériorité face à la méthode classique d'estimation. Cependant, la formulation bayésienne, ajoutant un *a priori* sur la solution recherchée, n'a pas permis d'améliorer encore ces résultats. En effet, l'hypothèse simple de distribution gaussienne des projections dans l'espace propre, n'est pas vérifiée pour nos ensembles d'apprentissage. D'autres hypothèses plus complexes et mieux adaptées, pourront être envisagées entraînant toutefois une augmentation significative de la complexité algorithmique.



# Chapitre 7

## Détection robuste

Ce chapitre propose de nouvelles méthodes de détection d'une classe d'objets basées sur une représentation globale de l'apparence par espace propre.

Les techniques proposées s'inscrivent dans le cadre difficile où l'on dispose d'une hypothèse unique, c'est-à-dire la classe d'intérêt  $\mathcal{B}$  est parfaitement définie, tandis les autres classes, notamment  $\bar{\mathcal{B}}$ , ne sont pas modélisées de par leur trop grande variabilité [46]. Typiquement, les détecteurs reposant sur une hypothèse unique proposent de mesurer la ressemblance, notée  $S_{ij}(\mathbf{y}|\mathcal{B})$  entre l'observation  $\mathbf{y}$  extraite de la fenêtre centrée sur la position  $(i, j)$ , et celles préalablement apprises,  $\mathcal{B} = \{\mathbf{x}^k\}_{k \in \{1, \dots, K\}}$  (cf. figure 7.1). Toutes les valeurs  $S_{ij}$  sont rassemblées dans une carte. Les positions  $(i, j)$  dont les valeurs de ressemblance sont supérieures à un certain seuil permettent de localiser les objets d'intérêt.

Le paragraphe 7.1 rappelle les mesures de ressemblance classiquement utilisées pour la détection. Celles-ci supposent que l'observation est une instance d'un des objets de la base, et utilisent la vraisemblance  $\mathcal{P}(\mathbf{y}|\mathcal{B})$  comme mesure de ressemblance.

Ces mesures de ressemblance ne sont pas robustes à la présence de bruit structuré sur l'observation, induite par des effets spéculaires locaux, ou des occultations partielles. De manière analogue à la reconnaissance robuste étudiée au chapitre 6, on propose un nouveau modèle où l'on suppose explicitement que l'observation  $\mathbf{y}$  est une occurrence bruitée d'une image d'apprentissage  $\mathbf{x}$ , soit  $\mathbf{y} = \mathbf{x} + \mathbf{w}^o$ . Le bruit d'observation  $\mathbf{w}^o$  correspond alors à des perturbations sur l'observation telles que les occultations

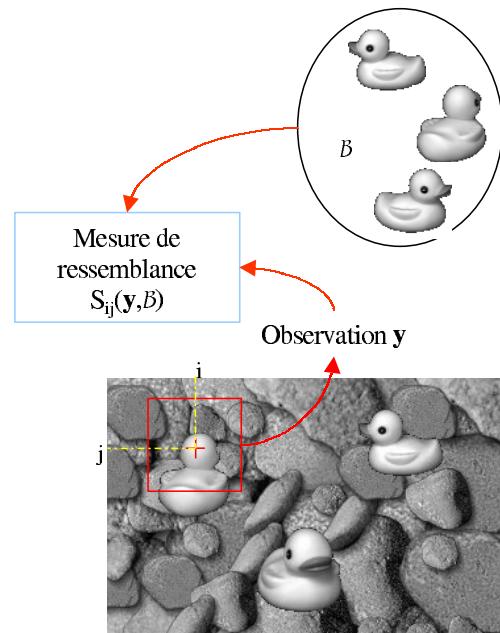


FIG. 7.1:Détection.

partielles, bruit d’acquisition, etc. Nous proposons à la section 7.2, des détecteurs robustes basés sur cette nouvelle approche [31].

Les méthodes, classiques et robustes, sont comparées à la section 7.3 pour la détection de l’objet de la base COIL dans des images en niveaux de gris, puis sur la détection des panneaux triangulaires de la base A43, dans des images en couleur.

## 7.1 Méthodes classiques de détection

**Modèle classique.** Le modèle classique considère que l’observation  $\mathbf{y}$  est une réalisation d’une image d’apprentissage soit  $\mathbf{y} = \mathbf{x}$ . Aussi, un critère simple de ressemblance consiste à estimer la probabilité que l’observation  $\mathbf{y}$  soit une instance d’un des objets de la base  $\mathcal{B}$ , notée  $\mathcal{P}(\mathbf{y}|\mathcal{B})$ . Celle-ci est alors calculée grâce aux vraisemblances proposées au chapitre 5 (paragraphe 5.2). On rappelle que, dans ce cas, l’observation  $\mathbf{y}$  est liée à ses coordonnées  $\mathbf{c}$  par la relation :

$$\tilde{\mathbf{y}} = \mathbf{U}\mathbf{c} + \mathbf{w}^r$$

où  $\mathbf{w}^r$  est l’erreur de reconstruction dans  $F^\perp$ . Cette section rappelle deux détecteurs classiques basés sur une représentation d’une classe d’objets par espace propre [111, 97].

### 7.1.1 Distance From Feature Space

La DFFS ou Distance From Feature Space est définie comme la distance euclidienne entre l’observation  $\mathbf{y}$  et l’espace propre  $F$  [111] :

$$DFFS(\mathbf{y}) = \|\tilde{\mathbf{y}} - \mathbf{U}\mathbf{U}^T\tilde{\mathbf{y}}\|^2 \quad (7.1)$$

Cette mesure est liée à la vraisemblance proposée au chapitre 5 (section 5.2), lorsqu’on suppose que les coordonnées  $\mathbf{c} = \mathbf{U}^T\tilde{\mathbf{y}}$  sont distribuées uniformément dans l’espace propre  $F$ , et que leurs erreurs de reconstruction ont une distribution gaussienne de moyenne nulle, indépendante identiquement distribuée dans  $F^\perp$ , ou :

$$\mathcal{P}(\mathbf{y}|\mathcal{B}) \propto \frac{\exp\left[\frac{-\|\mathbf{w}^r\|^2}{2\sigma^2}\right]}{(\sqrt{2\pi}\sigma)^{NP-J}} \quad (7.2)$$

avec  $\mathbf{w}^r = \tilde{\mathbf{y}} - \mathbf{U}\mathbf{U}^T\tilde{\mathbf{y}}$ . On a l’équivalence :  $DFFS(\mathbf{y}) = -\log \mathcal{P}(\mathbf{y}|\mathcal{B}) + cste$ .

### 7.1.2 Vraisemblance de Moghaddam et Pentland

La vraisemblance proposée par Moghaddam et Pentland [97] suppose que les images d'apprentissage ont une distribution gaussienne :

$$\mathcal{P}(\mathbf{y}|\mathcal{B}) = \frac{\exp\left[-\frac{1}{2}\tilde{\mathbf{y}}^T\hat{\Sigma}^{-1}\tilde{\mathbf{y}}\right]}{(2\pi)^{\frac{NP}{2}}|\det(\hat{\Sigma})|^{\frac{1}{2}}}$$

Pour réduire le temps de calcul, cette probabilité est approchée par (cf. paragraphe 5.2) :

$$\hat{\mathcal{P}}(\mathbf{y}|\mathcal{B}) = \frac{\exp\left[-\sum_{j=1}^J \frac{c_j^2}{2\lambda_j}\right]}{\prod_{j=1}^J \sqrt{2\pi\lambda_j}} \cdot \frac{\exp\left[-\frac{\|\mathbf{w}^r\|^2}{2\sigma^2}\right]}{(\sqrt{2\pi}\sigma)^{NP-J}} \quad (7.3)$$

avec

$$\begin{cases} \mathbf{c} = \mathbf{U}^T\tilde{\mathbf{y}} \\ \mathbf{w}^r = \tilde{\mathbf{y}} - \mathbf{U}\mathbf{U}^T\tilde{\mathbf{y}} \end{cases}$$

### 7.1.3 Remarques

La distance de Mahalanobis définie par  $d^2(\mathbf{y}) = \tilde{\mathbf{y}}^T\hat{\Sigma}^{-1}\tilde{\mathbf{y}}$ , est directement liée à la vraisemblance de Moghaddam et Pentland :  $\hat{\mathcal{P}}(\mathbf{y}|\mathcal{B}) \simeq \mathcal{P}(\mathbf{y}|\mathcal{B}) \propto \exp[-d^2(\mathbf{y})]$ . Or, Fukunaga remarque que cette distance a une erreur de classification qui augmente en fonction de la dimension  $NP$  des vecteurs considérés [46]. Ainsi lorsque l'on considère des variables de grande dimension  $NP$ , seuiller la distance de Mahalanobis (ou la vraisemblance de Moghaddam et Pentland) pour détecter une classe d'objets, peut provoquer de nombreuses fausses alarmes.

Une méthode pour minimiser les erreurs de détection consiste à classer les valeurs de distance  $d_{ij}^2$  calculées en chaque position  $(i, j)$  de l'image à analyser, et à ne retenir que la valeur la plus petite. Cette procédure de classement réduit sensiblement les erreurs de détection [46]. Celle-ci est utilisée par Pentland et al. [111], pour évaluer les performances de ses détecteurs dans le cadre de leur application. Sachant qu'il n'y a qu'une seule occurrence de l'objet recherché dans chaque image de test, ils ne considèrent que la position présentant la meilleure valeur de ressemblance de chacune des cartes calculées sur une série d'images tests. Ainsi, pour un seuil  $s$  :

- si la plus grande mesure de ressemblance de la carte est supérieure à ce seuil  $s$  et que sa position  $(i, j)$  est dans le voisinage (5 pixels) de l'objet recherché, alors elle est comptabilisée comme une bonne détection.
- Par contre, si la plus grande mesure de ressemblance de la carte, est supérieure à  $s$  mais que sa position  $(i, j)$  n'est pas dans le voisinage à 5 pixels près de l'objet recherché, alors elle est comptabilisée comme une fausse alarme.
- Si la plus grande mesure de ressemblance est inférieure au seuil  $s$ , alors elle n'est pas prise en compte.

Cette méthode permet de calculer une courbe COR ou caractéristique opérationnelle du récepteur, paramétrée par le seuil  $s$ , qui présente le pourcentage de bonnes détections en fonction du pourcentage de fausses alarmes. Cependant, considérer uniquement la position de plus forte ressemblance, implique que l'on ne recherche dans l'image qu'une seule occurrence des objets d'intérêt, ce qui n'est pas le cas dans notre application. On propose à l'annexe I une autre méthode pour calculer les courbes COR dans la situation plus difficile où l'on ne connaît pas le nombre d'occurrences des objets à détecter dans chaque image de test.

## 7.2 Nouvelles méthodes de détection

Les méthodes classiques nécessitent une segmentation préalable et ne tolèrent pas la présence de bruit structuré induit par des perturbations locales telles que les occultations partielles ou les effets spéculaires. Afin de pallier ces limitations, nous proposons un nouveau modèle liant l'observation  $\mathbf{y}$ , aux images d'apprentissage  $\mathbf{x}$ .

**Nouveau modèle.** De manière analogue à la reconnaissance robuste étudiée au chapitre 6, on suppose que l'observation  $\mathbf{y}$  est une occurrence bruitée d'une image d'apprentissage  $\mathbf{x}$ , soit  $\mathbf{y} = \mathbf{x} + \mathbf{w}^o$  où  $\mathbf{w}^o$  est le bruit d'observation qui correspond à toutes les perturbations sur l'observation : occultations partielles, bruit d'acquisition, etc. Ainsi, l'image centrée  $\tilde{\mathbf{y}}$  est liée à ses paramètres réduits  $\mathbf{c}$  par la relation :

$$\tilde{\mathbf{y}} = \mathbf{U}\mathbf{c} + \mathbf{w}$$

où  $\mathbf{w}$  est la somme de l'erreur de reconstruction  $\mathbf{w}^r$  et du bruit d'observation  $\mathbf{w}^o$ . La vraisemblance conjointe qui lie l'observation  $\mathbf{y}$  à ses coordonnées  $\mathbf{c}$ ,  $\mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$ , s'écrit :

$$\mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B}) = \mathcal{P}(\mathbf{w}|\mathcal{B}) \mathcal{P}(\mathbf{c}|\mathcal{B})$$

Lorsque l'on dispose d'expressions analytiques pour les distributions du bruit  $\mathcal{P}(\mathbf{w}|\mathcal{B})$  et des coordonnées  $\mathcal{P}(\mathbf{c}|\mathcal{B})$ , alors la vraisemblance  $\mathcal{P}(\mathbf{y}|\mathcal{B})$ , peut être calculée en intégrant  $\mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$  par rapport à  $\mathbf{c}$ , pour servir de détecteur. Cependant ce calcul peut être complexe. Aussi, proposons-nous d'utiliser comme mesure de ressemblance :

$$\mathcal{P}(\mathbf{y}, \hat{\mathbf{c}}|\mathcal{B}) = \max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$$

On note  $\hat{\mathbf{c}} = \arg \max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$  et  $\hat{\mathbf{w}} = \tilde{\mathbf{y}} - \mathbf{U}\hat{\mathbf{c}}$ , alors :

$$\max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B}) = \mathcal{P}(\hat{\mathbf{w}}|\mathcal{B}) \mathcal{P}(\hat{\mathbf{c}}|\mathcal{B})$$

On considère deux hypothèses sur la distribution du bruit, ainsi que deux hypothèses sur la distribution des composantes  $\mathbf{c}$ .

**Distribution du bruit  $\mathcal{P}(\mathbf{w}|\mathcal{B})$ .**

- *Hypothèse gaussienne.* On suppose que l'erreur a une distribution gaussienne de moyenne nulle, indépendante et identiquement distribuée, notée  $\mathcal{N}(\mathbf{0}, \sigma_g^2 \mathbf{I}_{NP})$ , telle que :

$$\mathcal{P}(\mathbf{w}|\mathcal{B}) \propto \exp\left[-\frac{\|\mathbf{w}\|^2}{2\sigma_g^2}\right]$$

- *Hypothèse robuste.* L'hypothèse d'un bruit  $\mathbf{w}$  gaussien n'est pas satisfaisante lorsque l'on souhaite reconnaître des objets apparaissant sur un fond texturé, ou partiellement occultés. Pour prendre en compte l'apparition éventuelle d'erreurs grossières (*outliers*), nous modélisons la distribution de l'erreur par :

$$\mathcal{P}(\mathbf{w}|\mathcal{B}) \propto \prod_{n=1}^N \exp\left[-\frac{1}{2} \rho\left(\frac{\|w_n\|}{\sigma_\rho}\right)\right]$$

**Distribution des coordonnées  $\mathbf{c}$ .** Deux hypothèses ont été considérées pour la distribution de  $\mathcal{P}(\mathbf{c}|\mathcal{B})$  :

- *Hypothèse uniforme.* On suppose la distribution uniforme de  $\mathbf{c}$  dans l'espace propre :

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

- *Hypothèse gaussienne.* La variable aléatoire  $\mathbf{c}$  de dimension  $J$  suit la loi normale de moyenne nulle et de matrice de covariance  $\Lambda$ , notée  $\mathcal{N}(\mathbf{0}, \Lambda)$ . La matrice de covariance  $\Lambda$  est la matrice diagonale des  $J$  premières valeurs propres  $\lambda_j$  issues de l'apprentissage par ACP, et la densité de probabilité de  $\mathbf{c}$  s'écrit :

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

Suivant ces hypothèses, quatre détecteurs sont définis dans les paragraphes suivants. Comme on l'a vu au chapitre 6, la maximisation de la vraisemblance conjointe  $\mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$ , sous l'hypothèse robuste de distribution du bruit, met en œuvre des techniques de régression introduisant des variables auxiliaires. Nous proposons à la section 7.2.5, un dernier détecteur robuste utilisant la pondération  $\mathbf{b}^A$  introduite dans l'algorithme ARTUR.

### 7.2.1 Hypothèses : distribution uniforme des coordonnées $\mathbf{c}$ et distribution gaussienne du bruit $\mathbf{w}$

On suppose que les coordonnées  $\mathbf{c}$  ont une distribution uniforme dans le plan propre, soit  $\mathcal{P}(\mathbf{c}|\mathcal{B}) = cste$ , et que le bruit  $\mathbf{w}$  a une distribution gaussienne indépendante identiquement distribuée  $\mathcal{N}(\mathbf{0}, \sigma_g^2 \mathbf{I}_{NP})$ . Alors, la distribution conjointe de l'observation  $\mathbf{y}$  et de ses coordonnées  $\mathbf{c}$  s'écrit :

$$\mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B}) \propto \exp\left[-\frac{\|\mathbf{w}\|^2}{2\sigma_g^2}\right]$$

L'estimée en  $\mathbf{c}$  qui maximise cette probabilité correspond à  $\hat{\mathbf{c}}_{MV} = \mathbf{U}^T \tilde{\mathbf{y}}$  (cf. sections 6.1.1.1 et 6.1.2.1). Dans ce cas, notre détecteur,  $\max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c} | \mathcal{B})$ , s'écrit :

$$\begin{aligned} \max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c} | \mathcal{B}) &\propto \exp - \left[ \frac{\|\tilde{\mathbf{y}} - \mathbf{U} \hat{\mathbf{c}}_{MV}\|^2}{2\sigma_g^2} \right] \\ &\propto \exp - \left[ \frac{\|\tilde{\mathbf{y}} - \mathbf{U} \mathbf{U}^T \tilde{\mathbf{y}}\|^2}{2\sigma_g^2} \right] \\ &\propto \exp - \left[ \frac{DFFS(\mathbf{y})}{2\sigma_g^2} \right] \end{aligned} \quad (7.4)$$

Dans le cadre des hypothèses de distribution uniforme des coordonnées  $\mathbf{c}$ , et de distribution gaussienne  $\mathcal{N}(\mathbf{0}, \sigma_g^2 \mathbf{I}_{NP})$  du bruit  $\mathbf{w}$ , alors notre détecteur,  $\max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c} | \mathcal{B})$ , est équivalent au détecteur classique DFFS.

### 7.2.2 Hypothèses : distribution uniforme des coordonnées $\mathbf{c}$ et distribution robuste du bruit $\mathbf{w}$

Les coordonnées  $\mathbf{c}$  ont une distribution uniforme dans le plan propre, et le bruit  $\mathbf{w}$  a une distribution qui permet l'apparition d'erreurs grossières. Alors la vraisemblance conjointe de l'observation  $\mathbf{y}$  et de ses coordonnées  $\mathbf{c}$  s'écrit :

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

L'estimée en  $\mathbf{c}$  qui maximise cette probabilité correspond à  $\hat{\mathbf{c}}_{RMV}$  (cf. paragraphes 6.1.1.2 et 6.1.2.2). Dans ce cas, notre détecteur,  $\max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c} | \mathcal{B})$ , dépend du terme :

$$RDFFS(\mathbf{y}) = \sum_{n=1}^N \rho \left( \frac{\|\hat{w}_n\|}{\sigma_\rho} \right)$$

avec  $\hat{\mathbf{w}} = \tilde{\mathbf{y}} - \mathbf{U} \hat{\mathbf{c}}_{RMV}$ . Ce détecteur peut être perçu comme une distance robuste à l'espace propre.

### 7.2.3 Hypothèses : distribution gaussienne des coordonnées $\mathbf{c}$ et distribution gaussienne du bruit $\mathbf{w}$

Le vecteur  $\mathbf{c}$  a une distribution gaussienne :  $\mathcal{N}(\mathbf{0}, \Lambda)$  dans le plan propre. Le bruit  $\mathbf{w}$  a une distribution gaussienne :  $\mathcal{N}(\mathbf{0}, \sigma_g^2 \mathbf{I}_{NP})$ . Alors, notre détecteur est calculé par :

$$\max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c} | \mathcal{B}) = \max_{\mathbf{c}} \left\{ \frac{\exp \left[ -\sum_{j=1}^J \frac{c_j^2}{2\lambda_j} \right]}{\prod_{j=1}^J \sqrt{2\pi \lambda_j}} \cdot \frac{\exp \left[ \frac{-\|\mathbf{w}\|^2}{2\sigma_g^2} \right]}{(\sqrt{2\pi} \sigma_g)^{NP}} \right\}$$

Dans ce cas, l'estimée qui maximise la vraisemblance  $\mathcal{P}(\mathbf{y}, \mathbf{c} | \mathcal{B})$  est  $\hat{\mathbf{c}}_{MAP}$ , présentée au paragraphe 6.1.2.3.

**Remarques.** Dans le cadre de ces hypothèses simples, la densité de probabilité  $\mathcal{P}(\mathbf{y}|\mathcal{B})$  peut être calculée en intégrant  $\mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$  en  $\mathbf{c}$  [153] :

$$\mathcal{P}(\mathbf{y}|\mathcal{B}) = \int_{-\infty}^{+\infty} \mathcal{P}(\mathbf{c}|\mathcal{B}) \mathcal{P}(\mathbf{w}|\mathcal{B}) d\mathbf{c}$$

L'observation centrée  $\tilde{\mathbf{y}}$  est liée aux variables aléatoires  $\mathbf{c}$  et  $\mathbf{w}$  par la relation  $\tilde{\mathbf{y}} = \mathbf{U}\mathbf{c} + \mathbf{w}$ . Les coordonnées  $\mathbf{c}$  et le bruit  $\mathbf{w}$  sont indépendants et ont respectivement les distributions gaussiennes  $\mathcal{N}(\mathbf{0}, \Lambda)$  et  $\mathcal{N}(\mathbf{0}, \sigma_g^2 \mathbf{I}_{NP})$ . De manière similaire à [153], on peut montrer que la distribution de  $\tilde{\mathbf{y}}$  est une distribution gaussienne centrée de matrice de covariance  $\Sigma_{\mathbf{y}} = \mathbf{U} \Lambda \mathbf{U}^T + \sigma_g^2 \mathbf{I}_{NP}$  :

$$\mathcal{P}(\mathbf{y}|\mathcal{B}) = \frac{\exp\left[-\frac{1}{2} \tilde{\mathbf{y}}^T \Sigma_{\mathbf{y}}^{-1} \tilde{\mathbf{y}}\right]}{(2\pi)^{\frac{NP}{2}} |\det(\Sigma_{\mathbf{y}})|^{\frac{1}{2}}}$$

La matrice de covariance  $\Sigma_{\mathbf{y}}$  se diagonalise grâce à la matrice des vecteurs propres issus de l'apprentissage,  $\mathbf{U}_{NP}$  :

$$\mathbf{U}_{NP}^T \Sigma_{\mathbf{y}} \mathbf{U}_{NP} = \begin{bmatrix} \lambda_1 + \sigma_g^2 & & & (0) \\ & \ddots & & \\ & & \lambda_J + \sigma_g^2 & \\ & & & \sigma_g^2 \\ (0) & & & \ddots \\ & & & \sigma_g^2 \end{bmatrix}$$

Cette vraisemblance gaussienne  $\mathcal{P}(\mathbf{y}|\mathcal{B})$  a une forme similaire à celle proposée par Moghaddam et Pentland  $\hat{\mathcal{P}}(\mathbf{y}|\mathcal{B})$  [97].

#### 7.2.4 Hypothèses : distribution gaussienne des coordonnées $\mathbf{c}$ et distribution robuste du bruit $\mathbf{w}$

Le vecteur aléatoire  $\mathbf{c}$  a une distribution gaussienne  $\mathcal{N}(\mathbf{0}, \Lambda)$  dans le plan propre, et le bruit  $\mathbf{w}$  a une distribution robuste qui permet l'apparition d'erreurs grossières. Alors la vraisemblance conjointe de l'observation  $\mathbf{y}$  et de ses coordonnées  $\mathbf{c}$  s'écrit :

$$\mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B}) \propto \exp\left[-\frac{1}{2} \sum_{n=1}^N \rho\left(\frac{\|w_n\|}{\sigma_\rho}\right)\right] \frac{\exp\left[-\sum_{j=1}^J \frac{c_j^2}{2\lambda_j}\right]}{\prod_{j=1}^J \sqrt{2\pi \lambda_j}}$$

L'estimée en  $\mathbf{c}$  qui maximise cette probabilité correspond à  $\hat{\mathbf{c}}_{RMAP}$  (cf. paragraphe 6.1.2.4). Pour simplifier les notations, le détecteur défini  $\mathcal{P}(\mathbf{y}, \hat{\mathbf{c}}_{RMAP}|\mathcal{B}) = \max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$ , est appelé RG.

#### 7.2.5 Utilisation des pondérations

On a vu au chapitre 6, que la maximisation de la vraisemblance conjointe  $\mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$  dans le cadre de l'hypothèse robuste de distribution du bruit  $\mathbf{w}$ , nécessite l'introduction d'une variable auxiliaire. La

pondération  $b_n^A$ , introduite dans l'algorithme ARTUR, est proche de 1 lorsque l'erreur  $\|w_n\|$  sur le pixel  $n$ , est petite, et tend vers 0 pour les forts résidus. Ainsi, l'image  $\mathbf{b}^A$  peut être interprétée comme une carte de confiance aux données. Comme second détecteur, nous proposons d'utiliser sa norme comme une mesure de similarité :

$$\|\mathbf{b}^A\|^2 = \sum_{n=1}^N (b_n^A)^2$$

La variable  $\mathbf{b}^A$  est calculée en même temps que les détecteurs RDFFS, pour le calcul de  $\hat{\mathbf{c}}_{RMV}$ , et RG, pour le calcul de  $\hat{\mathbf{c}}_{RMAP}$ . On distingue les deux cas, en spécifiant  $\|\mathbf{b}^A\|_G$  lorsque la pondération est issue du calcul du détecteur RG.

### 7.2.6 Remarques

L'hypothèse de distribution gaussienne du bruit  $\mathbf{w}$  permet de retrouver des expressions similaires aux détecteurs classiques : la DFFS lorsque les coordonnées  $\mathbf{c}$  ont une distribution uniforme dans l'espace propre  $F$ , et la vraisemblance de Moghaddam et Pentland lorsque les coordonnées  $\mathbf{c}$  ont une densité de probabilité gaussienne.

L'hypothèse gaussienne pour la distribution du bruit  $\mathbf{w}$  n'est pas satisfaisante lorsque l'on recherche les occurrences fortement dégradées des objets d'intérêt. Aussi, avec une nouvelle hypothèse sur la distribution de l'erreur permettant l'apparition d'erreurs grossières dans l'observation, on a proposé trois nouveaux détecteurs robustes :

- RDFFS basé sur une hypothèse de distribution uniforme de  $\mathbf{c}$  dans le plan propre, et une hypothèse robuste sur la distribution du bruit  $\mathbf{w}$ ,
- RG qui suppose la distribution gaussienne de  $\mathbf{c}$  dans le plan propre, et une hypothèse robuste sur la distribution du bruit  $\mathbf{w}$ ,
- la norme des pondérations,  $\|\mathbf{b}^A\|$  et  $\|\mathbf{b}^A\|_G$ , calculée en même temps que les détecteurs RDFFS et RG.

Le tableau 7.1 rappelle les détecteurs classiques et robustes. Ceux-ci sont comparés expérimentalement à la section suivante.

	distribution gaussienne de $\mathbf{w}$	distribution robuste de $\mathbf{w}$
$\mathcal{P}(\mathbf{c} \mathcal{B})$ uniforme	DFFS	RDFFS et $\ \mathbf{b}^A\ $
$\mathcal{P}(\mathbf{c} \mathcal{B})$ gaussienne	Vraisemblance de Moghaddam et Pentland	RG et $\ \mathbf{b}^A\ _G$

TAB. 7.1:Détecteurs classiques et robustes.

## 7.3 Expériences de détection

Les expériences de détection sont réalisées sur la base COIL des images en niveaux de gris, et la base A43 d'images en couleur. Les performances des détecteurs sont évaluées par des courbes COR qui présentent le pourcentage de bonnes détections en fonction du pourcentage de fausses alarmes. Notre méthode de calcul des courbes COR est présentée en annexe I et diffère de celle utilisée par Pentland et al. [111, 97] (cf. section 7.1.3). En effet, notre procédure ne fait aucune hypothèse quant au nombre d'objets présents dans une image.

De manière identique aux expériences de reconnaissance du chapitre 6 (cf. section 6.4), le paramètre d'échelle  $\sigma_\rho$  des détecteurs robustes est estimé *a priori* grâce aux images d'apprentissage.

### 7.3.1 Images en niveaux de gris : base COIL

On considère la base d'apprentissage COIL. Les images de ce test sont présentées à l'annexe J. La figure 7.2 présente les cartes de ressemblance obtenues sur les images  $I_{10}$  et  $I_{15}$  par les détecteurs robustes, et non robustes. Le contour bleu des cartes indique les positions où l'on ne peut pas extraire un échantillon y complet. Les pixels sombres indiquent la présence éventuelle d'un objet. On remarque, notamment pour l'image  $I_{10}$ , que les détecteurs robustes permettent de mieux localiser les occurrences des objets d'intérêt.

#### 7.3.1.1 DéTECTEURS CLASSIQUES

La figure 7.3 présente les courbes COR obtenues avec les détecteurs classiques DFFS et la Vraisemblance de Moghaddam et Pentland sur un espace propre  $F$  de dimension  $J = 10$ . Les résultats de détection sont légèrement améliorés par l'apprentissage réalisé avec le nouvel estimateur de la moyenne  $(\hat{\mu}^n, \Sigma^n)$ . On remarque également que les performances des deux détecteurs classiques, DFFS et la vraisemblance de Moghaddam et al., sont similaires. On rappelle que la vraisemblance de Moghaddam et Pentland utilise, en plus de la DFFS, un *a priori* de distribution gaussienne sur les coordonnées  $\mathbf{c}$  dans l'espace propre  $F$ . Celui-ci n'est pas vérifié pour notre base d'apprentissage (cf. section 5.2.2.2).

Les performances de ces détecteurs classiques ne sont pas très satisfaisantes car il faut accepter de nombreuses fausses alarmes pour que nos objets d'intérêt soient détectés. Ainsi, la courbe COR nous indique que pour détecter 80 % des objets d'intérêt dans les images de test, il faut accepter 10 % de fausses alarmes.

#### 7.3.1.2 DéTECTEURS ROBUSTES

**Effet du choix de l'estimateur de la moyenne dans l'apprentissage.** La figure 7.4 présente les courbes COR obtenues pour un apprentissage réalisé avec les estimateurs classiques  $(\hat{\mu}^c, \hat{\Sigma}^c)$  et avec

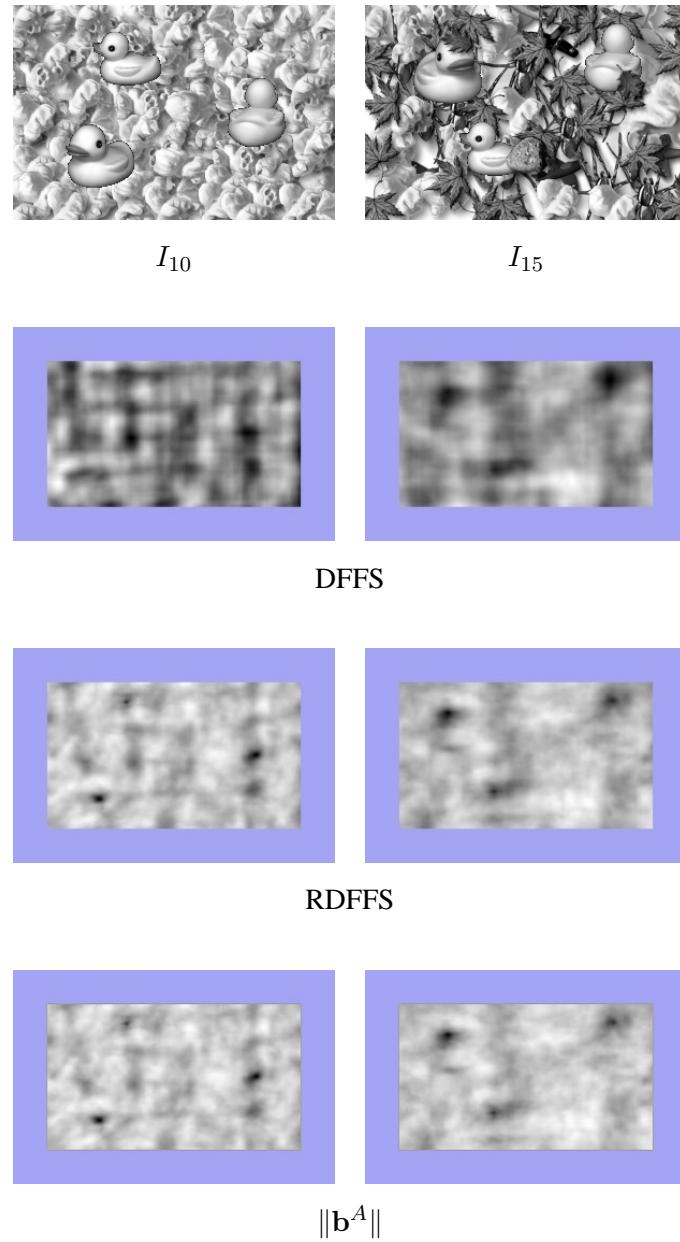


FIG. 7.2: Exemple de cartes de détection DFFS, RDFFS et  $\|\mathbf{b}^A\|$  obtenues sur les images  $I_{10}$  et  $I_{15}$  pour un espace propre  $F$  de dimension  $J = 10$ .

le nouvel estimateur de la moyenne  $(\hat{\mu}^n, \hat{\Sigma}^c)$ . Les cartes de ressemblance robustes ont été calculées par l'algorithme en continuation présenté au paragraphe 6.3 jusqu'à la fonction de Geman et Mc Clure  $\rho_{GM}$ .

On remarque que l'apprentissage avec l'estimateur  $\hat{\mu}^n$  de la moyenne permet d'obtenir de meilleurs résultats de détection. Aussi, celui-ci est utilisé systématiquement dans les expériences suivantes. Les performances de détection obtenues par les détecteurs RDFFS et RG (respectivement  $\|\mathbf{b}^A\|$  et  $\|\mathbf{b}^A\|_G$ ),

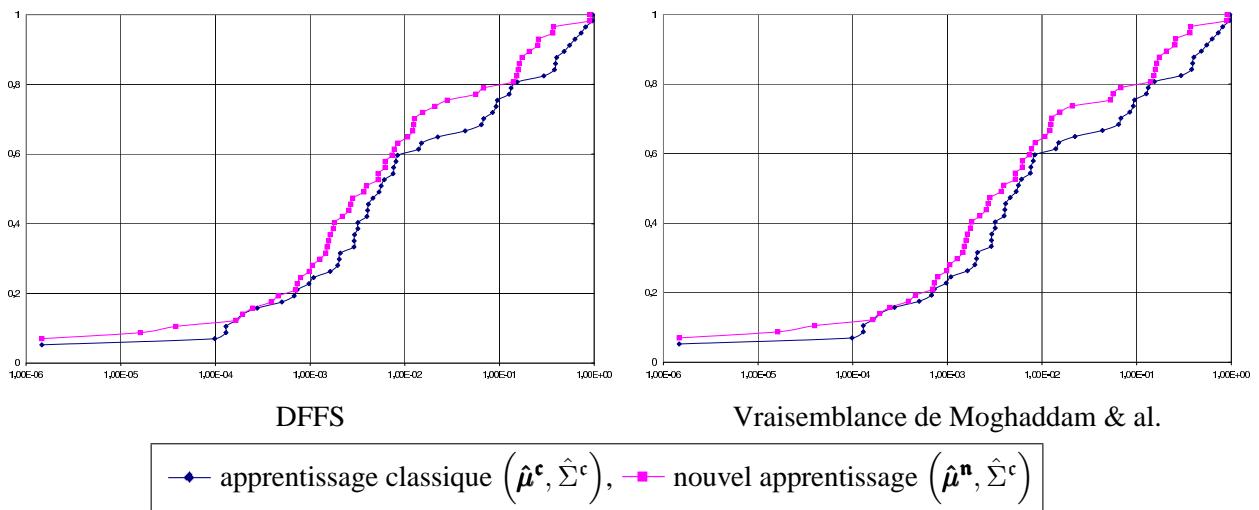


FIG. 7.3: Courbes COR calculées avec les détecteurs classiques pour un espace propre  $F$  de dimension  $J = 10$ .

sont sensiblement équivalents.

L'utilisation d'une hypothèse *a priori* de distribution gaussienne des coordonnées  $\mathbf{c}$  dans l'espace propre n'est pas adaptée à la classe d'objets considérée. De fait on constate qu'elle n'apporte pas d'améliorations pour la détection, aussi bien pour les détecteurs classiques (DFFS et vraisemblance de Moghaddam et Pentland), et les détecteurs robustes (RDFFS et RG,  $\|\mathbf{b}^A\|$  et  $\|\mathbf{b}^A\|_G$ ).

**Comparaison entre les détecteurs RDFFS et  $\|\mathbf{b}^A\|$ .** A la figure 7.4, on remarque également que les performances du détecteur  $\|\mathbf{b}^A\|$  surpassent celles de la RDFFS. Ainsi pour le nouvel apprentissage  $(\hat{\boldsymbol{\mu}}^n, \hat{\Sigma}^c)$ , la RDFFS détecte 80% des objets pour 0,15% de fausses alarmes, tandis que la norme des pondérations,  $\|\mathbf{b}^A\|$ , réalise seulement 0,04% de fausses alarmes pour le même pourcentage d'objets détectés.

**Influence du choix de la fonction  $\rho$ .** Les fonctions robustes HS, HL et GM, sont utilisées en continuation. La figure 7.5 présente les courbes COR obtenues par ces différentes fonctions utilisées successivement. Les résultats de détection s'améliorent en passant d'une fonction convexe (HS), peu pénalisante pour les outliers, à des fonctions non-convexes qui rejettent plus durement les données aberrantes (HL puis GM).

**Influence du choix de la dimension de l'espace d'apprentissage.** La figure 7.6 présente trois courbes COR obtenues avec la RDFFS pour un espace propre  $F$  de dimension 5, 10 et 30. L'amélioration des résultats de détection lorsque la dimension de  $F$  augmente est significative. On a déjà remarqué à la section 6.4.1, que l'augmentation de la dimension de l'espace propre permettait d'obtenir de meilleurs

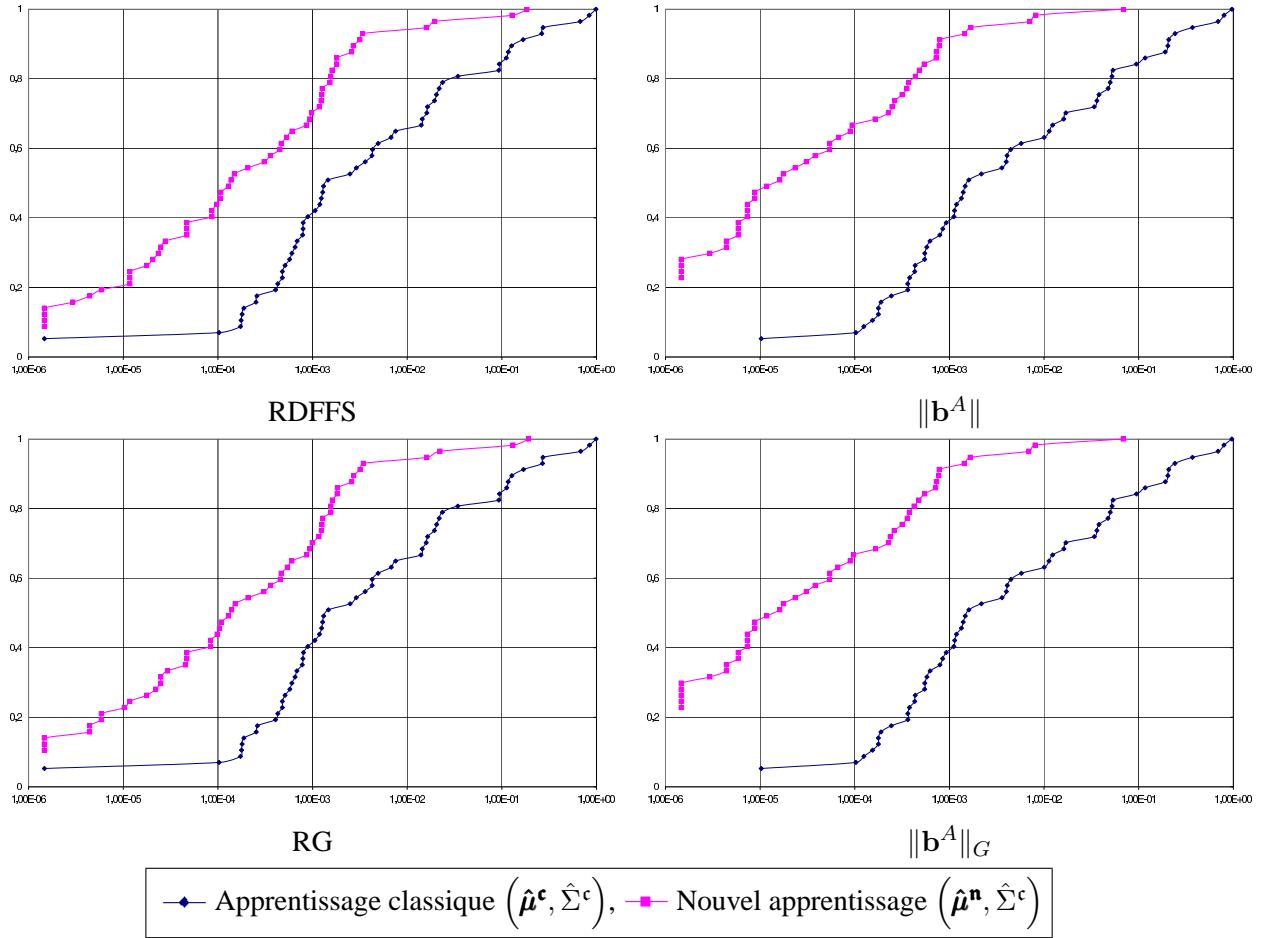


FIG. 7.4: Influence de l'apprentissage : courbes COR obtenues avec les détecteurs robustes sur un espace propre  $F$  de dimension  $J = 10$  pour les apprentissages  $(\hat{\mu}^c, \hat{\Sigma}^c)$  et  $(\hat{\mu}^n, \hat{\Sigma}^c)$ .

résultats de reconnaissance.

**Influence d'un bruit additif gaussien.** Un bruit gaussien de moyenne nulle et d'écart type égale à 20 est ajouté aux images de test pour simuler un bruit d'acquisition. Cela correspond à un rapport signal à bruit de 22 dB. La figure 7.7 compare les courbes ROC obtenues avec et sans ce bruit additif. On remarque une légère dégradation des performances des détecteurs robustes : pour 80% des objets détectés, le taux de fausses alarmes de  $\|\mathbf{b}^A\|$  (respectivement RDFFS), augmente de 0,04% à 0,1% (resp. de 0,15% à 0,26%).

### 7.3.1.3 Remarques

L'hypothèse de distribution gaussienne des coordonnées  $c$  dans l'espace propre  $F$  n'est pas vérifiée pour notre base d'apprentissage COIL. C'est pourquoi on observe des résultats similaires pour les deux

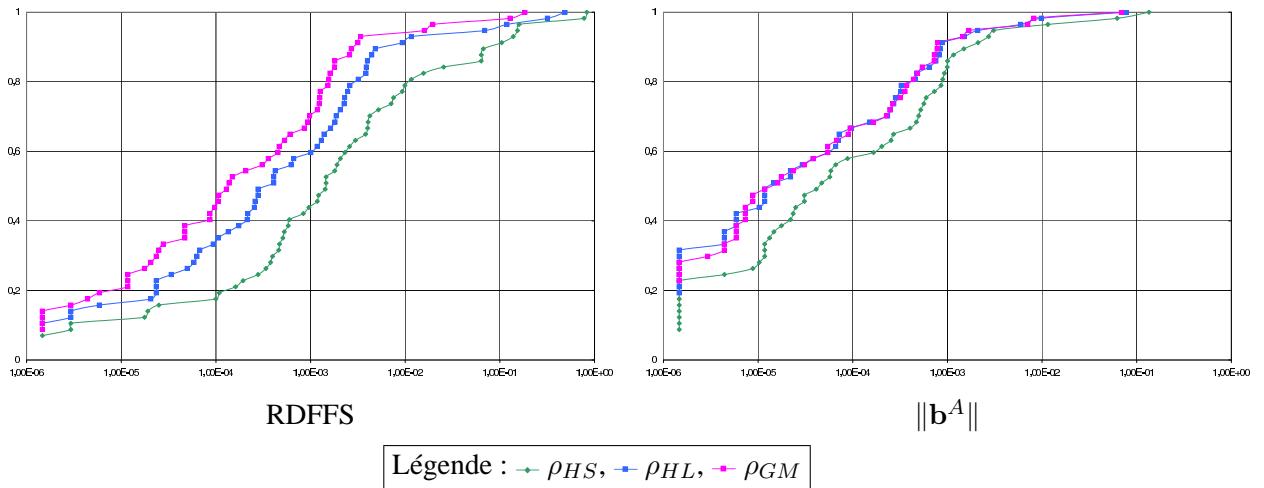


FIG. 7.5: Influence du choix de la fonction  $\rho$  : courbes COR calculées avec les fonctions  $\rho_{HS}$ ,  $\rho_{HL}$  et  $\rho_{GM}$  pour un espace propre  $F$  de dimension  $J = 10$ ).

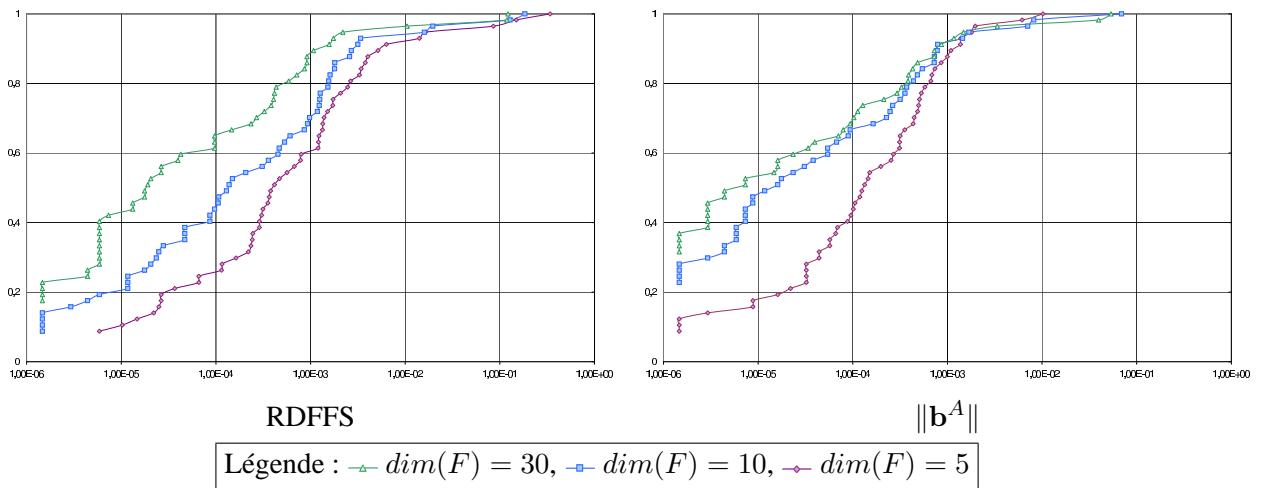


FIG. 7.6: Influence de la dimension de l'espace propre  $F$  : courbes COR obtenues avec les détecteurs robustes.

détecteurs classiques DFFS et la vraisemblance de Moghaddam et Pentland, ainsi que pour les détecteurs robustes RDFFS (respectivement  $\|\mathbf{b}^A\|$ ) et RG (resp.  $\|\mathbf{b}^A\|_G$ ). On remarque également que les performances de la norme de la pondération  $\|\mathbf{b}^A\|$ , surpassent celles de la vraisemblance robuste RDFFS.

Les résultats de détection sont meilleurs dans le cas de l'apprentissage réalisé avec le nouvel estimateur de la moyenne  $(\hat{\mu}^n, \hat{\Sigma}^c)$ . Ils s'améliorent également en augmentant la dimension de l'espace propre  $F$ . Les détecteurs robustes sont efficaces pour détecter les occurrences partiellement occultées, bruitées et apparaissant sur des fonds texturés. Le passage en continuation à des fonctions non-convexes qui rejettent mieux les erreurs grossières, améliore les résultats de détection.

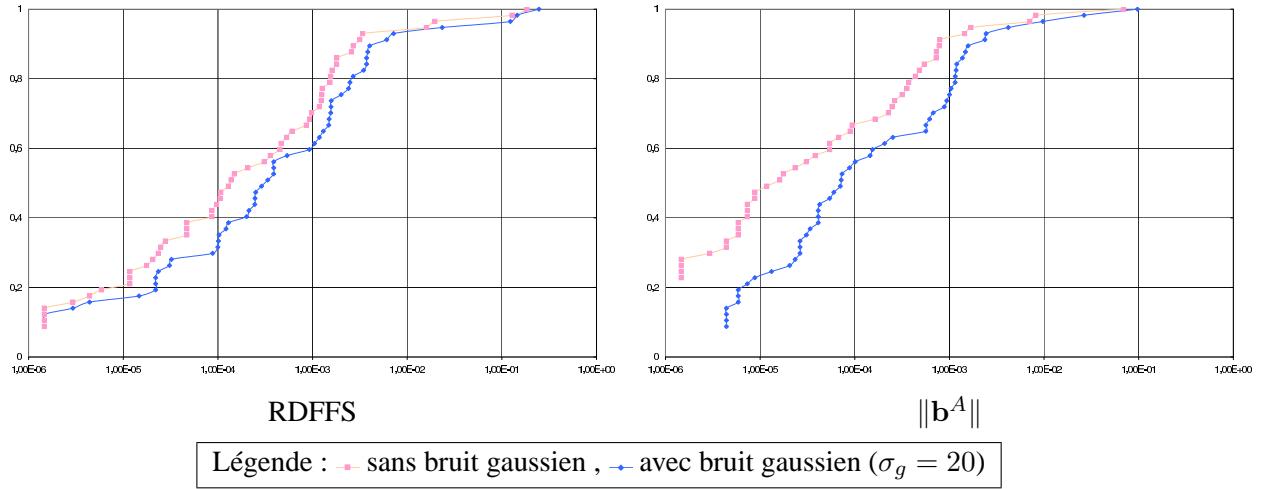


FIG. 7.7: Influence d'un bruit gaussien : courbes COR obtenues par les détecteurs robustes pour un espace propre  $F$  de dimension  $J = 10$ .

### 7.3.2 Expérience de détection avec des images en couleurs

#### 7.3.2.1 Détection avec la base A43

Le test est réalisé avec la base d'apprentissage A43 qui collecte les images de 43 panneaux triangulaires variant suivant leur rotation dans le plan de l'image (cf. annexe H). Les images utilisées pour ce test de détection, sont présentées à l'annexe K (section K.1). Tous les détecteurs, classiques (DFFS, vraisemblance de Moghaddam et Pentland) et robustes (RDFFS, RG,  $\|b^A\|$  et  $\|b^A\|_G$ ), ont été évalués avec les deux types d'apprentissage  $(\hat{\mu}^c, \hat{\Sigma}^c)$  et  $(\hat{\mu}^n, \hat{\Sigma}^c)$ . Les meilleurs résultats ont été obtenus pour les deux détecteurs robustes RDFFS et  $\|b^A\|$  avec un apprentissage réalisé avec le nouvel estimateur de la moyenne  $(\hat{\mu}^n, \hat{\Sigma}^c)$ . La figure 7.8 présente leurs courbes COR. Contrairement aux expériences réalisées avec la base COIL, ces résultats ne sont pas très satisfaisants : choisir un seuil permettant un bon pourcentage de bonnes détection impose un nombre important de fausses alertes. Par exemple, détecter 80% des objets recherchés impose un pourcentage de fausses alertes supérieur à 50% pour nos deux détecteurs robustes RDFFS et  $\|b^A\|$ .

La figure 7.9 présente une carte de détection (en fausses couleurs) obtenue avec la norme des pondérations  $\|b^A\|$  : les positions des pixels rouges foncés (voire noirs) sont les positions des observations les plus ressemblantes avec les images d'apprentissage. Les bordures grises indiquent les positions  $(i, j)$  où l'on ne peut pas extraire une observation y complète. Beaucoup de fausses alertes apparaissent notamment sur les régions de l'image  $I_2$  où il y a de l'herbe. Les détecteurs robustes, RDFFS et  $\|b^A\|$ , ne sont pas capables de discriminer ces fausses alertes des objets d'intérêt. Par ailleurs, l'*a priori* de distribution gaussienne sur les coordonnées  $c$  utilisé par le détecteur RG, n'est pas adapté à notre base d'apprentissage. C'est pourquoi RG ne permet pas d'obtenir de meilleurs résultats de détection que la

RDFFS.

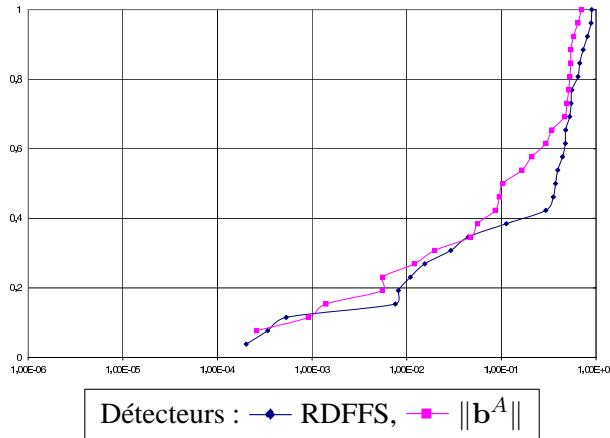


FIG. 7.8: Courbes COR obtenues pour les détecteurs robustes RDFFS et  $\|\mathbf{b}^A\|$ , sur un espace propre  $F$  de dimension  $J = 60$ , avec l'apprentissage  $(\hat{\mu}^n, \hat{\Sigma}^c)$ .

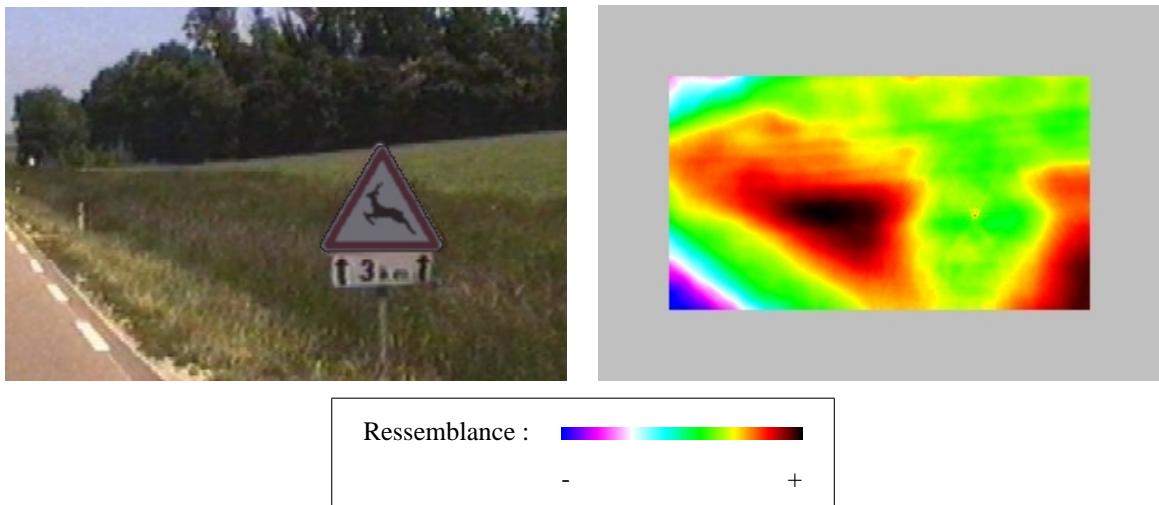


FIG. 7.9: Carte de détection de l'image  $I_2$  obtenue avec le détecteur robuste  $\|\mathbf{b}^A\|$ .

Dans nos détecteurs, seule la vraisemblance sur la distribution de l'erreur,  $\mathcal{P}(\mathbf{w}|\mathbf{B})$ , est réellement utilisée. La distribution  $\mathcal{P}(\mathbf{c}|\mathcal{B})$  n'est pas exploitée correctement puisque les hypothèses simples : distribution uniforme et gaussienne, ne sont pas vérifiées par notre base d'apprentissage. Plusieurs solutions peuvent être envisagées pour améliorer ces résultats de détection.

- En choisissant une hypothèse adaptée pour la vraisemblance  $\mathcal{P}(\mathbf{c}|\mathcal{B})$ , on peut calculer explicitement la vraisemblance  $\mathcal{P}(\mathbf{y}|\mathcal{B})$  en résolvant l'intégrale :

$$\mathcal{P}(\mathbf{y}|\mathcal{B}) = \int \mathcal{P}(\mathbf{c}|\mathcal{B}) \mathcal{P}(\mathbf{w}|\mathcal{B}) d\mathbf{c}$$

La résolution de cette intégrale peut être difficile dans le cas général.

- La seconde solution consiste à utiliser notre détecteur  $\max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$ , en utilisant une hypothèse sur la distribution  $\mathcal{P}(\mathbf{c}|\mathcal{B})$  mieux adaptée à la base d'objets  $\mathcal{B}$ . Ce détecteur implique le calcul de l'estimée  $\hat{\mathbf{c}}_{RMAP} = \arg \max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$ , qui nécessite le développement d'un algorithme spécifique selon l'hypothèse  $\mathcal{P}(\mathbf{c}|\mathcal{B})$  choisie. Cette solution a permis de définir le détecteur RG dans le cas simple de distribution gaussienne de  $\mathcal{P}(\mathbf{c}|\mathcal{B})$ . Cependant le détecteur  $\max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B}) = \mathcal{P}(\mathbf{y}, \hat{\mathbf{c}}_{RMAP}|\mathcal{B})$  peut avoir un calcul complexe dans le cas général.
- Afin d'éviter l'estimation difficile de  $\hat{\mathbf{c}}_{RMAP}$ , on propose comme dernière alternative, d'utiliser l'estimée robuste au sens du maximum de vraisemblance  $\hat{\mathbf{c}}_{RMV}$ . Cette estimée a permis d'obtenir de bons résultats de reconnaissance (cf. chapitre 6). Aussi on approche le détecteur robuste  $\max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B})$  par :

$$\begin{aligned}\max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c}|\mathcal{B}) &= \mathcal{P}(\mathbf{y}, \hat{\mathbf{c}}_{RMAP}|\mathcal{B}) \approx \mathcal{P}(\mathbf{y}, \hat{\mathbf{c}}_{RMV}|\mathcal{B}) \\ &\approx \mathcal{P}(\hat{\mathbf{w}}|\mathcal{B}) \mathcal{P}(\hat{\mathbf{c}}_{RMV}|\mathcal{B})\end{aligned}$$

avec  $\hat{\mathbf{w}} = \tilde{\mathbf{y}} - \mathbf{U} \hat{\mathbf{c}}_{RMV}$  et l'hypothèse robuste de distribution du bruit,  $\mathcal{P}(\mathbf{w}|\mathcal{B})$ . La vraisemblance  $\mathcal{P}(\hat{\mathbf{w}}|\mathcal{B})$  correspond alors au détecteur RDFFS (cf. paragraphe 7.2.2). En choisissant une vraisemblance  $\mathcal{P}(\mathbf{c}|\mathcal{B})$  mieux adapté à notre base d'apprentissage, on espère améliorer les résultats de détection. Cela est mis en évidence au paragraphe suivant.

### 7.3.2.2 Détection avec la base AVG

Pour illustrer ce que peut apporter la prise en compte de  $\mathcal{P}(\mathbf{c}|\mathcal{B})$ , on se place dans une situation dans laquelle on peut approcher facilement cette distribution par une expression analytique simple. Pour cela, on considère une nouvelle base d'apprentissage notée AVG.

**Base d'apprentissage AVG.** On considère la base d'apprentissage constituée d'un seul objet variant suivant ses rotations dans le plan de l'image (une image tous les 2 degrés de rotation). On a choisi l'image moyenne des panneaux blancs triangulaires (cf. figure 7.10). On suppose que celle-ci permet de représenter les panneaux blancs de manière suffisante pour les détecter.

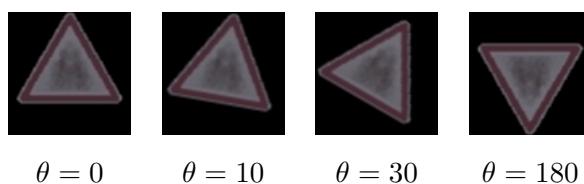


FIG. 7.10:Base AVG : image moyenne des panneaux blancs apprise suivant ses rotations dans le plan de l'image ( $\theta$  angle de rotation).

Au chapitre 5 (Base A1, section 5.2.2.2), on a montré la distribution particulière des projections des images d'apprentissage sur les plans propres associés à une valeur propre double. La figure 7.11 présente la répartition des images d'apprentissage de la base AVG dans les trois premiers plans propres (en vert). On rappelle que le pas de rotation entre les images de la base d'apprentissage est de 2 degrés. On visualise également les projections (en rouge) des images de panneaux triangulaires blancs variant avec un pas de 10 degrés de rotation dans le plan de l'image. On constate que, à quelques exceptions près, les projections des images des panneaux triangulaires se regroupent autour de celle du panneau moyen selon une couronne circulaire dans chaque plan propre associé à une valeur propre double. L'annexe D précise la répartition particulière de ces images (cf. équation D.5).

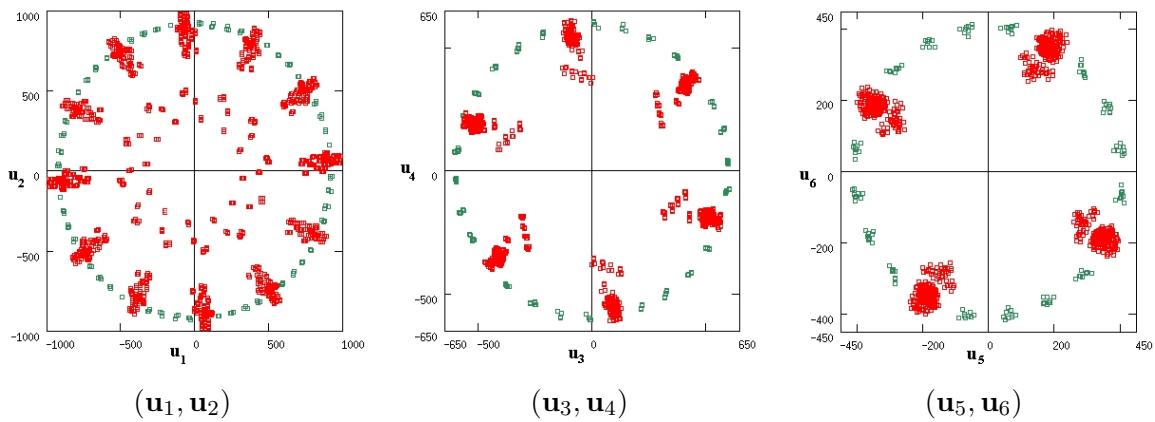


FIG. 7.11: Distribution des images d'apprentissage (en vert) et des panneaux blancs (en rouge) dans les plans propres.

**Hypothèse sur la distribution des coordonnées c :  $\mathcal{P}(c|\mathcal{B})$ .** Dans l'espace propre défini par les  $J$  premiers vecteurs propres issus de la base d'apprentissage AVG, on choisit d'approcher grossièrement la vraisemblance des coordonnées  $c$  des panneaux blancs par :

$$\mathcal{P}(c_1, \dots, c_J | \mathcal{B}) \propto \prod_{j=1}^{\frac{J}{2}} \exp - \left[ \frac{1}{2\gamma} |c_{2j-1}^2 + c_{2j}^2 - \lambda_{2j} - \lambda_{2j-1}| \right] \quad (7.5)$$

Les coordonnées  $(c_{2j-1}, c_{2j})$  dans le plan propre défini par les vecteurs  $(\mathbf{u}_{2j-1}, \mathbf{u}_{2j})$  associées à une valeur propre double ( $\lambda_{2j} = \lambda_{2j-1}$ ), sont théoriquement réparties suivant un cercle de rayon  $R^2 = \lambda_{2j} + \lambda_{2j-1}$ . La dimension  $J$  de l'espace propre  $F$ , est choisie paire.

**Hypothèse sur la distribution du bruit :  $\mathcal{P}(w|\mathcal{B})$ .** On se place dans le cadre de la distribution robuste du bruit. Suivant ces hypothèses sur la densité de probabilité de  $c$  et  $w$ , on évalue les performances

du détecteur  $\max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c} | \mathcal{B})$  approché par :

$$\begin{aligned}\max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c} | \mathcal{B}) &\approx \mathcal{P}(\mathbf{y}, \hat{\mathbf{c}}_{RMV} | \mathcal{B}) \\ &\approx \mathcal{P}(\hat{\mathbf{w}} | \mathcal{B}) \mathcal{P}(\hat{\mathbf{c}}_{RMV} | \mathcal{B})\end{aligned}$$

**Images de test.** Les images du test de détection sont présentées à l'annexe K.2. Dans cette expérience, on se place uniquement dans le cas de l'apprentissage  $(\hat{\mu}^n, \hat{\Sigma}^\epsilon)$ , qui donne les meilleurs résultats à la fois en reconnaissance et en détection. L'image moyenne de la base AVG,  $\hat{\mu}^n$ , a été introduite dans l'image  $I_{17}$  (cf. figure 7.12). Elle ne ressemble pas à un panneau blanc triangulaire. Cependant, elle est dans le plan propre ce qui implique que la probabilité de son erreur,  $\mathcal{P}(\mathbf{w} | \mathcal{B})$ , est maximum quelle que soit l'hypothèse de distribution choisie (robuste ou gaussienne). L'image moyenne  $\hat{\mu}^n$  de la base AVG représente donc unurre pour nos détecteurs robustes RDFFS et  $\|\mathbf{b}^A\|$ . De plus, les coordonnées  $\mathbf{c}$  de l'image moyenne  $\hat{\mu}^n$ , correspondent au vecteur nul de  $F$ , qui a une probabilité maximum lorsque l'on considère l'hypothèse de distribution gaussienne pour  $\mathcal{P}(\mathbf{c} | \mathcal{B})$ . Cela implique que le détecteur RG (et  $\|\mathbf{b}^A\|_G$ ) est également incapable d'écartier leurre  $\hat{\mu}^n$ .

**Résultats de détection sur un espace propre de dimension 20.** La figure 7.12 présente les cartes de ressemblance obtenues avec nos différents détecteurs robustes, RDFFS et  $\|\mathbf{b}^A\|$ , ainsi que les cartes de ressemblance calculées avec la vraisemblance  $\mathcal{P}(\hat{\mathbf{c}}_{RMV} | \mathcal{B})$ , pour deux images appartenant au test de détection,  $I_2$  et  $I_{17}$ . On rappelle que les bordures grises des cartes indique les positions où l'on ne peut pas extraire d'observation y complète de l'image  $I$ . Les cartes de ressemblance sont présentées en fausses couleurs : la couleur rouge foncée voire noire, indique de fortes valeurs de ressemblance, tandis que la couleur bleue caractérise les faibles mesures de ressemblance. Comme prévu, l'image moyenne  $\hat{\mu}^n$  est parfaitement détectée dans l'image  $I_{17}$  par les détecteurs RDFFS et  $\|\mathbf{b}^A\|$ , avec des valeurs de ressemblance élevées. Seule la carte calculée uniquement grâce à la vraisemblance  $\mathcal{P}(\hat{\mathbf{c}}_{RMV} | \mathcal{B})$  permet de discriminer ceurre des deux occurrences de panneaux également présentes dans l'image  $I_{17}$ .

Les courbes COR présentées à la figure 7.13 permettent d'évaluer les détecteurs robustes RDFFS et  $\|\mathbf{b}^A\|$ . Avant de considérer le détecteur robuste  $\max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c} | \mathcal{B}) \approx \mathcal{P}(\mathbf{y}, \hat{\mathbf{c}}_{RMV} | \mathcal{B})$ , on évalue les performances obtenues par la vraisemblance  $\mathcal{P}(\hat{\mathbf{c}}_{RMV} | \mathcal{B})$  calculé avec l'expression (7.5). Les performances des détecteurs classiques, DFFS et la vraisemblance de Moghaddam et Pentland, sont également reportées à la figure 7.13. On constate qu'elles sont largement inférieures à celles des détecteurs robustes.

Sur ce test, le détecteur robuste  $\|\mathbf{b}^A\|$  permet de détecter 80% des objets pour un pourcentage de fausses alarmes de 3%. La RDFFS obtient des résultats très médiocres puisque pour détecter 80% des objets, le pourcentage de fausses alarmes atteint 50%. A contrario, on remarque les très bonnes performances de détection obtenues par la vraisemblance  $\mathcal{P}(\hat{\mathbf{c}}_{RMV} | \mathcal{B})$ . Choisir un seuil permettant la détection de 80 % de nos objets induit un faible pourcentage de fausses alarmes de 0,06%.

Le détecteur  $\mathcal{P}(\mathbf{y}, \hat{\mathbf{c}}_{RMV} | \mathcal{B})$  est calculé par :

$$\mathcal{P}(\mathbf{y}, \hat{\mathbf{c}}_{RMV} | \mathcal{B}) \propto \exp \left[ -\frac{RDFFS(\mathbf{y})}{2} \right] \cdot \exp - \left[ \frac{1}{2\gamma} \sum_{j=1}^{\frac{J}{2}} |\hat{c}_{2j-1}^2 + \hat{c}_{2j}^2 - \lambda_{2j} - \lambda_{2j-1}| \right] \quad (7.6)$$

où  $\hat{\mathbf{c}} = \hat{\mathbf{c}}_{RMV}$ . L'énergie  $\mathcal{J}(\mathbf{y}) = -\ln \mathcal{P}(\mathbf{y}, \hat{\mathbf{c}}_{RMV} | \mathcal{B})$  s'écrit à une constante additive près :

$$\mathcal{J}(\mathbf{y}) = RDFFS(\mathbf{y}) + \frac{1}{\gamma} \sum_{j=1}^{\frac{J}{2}} |\hat{c}_{2j-1}^2 + \hat{c}_{2j}^2 - \lambda_{2j} - \lambda_{2j-1}|$$

Le paramètre  $\gamma$  de la distribution  $\mathcal{P}(\mathbf{c} | \mathcal{B})$  permet de pondérer les termes  $\mathcal{P}(\hat{\mathbf{w}} | \mathcal{B})$  (ou RDFFS), et  $\mathcal{P}(\hat{\mathbf{c}}_{RMV} | \mathcal{B})$ . Les fausses alarmes obtenues par la RDFFS et la vraisemblance dans l'espace propre  $\mathcal{P}(\hat{\mathbf{c}}_{RMV} | \mathcal{B})$  ne sont pas les mêmes comme en témoigne leurs comportements différents face auurre  $\hat{\mu}^n$  de l'image  $I_{17}$  (cf. figure 7.12). Plusieurs valeurs de  $\gamma$  ont été testées et la meilleure a été retenue. Le détecteur robuste  $\mathcal{P}(\mathbf{y}, \hat{\mathbf{c}}_{RMV} | \mathcal{B})$  améliore encore les résultats de détection (déjà bons) obtenus par la vraisemblance  $\mathcal{P}(\hat{\mathbf{c}}_{RMV} | \mathcal{B})$  (cf. figure 7.14). Le taux de fausses alarmes est diminué de 0,06% à 0,04% pour 80% d'objets détectés.

**Conclusion.** On a modélisé un espace propre  $F$  avec la base d'apprentissage AVG. Dans cet espace  $F$ , les projections, ou coordonnées  $\mathbf{c}$ , des panneaux blancs ont une densité de probabilité,  $\mathcal{P}(\mathbf{c} | \mathcal{B})$ , que l'on a pu approcher grossièrement par une expression analytique. Alors la mesure de probabilité  $\mathcal{P}(\hat{\mathbf{c}}_{RMV} | \mathcal{B})$  permet de détecter efficacement les occurrences de  $\mathcal{B}$  présentes dans nos images tests, pour un très faible pourcentage de fausses alarmes. Cela indique que la mesure  $\mathcal{P}(\hat{\mathbf{c}}_{RMV} | \mathcal{B})$  permet à elle seule dans cet exemple, de classer efficacement les observations  $\mathbf{y}$  dans les deux classes  $\mathcal{B}$  et  $\bar{\mathcal{B}}$ . La vraisemblance conjointe  $\mathcal{P}_\gamma(\mathbf{y}, \hat{\mathbf{c}}_{RMV} | \mathcal{B})$  considère en plus de  $\mathcal{P}(\hat{\mathbf{c}}_{RMV} | \mathcal{B})$ , la vraisemblance robuste de l'erreur, ce qui permet d'améliorer encore les résultats de détection.

## 7.4 Conclusion

Nous avons proposé dans ce chapitre de nouvelles méthodes de détection robustes basées sur une représentation par espace propre d'une classe d'objets d'intérêt  $\mathcal{B}$ . Les méthodes classiques, la DFFS et la vraisemblance de Moghaddam et Pentland, reposent sur une hypothèse de bruit gaussien sur l'observation. Cette hypothèse n'est plus valide lorsque l'on recherche également des occurrences dégradées des objets, par du bruit structuré (fond texturé, occultations partielles, etc.). On a donc proposé d'utiliser une hypothèse mieux adaptée à la distribution du bruit  $\mathcal{P}(\mathbf{w} | \mathcal{B})$  pour autoriser l'apparition d'erreurs grossières ou outliers. Celle-ci permet de définir deux familles de nouveaux détecteurs. La première correspond au maximum de la vraisemblance conjointe de l'observation  $\mathbf{y}$  et ses composantes  $\mathbf{c}$  sur l'espace propre :

$$\max_{\mathbf{c}} \mathcal{P}(\mathbf{y}, \mathbf{c} | \mathcal{B}) = \max_{\mathbf{c}} \{ \mathcal{P}(\mathbf{w} | \mathcal{B}) \mathcal{P}(\mathbf{c} | \mathcal{B}) \}$$

Dans le cadre de la théorie semi-quadratique, l'estimation de ce maximum nécessite l'introduction de variables auxiliaires. Celle introduite par l'algorithme ARTUR proposé au chapitre 6, les pondérations  $\mathbf{b}^A$  sur l'observation, définit une carte de confiance aux données observées. Comme seconde famille de détecteurs, on a proposé d'utiliser la norme des pondérations  $\|\mathbf{b}^A\|$ .

Ces deux familles de détecteurs ont été développés dans un premier temps, dans le cadre de deux hypothèses simples de la distribution  $\mathcal{P}(\mathbf{c}|\mathcal{B})$  : la distribution uniforme de  $\mathbf{c}$  dans  $F$  (détecteurs RDFFS et  $\|\mathbf{b}^A\|$ ) et la distribution gaussienne (RG et  $\|\mathbf{b}^A\|_G$ ). Les détecteurs RDFFS et RG, peuvent être perçus comme des versions robustes des détecteurs classique, DFFS et la vraisemblance de Moghaddam et Pentland.

Ces hypothèses simples sur  $\mathcal{P}(\mathbf{c}|\mathcal{B})$  ne sont pas adaptées aux classes d'images utilisées dans nos expérimentations. On a montré qu'en utilisant un modèle approprié pour la distribution de  $\mathcal{P}(\mathbf{c}|\mathcal{B})$ , on pouvait obtenir d'excellentes performances de détection. Dans notre exemple, la vraisemblance  $\mathcal{P}(\mathbf{c}|\mathcal{B})$  a été approchée grossièrement par une expression analytique. Dans le cas général, d'autres techniques, notamment non-paramétriques telles que les histogrammes multidimensionnels évoqués au chapitre 3, pourraient être également utilisées pour modéliser cette distribution.

Les méthodes de détection proposées dans ce chapitre, reposent sur une hypothèse unique : la classe  $\mathcal{B}$  est connue mais l'on ne dispose d'aucun a priori sur  $\bar{\mathcal{B}}$ . Les résultats de reconnaissance et de détection pourraient être améliorés en réalisant un apprentissage sur la non classe  $\bar{\mathcal{B}}$  ou en mettant plusieurs classes d'intérêt en compétition [96, 111, 46].

Une carte de vraisemblance robuste de dimension  $225 \times 125$  pixels est calculée en moyenne en 60 minutes pour l'expérience la plus longue, A43 (cf. section 7.3.2.1). Cette durée importante est liée notamment à la dimension des vecteurs manipulés, dans ce cas  $N \times P = 76 \times 76 \times 3 = 17328$ . Elle tombe à 14 minutes pour des images de dimension  $N \times P = 64 \times 64 \times 1 = 4096$  dans l'expérience COIL pour le même nombre  $J$  de composantes principales retenues (cf. section 7.3.1). Ainsi, le temps de calcul peut être réduit en considérant des images de plus faible résolution. Cette vitesse d'estimation robuste, limitante pour la détection dans de grandes images, est cependant acceptable pour analyser plus finement de petites régions d'intérêt préalablement détectées. Dans le cadre de la reconnaissance, l'estimation robuste permet d'identifier l'observation en moins d'une seconde.

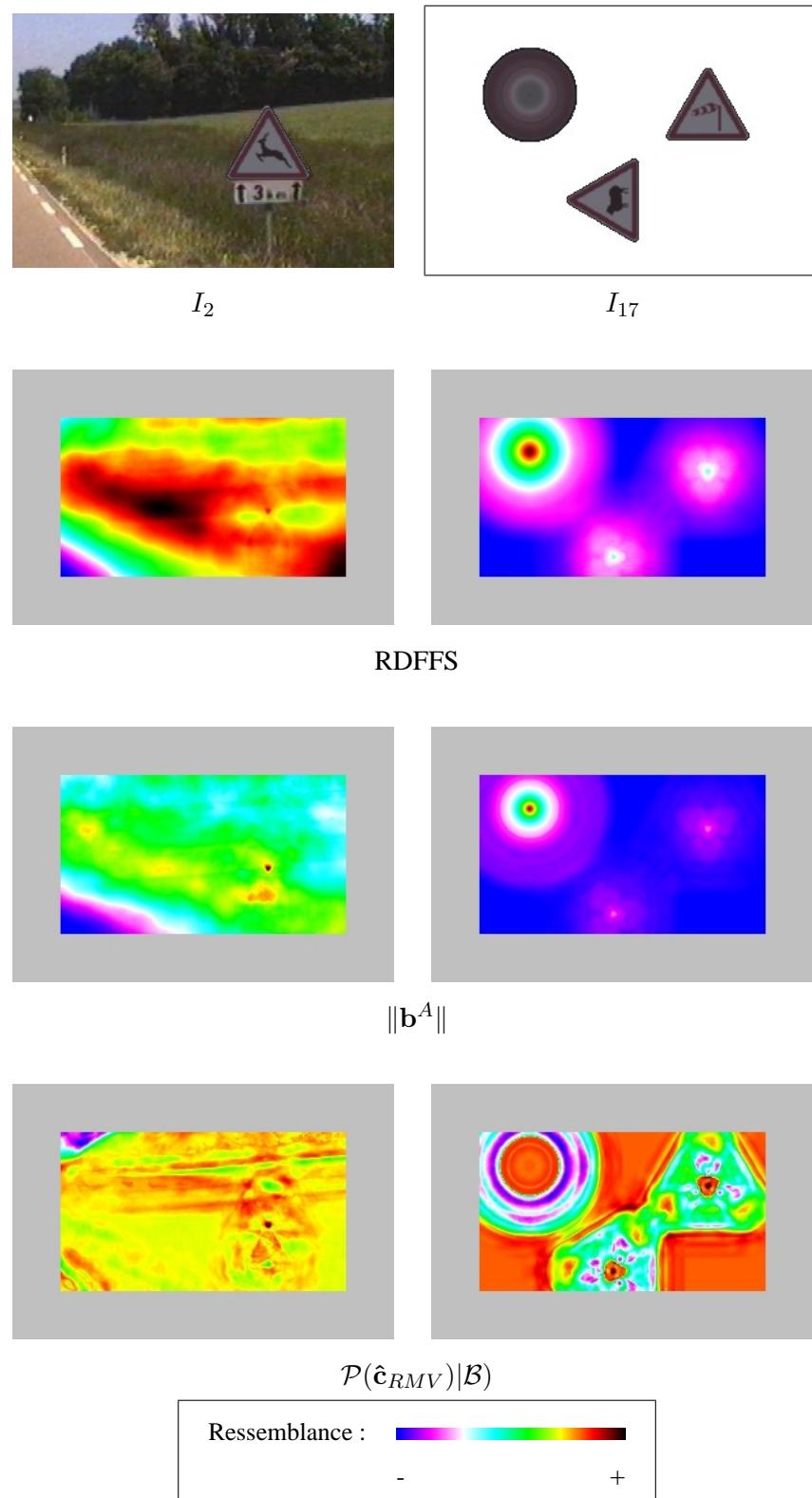


FIG. 7.12: Cartes de ressemblance obtenues sur les images  $I_2$  et  $I_{17}$ .

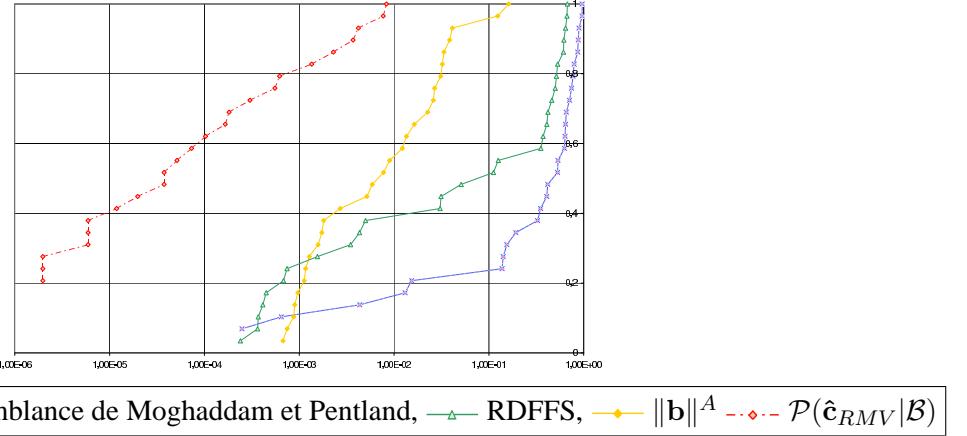


FIG. 7.13: Courbes COR obtenues pour les différents détecteurs classique et robustes sur un espace de dimension 20.

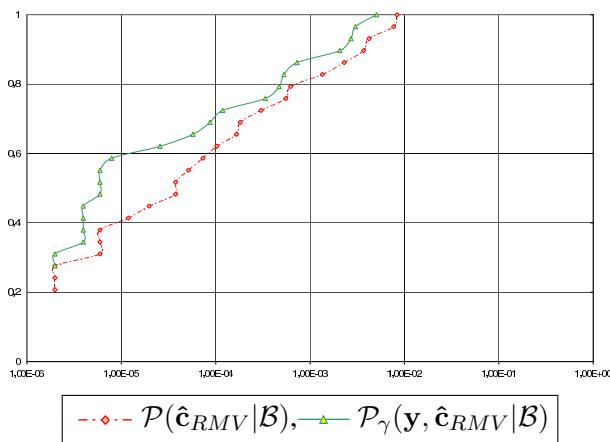


FIG. 7.14: Courbes COR obtenues avec  $\mathcal{P}(\hat{\mathbf{c}}_{RMV}|\mathcal{B})$  et  $\mathcal{P}_\gamma(\mathbf{y}, \hat{\mathbf{c}}_{RMV}|\mathcal{B})$  sur un espace propre  $F$  de dimension  $J = 20$ .

## Chapitre 8

# Conclusion et perspectives

Les travaux réalisés dans le cadre de cette thèse ont porté sur la description automatique de bases d'images séquentielles de scènes routières. Ils concernent spécifiquement la recherche d'objets fixes sur l'accotement tels que les arbres, les poteaux, etc. La stratégie générale envisagée pour cette application consiste à faire une pré-détection rapide des zones d'intérêt dans les images, avec éventuellement des fausses alarmes mais sans oubli, puis d'affiner ces premiers résultats. Suite à un état de l'art portant sur les techniques d'indexation et les systèmes de reconnaissance de panneaux, notre choix s'est porté sur des caractéristiques d'apparence utilisées avec des méthodes statistiques qui formalisent précisément la connaissance *a priori* disponible. Les caractéristiques locales d'apparence, simples, de petite dimension et rapides à extraire, ont été utilisées pour proposer une méthode de pré-sélection des zones d'intérêt. Les signatures globales, plus informatives, ont été utilisées pour proposer des méthodes de détection et de reconnaissance robustes afin d'affiner les résultats de la pré-détection.

Dans la première partie, nous avons proposé une méthode de pré-détection de zones d'intérêt par analyse des changements statistiques temporels des mesures locales d'apparence, dans les séquences d'images prises par une caméra en mouvement. Les événements détectés sont l'apparition, la disparition et le changement d'apparence d'objets dans une séquence d'images acquises par une caméra en mouvement. Cette méthode est entièrement non supervisée car elle ne requiert aucun apprentissage préalable des mesures sur notre classe d'objets d'intérêt. De plus, elle est rapide et elle ne nécessite ni estimation, ni compensation de mouvement. Cette pré-détection a été validée sur une séquence d'images réelles de scènes routières présentant de grands mouvements de caméra : les objets à fûts verticaux tels que les arbres, les délinéateurs ou les panneaux, sont correctement pré-détectés.

Dans la seconde partie, nous utilisons une représentation globale de l'apparence par espace propre. Les signatures globales de l'apparence sont de très grande dimension, et sont naturellement sensibles aux perturbations provoquées par exemple, par des occultations partielles. Les techniques classiques de détection et de reconnaissance sur les espaces propres ne sont pas robustes à la présence de ce bruit structuré. A l'aide des statistiques robustes, notamment les *M-estimateurs*, nous avons proposé des méthodes

de détection et de reconnaissance robustes à de tels effets. L'ensemble des techniques proposées s'inscrit dans un formalisme statistique unifié et original. Ces méthodes ont été validées expérimentalement pour la détection et la reconnaissance de panneaux routiers.

Dans le cadre de l'application de ces méthodes pour l'analyse de bases de données d'images de scènes routières, l'ensemble de la chaîne complète de traitement commence par définir dans la séquence des zones d'intérêt grâce à la pré-détection par mesures locales. Puis cette première segmentation est affinée par les techniques de détection et de reconnaissance robustes basées sur une représentation globale de l'apparence. Ces dernières sont bien adaptées à la recherche d'objets manufacturés et normalisés (panneaux, balises) dont l'apparence peut être apprise *a priori*. D'autres types d'objets tels que les arbres ou les poteaux, difficiles à apprendre dans leur globalité par leur trop grande variabilité, pourront être modélisés par apprentissage de leur mesures locales, et reconnus par des méthodes classiques telles que la comparaison d'histogrammes.

#### Perspectives :

- ✓ La méthode de pré-détection proposée en première partie est générique et n'est pas spécifique aux mesures locales de l'apparence. Les mesures doivent être choisies pour discriminer au mieux les objets d'intérêt et leur environnement. D'autres signatures locales proposées dans l'état de l'art, comme la texture, pourrait également être considérée.
- ✓ Comme on l'a déjà évoqué au paragraphe 4.4, l'utilisation des changements statistiques temporels peut être étendue au suivi des objets dans la séquence. De plus, une analyse approfondie de leur signature « statistico-temporelle » pourrait améliorer les méthodes de détection et de reconnaissance des objets en rives.
- ✓ Une autre extension de ces travaux consisterait à détecter les changements d'environnement notamment rural ou urbain. Une telle classification permettrait de segmenter la séquence en considérant l'entrée et la sortie dans les agglomérations qui sont des indicateurs intéressants d'un point de vue opérationnel.
- ✓ En plus des comportements statistiques temporels différents entre la classe d'objets d'intérêt et son contexte, d'autres types d'*a priori* peuvent permettre de discriminer ces deux classes. Par exemple, certains objets d'intérêt, tels que les panneaux, présentent certaines propriétés de symétrie qui sont peu courantes dans un environnement naturel. La recherche de symétrie particulière dans les images pourrait également permettre de définir des zones d'intérêt.
- ✓ La méthode de pré-détection d'objets dans les séquences d'images repose sur la connaissance *a priori* de comportements statistiques temporels différents entre notre classe d'objets d'intérêt et son environnement. Au contraire, les méthodes robustes de détection et de reconnaissance proposée dans la seconde partie, reposent sur un apprentissage préalable d'échantillons caractéristiques de notre classe d'intérêt.

Aucune hypothèse n'est faite sur le contexte dans lequel apparaissent nos objets. Une première amélioration possible consisterait à augmenter la connaissance *a priori* en considérant non seulement nos objets d'intérêt, mais aussi l'environnement particulier des scènes routières dans lequel ils apparaissent.

- ✓ L'apprentissage nécessaire aux méthodes de détection et de reconnaissance robuste par espace propre, doit être suffisamment exhaustif pour prendre en compte l'ensemble des variations possibles d'apparence des objets dans les images, notamment les changements de conditions d'éclairage. Pour rendre le système invariant à ces perturbations, on pourra envisager de transformer les données initiales couleurs codées en RGB, pour les rendre moins sensibles à de telles variations.
- ✓ Enfin, le modèle d'apparence globale utilisé dans cette thèse lie l'observation à une variable de plus petite dimension via une relation linéaire. L'état de l'art a rappelé d'autres relations, notamment non linéaires, éventuellement plus performantes. Le formalisme statistique utilisé dans nos méthodes de détection et de reconnaissance robustes, est générique et pourrait être également étendue aux relations non-linéaires.



# Bibliographie

- [1] D. ANDROUTSOS, K.N. PLATANIOTIS, and A.N. VENETSANOPoulos. « Image region extraction for content-based image retrieval ». In *proceedings of IX European Signal Processing Conference (EUPSICO 98)*, September 1998.
- [2] K. ARBTER, W.E. SNYDER, H. BURKHARDT, and G. HIRZINGER. « Application of Affine-Invariant Fourier Descriptors to Recognition of 3D Objects ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7) :640–646, July 1990.
- [3] A. BACCINI and P. BESSE. « Statistique Descriptive Multidimensionnelle ». Technical Report, Université Paul Sabatier, Laboratoire de Statistique et Probabilités, Toulouse, France, juillet 1999.
- [4] M. BASSEVILLE. « Distance measures for signal processing and pattern recognition ». *Signal Processing*, 18(4) :349–369, 1989.
- [5] M. BEATTY and B.S. MANJUNATH. « Dimensionality Reduction Using Multi-Dimensional Scaling for Content-Based Retrieval ». In *IEEE proceedings of the International Conference on Image Processing*, Santa Barbara, CA, October 1997.
- [6] P. BELHUMEUR, J. HESSPANHA, and J. KRIEGMAN. « Eigenfaces vs. Fisherfaces : recognition using class specific linear projection ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7) :711–720, july 1997.
- [7] S. BELONGIE, C. CARSON, H. GREENSPAN, and J. MALIK. « Color-and Texture-Based Image Segmentation Using EM and Its Application to Content-Based Image Retrieval ». In *IEEE proceedings of the International Conference on Computer Vision*, Bombay, India, 1998.
- [8] C. BIERNAKI and R. MOHR. « Indexation et appariement d’images par modèle de mélange gaussien ». Technical Report 3600, INRIA, France, January 1999.
- [9] A. Del BIMBO. *Visual Information Retrieval*. Academic Press Morgan Kaufmann Publishers, 1999.
- [10] Horst BISCHOF and Ales LEONARDIS. « Robust recognition of scaled eigenimages through a hierarchical approach ». In *IEEE proceedings of the conference on Computer Vision and Pattern Recognition*, pages 664–670, Santa Barbara, California, 23-25 June 1998.

- [11] C.M. BISHOP. *Neural Networks for Pattern Recognition*. Oxford University Press, 1995.
- [12] Å. BJÖRK. *Numerical Methods for Least Squares Problems*. SIAM, 1996.
- [13] M. J. BLACK and P. ANANDAN. « A Framework for the robust estimation of optical flow ». In *Proceedings of the 4th International Conference on Computer Vision*, pages 231–236, Berlin, Allemagne, May 1993.
- [14] M. J. BLACK, D. J. FLEET, and Y. YACOOB. « Robustly Estimating Changes in Image Appearance ». *Computer Vision and Image Understanding*, 78(1) :8–31, 2000.
- [15] M. J. BLACK and 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, January 1998.
- [16] M. J. BLACK and A. RANGARAJAN. « On the unification of line processes, outlier rejection, and robust statistics with applications in early vision ». *International Journal of Computer Vision*, 19(1) :57–92, July 1996.
- [17] John S. BORECZKY and Lawrence A. ROWE. « Comparison of Video Shot Boundary Detection Techniques ». In *SPIE proceedings of Storage and Retrieval for Image and Video Databases*, pages 170–179, 1996.
- [18] K. L. BOYER, M.J. MIRZA, and G. GANGULY. « The Robust Sequential Estimator : A General Approach and its Application to Surface Organization in Range Data ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(10) :987–1001, October 1994.
- [19] C. CARSON, S. BELONGIE, H. GREENSPAN, and J. MALIK. « Region-Based Image Querying ». In *Proceedings CVPR '97 Workshop on Content-Based Access of Image and Video Libraries*, Puerto Rico, June 1997.
- [20] C. CARSON and V.E. OGLE. « Storage and Retrieval of Feature Data for a very Large Online Image Collection ». *IEEE Computer Society Bulletin of the Technical Committee on Data Engineering*, 19(4) :19–27, December 1996.
- [21] K. R. CASTLEMAN. *Digital Image Processing*. Prentice-Hall, 1996.
- [22] K.-Y. CHANG and J. GHOSH. « A Unified Probabilistic Principal Surfaces ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(1) :22–41, January 2001.
- [23] T. CHANG and C-C.J. KUO. « Texture Analysis and Classification with Tree-Structured Wavelet Transform ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2(4) :429–441, October 1991.
- [24] P. CHARBONNIER, L. BLANC-FÉRAUD, G. AUBERT, and M. BARLAUD. « Deterministic Edge-Preserving Regularization in Computed Imaging ». *IEEE Transactions on Image Processing*, 6(2) :298–311, February 1997.

- [25] Pierre CHARBONNIER. *Reconstruction d'image : régularisation avec prise en compte des discontinuités*. Thèse de Doctorat, Université de Nice-Sophia Antipolis, France, 1994.
- [26] J.-Y. CHEN, C.A. BOUMAN, and J.P. ALLEBACH. « Fast Image Database Search using Tree-Structured VQ ». In *IEEE Proceedings of International Conference on Image Processing*, volume 2, pages 827–830, Santa Barbara, CA, October 1997.
- [27] R. DAHYOT, P. CHARBONNIER, and F. HEITZ. « Non-Supervised Robust Visual Recognition of Colour Images using Half-Quadratic Theory ». In *proceedings of European Workshop on Content-Based Multimedia Indexing (CBMI)*, Toulouse, FRANCE, October 1999.
- [28] R. DAHYOT, P. CHARBONNIER, and F. HEITZ. « Reconnaissance robuste non supervisée d'images en couleur utilisant la théorie semi-quadratique ». In *proceedings of GRETSI conference on signal and image processing*, volume 2, pages 295–298, Vannes, FRANCE, September 1999.
- [29] R. DAHYOT, P. CHARBONNIER, and F. HEITZ. « Robust Visual Recognition of Colour Images ». In *IEEE proceedings of the conference on Computer Vision and Pattern Recognition*, volume 1, pages 685–690, Hilton Head Island, USA, June 2000.
- [30] R. DAHYOT, P. CHARBONNIER, and F. HEITZ. « Détection d'évènements dans les séquences d'images avec caméra en mouvement ». In *to appear in proceedings of GRETSI conference on signal and image processing*, Toulouse, FRANCE, September 2001.
- [31] R. DAHYOT, P. CHARBONNIER, and F. HEITZ. « Détection robuste d'objets : une approche par modèle d'apparence ». In *to appear in proceedings of GRETSI conference on signal and image processing*, Toulouse, FRANCE, September 2001.
- [32] R. DAHYOT, P. CHARBONNIER, and F. HEITZ. « Unsupervised Statistical Change Detection in Camera-in-Motion Video ». In *IEEE proceedings of the International Conference on Image Processing*, Thessaloniki, GRECE, October 2001.
- [33] T.J. DARRELL, P. MAES, B. BLUMBERG, and A.P. PENTLAND. « A Novel Environment for Situated Vision and Behavior ». Technical Report 261, MIT Media Lab Perceptual Computing Section, June 1994.
- [34] T.J. DARRELL and A.P. PENTLAND. « Cooperative Robust Estimation Using Layers of Support ». Technical Report 163, MIT Media Lab Perceptual Computing Section, February 1991.
- [35] M. DAS, B.A. DRAPER, W.J. LIM, R. MANMATHA, and E.M. RISEMAN. « A Fast, Background-independant Retrieval Strategy for Color Image Databases ». Technical Report UM-CS-1996-079, UMass Computer Science Research Publications, November 1996.
- [36] M. DAS, E.M. RISEMAN, and B.A. DRAPER. « FOCUS : Searching for Multi-colored Objects in a Diverse Image Database ». In *IEEE proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR 97)*, pages 756–761, June 1997.

- [37] V. Colin de VERDIÈRE. « Visual Recognition using Local Appearance ». In *proceedings of the European Conference on Computer Vision*, University of Freiburg, Germany, June 1998.
- [38] Vincent Colin de VERDIÈRE. *Représentation et reconnaissance d'objets par champs réceptifs*. Thèse de Doctorat, Institut National Polytechnique de Grenoble, France, décembre 1999.
- [39] Y. DENG and B.S. MANJUNATH. « Content-based Search of Video Using Color, Texture, and Motion ». In *IEEE proceedings of the International Conference on Image Processing*, volume 2, Washington, DC, October 1997.
- [40] R. DERICHE. « Recursively implementing the gaussian and its derivatives ». Technical Report 1893, INRIA, France, April 1993.
- [41] Frank DIEBOLT. *Reconnaissance des marquages routiers*. Thèse de Doctorat, Université Louis Pasteur, Strasbourg I, France, décembre 1996.
- [42] F. ENNESSER and G. MEDIONI. « Finding Waldo, or Focus of Attention Using Local Color Information ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8) :805–809, August 1995.
- [43] S. ERCKERT. Détection et reconnaissance automatique de la signalisation verticale dans les scènes routières en couleur. Mémoire de D.E.A., Université Louis Pasteur, Strasbourg, juin 1997.
- [44] S. ESTABLE. *Reconnaissance D'objets en environnement extérieur dynamique - Application à la reconnaissance des panneaux routiers*. Thèse de Doctorat, Université Blaise Pascal, Clermont II, France, juillet 1996.
- [45] R. FABLET, P. BOUTHEMY, and P. PEREZ. « Statistical motion-based video indexing and retrieval ». In *6ème conférence internationale sur la Recherche d'Informations assistée par Ordinateur; RIAO'2000*, pages 602–619, Paris, Avril 2000.
- [46] K. FUKUNAGA. *Introduction to Statistical Pattern Recognition*. Academic Press, second edition, 1990.
- [47] B.V. FUNT and G.D. FINLAYSON. « Color Constant Color Indexing ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5) :522–529, May 1995.
- [48] Marc GELGON. *Segmentation spatio-temporelle et suivi dans une séquence d'images : application à la structuration et à l'indexation de vidéo*. Thèse de Doctorat, Université de Rennes I, novembre 1998.
- [49] D. GEMAN and C. YANG. « Nonlinear image recovery with half-quadratic regularization and FFT's ». *IEEE Transactions on Image Processing*, 4(7) :932–946, July 1995.
- [50] S. GEMAN and G. REYNOLDS. « Constrained restoration and the recovery of discontinuities ». *IEEE Transactions on Pattern Analysis Machine Intelligence*, 14(3) :367–383, March 1992.

- [51] M.M. GORKANI and R.W. PICARD. « Texture Orientation for Sorting Photos at a glance ». Technical Report 292, MIT Media Lab Perceptual Computing Section, October 1994.
- [52] P. GROS, G. MCLEAN, R. DELON, R. MOHR, C. SCHMID, and G. MISLER. « Utilisation de la couleur pour l'appariement et l'indexation d'images ». Technical Report 3269, INRIA, September 1997.
- [53] L. GUIBERT, G. KERYER, A. SERVEL, M. ATTIA, H.S. MACKENZIE, P. PELLAT-FINET, and J.-L. de Bougrenet de la TOCNAYE. « On-board optical joint transform correlator for real-time road sign recognition ». *Optical Engineering*, 34(1) :135–143, January 1995.
- [54] J. HAFNER, H.S. SWAHNEY, W. EQUITZ, M. FLICKNER, and W. NIBLACK. « Efficient Color Histogram Indexing for Quadratic Form Distance Functions ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7) :729–736, July 1995.
- [55] Raouf HAMDAN. *Détection, suivi et reconnaissance des formes et du mouvement par modèles probabilistes d'apparence*. Thèse de Doctorat, Université Louis Pasteur, Strasbourg I, France, janvier 2001.
- [56] F. R. HAMPEL, E. M. RONCHETTI, P. J. ROUSSEEUW, and W. A. STAHEL. *Robust Statistics : The Approach Based on Influence Functions*. John Wiley and Sons, 1986.
- [57] J. HARTHONG. *Probabilités & statistiques - De l'intuition aux applications*. Diderot Editeur, Arts et Sciences, 1996.
- [58] G. HEALEY and L. WANG. « Illumination-invariant recognition of texture in color images ». *Journal of the optical society of America*, 12(9) :1877–1883, September 1995.
- [59] Pierre HELLIER. *Recalage non rigide en imagerie cérébrale : méthodes et validations*. Thèse de Doctorat, Université de Rennes I, France, Décembre 2000.
- [60] N. HERODOTOU, K.N. PLATANIOTIS, and A.N. VENETSANOPoulos. « A colour segmentation and Classification Scheme for Facial Image and Video Retrieval Extended Summary ». In *proceedings of IX European Signal Processing Conference (EUPSICO 98)*, September 1998.
- [61] Z. HUANG and F.S. COHEN. « Affine-Invariant B-Spline Moments for Curve Matching ». *IEEE Transactions on Image Processing*, 5(10) :1473–1480, 1996.
- [62] P.J. HUBER. *Robust Statistics*. John Wiley and Sons, 1981.
- [63] D.P. HUTTENLOCHER, G.A. KLANDERMAN, and W.J. RUCKLIDGE. « Comparing Images Using the Hausdorff Distance ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9) :850–863, September 1993.
- [64] A.K. JAIN and A. VAILAYA. « Image Retrieval using color and shape ». *Pattern Recognition*, 29(8) :1233–1244, 1996.

- [65] P. JOLY. *Consultation et Analyse des Documents en Image Animée Numérique*. Thèse de Doctorat, Université des Sciences Paul Sabatier de Toulouse, juillet 1996.
- [66] D.S. KANG, N.C. GRISWOLD, and N. KEHTARNAVAZ. « An Invariant Traffic Sign Recognition System, Based on Sequential Color Processing and Geometrical Transformation ». In *IEEE proceedings of International Conference on Image Processing (ICIP)*, Texas, 1994.
- [67] H. KAUPPINEN, T. SEPPÄNEN, and M. PIETIKÄINEN. « An Experimental Comparison of Autoregressive and Fourier-Based Descriptors in 2D Shape Classification ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2) :201–207, February 1995.
- [68] S. M. KAY. *Fundamentals of Statistical signal processing - Estimation Theory*. Prentice Hall International Editions, 1993.
- [69] B. KÉGL, A. KRZYZAK, T. LINDER, and K. ZEGER. « Learning and Design of Principal Curves ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(3) :281–297, March 2000.
- [70] P.M. KELLY, M. CANNON, and D. HUSH. « Query by image example : the CANDID approach ». In *SPIE Proceedings Storage and Retrieval for Image and Video Databases III*, volume 2420, pages 238–248, 1995.
- [71] A. KHOTANZAD and Y.H. HONG. « Invariant Image Recognition by Zernike Moments ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5) :489–497, 1990.
- [72] M. KLIOT and E. RIVLIN. « Invariant-Based Shape Retrieval in Pictorial Databases ». In *IEEE proceedings of fifth European Conference on Computer Vision*, volume 1, pages 491–507, Freiburg, Germany, June 1998.
- [73] M.K. KWONG and B. LIN. « W-transform Method for Feature-oriented Multiresolution Image Retrieval ». Technical Report Preprint MCS-P489-0195, Mathematics and Computer Science Division, 1995.
- [74] F. De la TORRE and M. J. BLACK. « Robust Principal Component Analysis for Computer Vision ». In *Proceedings of the International Conference on Computer Vision*, Vancouver, Canada, July 2001.
- [75] M. LALONDE and Y. LI. « Detection of Road Signs using Color Indexing ». Technical Report CRIM-IIT-95/12-49, Centre de Recherche Informatique de Montréal, 1995.
- [76] M. LALONDE and Y. LI. « Road Sign Recognition ». Technical Report CRIM-IIT-95/09-35, Centre de Recherche Informatique de Montréal, August 1995.
- [77] M. LALONDE and Y. LI. « Road Sign Recognition ». Technical Report CRIM-IIT-95/12-48, Centre de Recherche Informatique de Montréal, December 1995.
- [78] Z. LEI, Y. CHAN, and D. LOPRESTI. « Image Curvelet Features Extraction and Matching ». In *IEEE Proceedings of International Conference on Image Processing*, Santa Barbara, CA, October 1997.

- [79] Z. LEI, T. TASDIZEN, and D.B. COOPER. « PIMs and Invariant Parts for Shape Recognition ». In *Proceedings of Sixth International Conference on Computer Vision (ICCV'98)*, pages 827–832, Bombay, India, 1998. also as LEMS Tech. Report 163, Brown University.
- [80] A. LEONARDIS and H. BISCHOF. « Robust recognition Using Eigenimages ». *Computer Vision and Image Understanding*, 78(1) :99–118, 2000.
- [81] W. LI. « Numerical Algorithms for the Huber M-estimator Problem ». In C.K. CHUI and L.L. SCHUMAKER, editors, *Approximation Theory VIII*, pages 325–334, New York, 1995.
- [82] K.-C. LIANG and C.-C.J. KUO. « Progressive Image Indexing and Retrieval Based on Embedded Wavelet Coding ». In *IEEE proceedings of the International Conference on Image Processing*, Santa Barbara, CA, October 1997.
- [83] F. LIU and R.W. PICARD. « Periodicity, directionality, and randomness : wold features for image modeling and retrieval ». Technical Report 320, MIT Media Lab Perceptual Computing Section, March 1995.
- [84] Wei-Ying MA. « *Netra : A Toolbox for Navigating Large Image Databases* ». PhD thesis, University of California, Santa Barbara, USA, June 1997.
- [85] W.Y. MA, Y. DENG, and B.S. MANJUNATH. « Tools for texture/color based search of images ». In *SPIE Conference on Human Vision and Electronic Imaging II*, pages 496–507, February 1997.
- [86] W.Y. MA and B.S. MANJUNATH. « NeTra : A Toolbox for Navigating Large Image Databases ». In *IEEE proceedings of International Conference on Image Processing*, volume 1, Washington, DC, October 1997.
- [87] E.C. MALTHOUSE. « Some Theoretical Results on Nonlinear Principal Components Analysis ». Technical Report, september 1996.
- [88] O. L. MANGASARIAN and D. R. MUSICANT. « Robust Linear and Support Vector Regression ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(9) :950–955, September 2000.
- [89] B.S. MANJUNATH and W.Y. MA. « Texture Features for Browsing and Retrieval of Image Data ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8) :837–842, August 1996.
- [90] R. MANMATHA and S. RAVELA. « A syntactic characterisation of appearance and its application to image retrieval ». In *SPIE Proceedings of the Conference on Human Vision and Electronic Imaging II*, volume 3016, San Jose, February 1997.
- [91] J. MAO and A.K. JAIN. « Texture classification and segmentation using multiresolution simultaneous autoregressive models ». *Pattern Recognition*, 25(2) :173–188, 1992.
- [92] F. MARQUES and G. GUTIERREZ. « Shape representation for object correspondence based on Sub-Graph matching and Fourier Descriptors ». In *proceedings of IX European Signal Processing Conference (EUPSICO 98)*, September 1998.

- [93] T.P. MINKA. « An image database browser that learns from user interaction ». Technical Report 365, MIT Media Lab Perceptual Computing Section Technical Report, 1996.
- [94] T.P. MINKA and R.W. PICARD. « Interactive learning using a society of models ». Technical Report 349, MIT Media Lab Perceptual Computing Section, 1995.
- [95] B. MOGHADDAM. « Principal Manifolds and Bayesian Subspaces for Visual Recognition ». Technical Report TR-99-35, Mitsubishi research Laboratory, Cambridge, Massachusetts, July 1999.
- [96] B. MOGHADDAM and A. PENTLAND. « Probabilistic Visual Learning for Object Detection ». In *IEEE Proceedings of International Conference on Computer Vision*, Cambridge, USA, June 1995.
- [97] B. MOGHADDAM and A. PENTLAND. « Probabilistic Visual Learning for Object Recognition ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7) :696–710, Juillet 1997.
- [98] R. MOHR, P. GROS, B. LAMIROY, S. PICARD, and C. SCHMID. « Indexation et Recherche d'Images ». In *proceedings of the GRETSI conference on signal and image processing*, Grenoble, France, 1997.
- [99] F. MOKHTARIAN. « Silhouette-Based Isolated Object Recognition through Curvature Scale Space ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(5) :539–544, May 1995.
- [100] F. MOKHTARIAN, S. ABBASI, and J. KITTLER. « Robust and Efficient Shape Indexing through Curvature Scale Space ». In *Proceedings of the 1996 British Machine and Vision Conference BMVC'96*, Edinburgh, Scotland, 1996.
- [101] F. MOKHTARIAN, S. ABBASI, and J. KITTLER. « Efficient Curvature-based Shape representation for Similarity Retrieval ». In *proceedings of IX European Signal Processing Conference (EUP-SICO 98)*, pages 597–600, Island of Rhodes, Greece, September 1998.
- [102] P. MONTESINOS, V. GOUET, R. DERICHE, and D. PELÉ. « Matching Color Uncalibrated Images using Differential Invariants ». *Image and Vision Computing*, 18(9) :659–671, June 2000.
- [103] H. MURASE and S. K. NAYAR. « Visual Learning and Recognition of 3-D Objects from Appearance ». *International Journal of Computer Vision*, 14(1) :5–24, January 1995.
- [104] C. NASTAR. « Indexation d'Images par le Contenu : un Etat de l'Art ». In *Journées CNET, CORESA'97*, Issy-les-Moulineaux, France, March 1997.
- [105] C. NASTAR, B. MOGHADDAM, and A. PENTLAND. « Generalized Matching for Recognition and Retrieval in an Image Database ». In *Proceedings of the Third International Conference Communicating by Image and Multimedia (Imagecom'96)*, Arcachon, France, May 1996.
- [106] S. A. NENE, S. K. NAYAR, and H. MURASE. « Columbia Object Image Library (COIL-20) ». Technical Report CU-CS-005-96, Department of Computer Science, Columbia University, 1996.
- [107] Christoforos NIKOU. *Contribution au recalage d'images médicales multimodales : approches par fonctions de similarité robustes et modèles déformables sous contraintes statistiques*. Thèse de Doctorat, Université Louis Pasteur, Strasbourg I, France, mai 1999.

- [108] K. OHBA and K. IKEUCHI. « Detectability, Uniqueness, and Reliability of Eigen Windows for Stable Verification of Partially Occluded Object ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(9) :1043–1048, September 1997.
- [109] S. PANCHANATHA, M.K. MANDAL, and T. ABOULNASR. « Video Indexing in the Wavelet Compressed Domain ». In *IEEE proceedings of the International Conference on Image Processing*, 1998.
- [110] A. PAPOULIS. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, third edition, 1991.
- [111] A. PENTLAND, B. MOGHADDAM, and T. STARNER. « View-Based and Modular Eigenspace for Face Recognition ». In *IEEE Proceedings of conference on Computer Vision and Pattern Recognition*, Seattle, USA, July 1994.
- [112] A. PENTLAND, R.W. PICARD, and S. SCLAROFF. « Photobook : Content-Based Manipulation of Image Databases ». Technical Report 255, MIT Media Lab Perceptual Computing Section, November 1993.
- [113] R.W. PICARD. « A Society of Models for Video and Image Libraries ». Technical Report 360, MIT Media Lab Perceptual Computing Section, 1996.
- [114] R.W. PICARD and T. KABIR. « Finding Similar Patterns in Large Image Databases ». Technical Report 205, MIT Media Lab Perceptual Computing Section, April 1993.
- [115] R.W. PICARD, T. KABIR, and F. LIU. « Real-time Recognition with the entire Brodatz Texture Database ». Technical Report 215, MIT Media Lab Perceptual Computing Section, June 1993.
- [116] R.W. PICARD and F. LIU. « A new Wold ordering for image similarity ». Technical Report 237, MIT Media Lab Perceptual Computing Section, April 1994.
- [117] R.W. PICARD and T.P. MINKA. « Vision Texture for Annotation ». Technical Report 302, MIT Media Lab Perceptual Computing Section, 1994.
- [118] H. Vincent POOR. *An Introduction to Signal Detection and Estimation*. Springer, 1994.
- [119] J. POUGET. « Probabilités ». Cours de l’Ecole Supérieure en Sciences informatiques, Université de Nice-Sophia Antipolis, 1988.
- [120] W. H. PRESS, S. A. TEUKOLSKY, W. T. VETTERLING, and B. P. FLANNERY. *Numerical Recipes in C - The Art of Scientific Computing*. Cambridge University Press, 1995.
- [121] L. PRIESE, J. KLIEBER, R. LAKMANN, V. REHRMANN, and R. SCHIAN. « New Results on Traffic Sign Recognition ». In *Proceedings Intelligent Vehicles Symposium '94*, pages 249–254, Paris, France, October, 14-16 1994.
- [122] R.J. QIAN and M.I. SEZAN. « Content-Scalable Shape Representation and Coding ». In *IEEE proceedings of the International Conference on Image Processing*, Santa Barbara, CA, October 1997.

- [123] T. RANDEN and J. HUSOY. « Image Content Search by Color and Texture Properties ». In *IEEE proceedings of the International Conference on Image Processing*, Santa Barbara, CA, October 1997.
- [124] R.P.N. RAO and D.H. BALLARD. « Object Indexing using an Iconic Sparse Distributed Memory ». Technical Report 559, Computer Science Department, University of Rochester, New York, 1995.
- [125] V. REHRMANN and L. PRIESE. « Fast and Robust Segmentation of Natural Color Scenes ». In *proceedings of the Third Asian Conference on Computer Vision*, Hong Kong, China, January, 8-10 1998.
- [126] W. RITTER. « Traffic Sign Recognition in Color Image Sequences ». In *IEEE proceedings of Intelligent Vehicles'92 Symposium*, pages 12–17, 1992.
- [127] Y. RUBNER, C. TOMASI, and L.J. GUIBAS. « The Earth Mover's Distance as a Metric for Image Retrieval ». Technical Report STAN-CS-TN-98-86, Stanford University, USA, September 1998.
- [128] E. SAHOURIA and A. ZAKHOR. « Content Analysis of Video Using Principal Components ». In *IEEE proceedings of International Conference on Image Processing*, volume 3, pages 541–545, Chicago, IL, 4-7 October 1998.
- [129] S. SANTINI and R. JAIN. « Similarity Queries in Image Databases ». In *IEEE proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR 96)*, pages 646–651, 1996.
- [130] R. SCHALKOFF. *Pattern Recognition - Statistical, Structural and Neural Approaches*. John Wiley & Sons, 1992.
- [131] Bernt SCHIELE. *Reconnaissance d'objets utilisant des histogrammes multidimensionnels de champs réceptifs*. Thèse de Doctorat, Institut National Polytechnique de Grenoble, France, 1997.
- [132] Cordelia SCHMID. *Appariement d'images par invariants locaux de niveaux de gris : Application à l'indexation d'une base d'objets*. Thèse de Doctorat, Institut National Polytechnique de Grenoble, France, 1996.
- [133] Henry SCHNEIDERMAN. « A Statistical Approach to 3D Object Detection Applied to Faces and Cars ». PhD thesis, Carnegie Mellon University, Pittsburg USA, 2000.
- [134] B. SCHOLKOPF. « Support Vector Learning ». PhD thesis, Berlin University, Allemagne, 1997.
- [135] S. SCLAROFF and A. PENTLAND. « Modal Matching for Correspondence and Recognition ». Technical Report 201, MIT Media Lab Perceptual Computing Section, 1993. also in proceedings of ICCV93.
- [136] S. SCLAROFF and A. PENTLAND. « Modal Matching for Correspondence and Recognition ». Technical Report 304, MIT Media Lab Perceptual Computing Section, 1993. also in PAMI.

- [137] N. SEBE, M.S. LEW, and D. HUIJSMANS. « Toward Improved Ranking Metrics ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10) :1132–1143, October 2000.
- [138] I. SEKITA, T. KURITA, and N. OTSU. « Complex Autoregressive Model for Shape Recognition ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(4) :489–496, April 1992.
- [139] T. SERRE, B. HEISELE, S. MUKHERJEE, and T. POGGIO. « Feature Selection for Face Detection ». Technical Report AI Memo 1697 and C.B.C.L. paper 192, Massachusetts Institute of Technology, Massachusetts, USA, September 2000.
- [140] C.J. SETCHELL. « Application of Computer Vision to Road-traffic Monitoring ». PhD thesis, Faculty of Engineering, Department of Electrical and Electronic Engineering, September 1997.
- [141] A. W. M. SMEULDERS, M. WORRING, S. SANTINI, A. GUPTA, and R. JAIN. « Content-Based Image Retrieval at the End of the Early Years ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12) :1349–1380, December 2000.
- [142] John R. SMITH. « Integrated Spatial and Feature Image Systems : Retrieval Analysis and Compression ». PhD thesis, Columbia University, USA, 1997.
- [143] Alexander SMOLA. « Regression Estimation with Support Vector Learning Machines ». Master's thesis, Université technologique de Munich, Allemagne, décembre 1996.
- [144] H. STARK and J. W. WOODS. *Probability, Random Processes, and Estimation Theory For Engineers*. Prentice Hall, second edition, 1994.
- [145] C. V. STEWART. « Bias in Robust Estimation Caused by Discontinuities and Multiple Structures ». *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(8) :818–833, August 1997.
- [146] C.V. STEWART. « Robust parameter estimation in computer vision ». *SIAM Reviews*, 41(3) :513–537, September 1999.
- [147] R. STOICA, J. ZERUBIA, and J.M. FRANCOS. « Indexing and Retrieval in Multimedia Libraries Through Parametric Texture Modeling using the 2D Wold Decomposition ». Technical Report 3594, INRIA, December 1998.
- [148] M. STRICKER and A. DIMAI. « Color Indexing with Weak Spatial Constraints ». In *SPIE proceedings of the conference on Storage and Retrieval for Image and Video Databases IV*, volume 2670, San Diego CA, USA, February 1996.
- [149] M.J. SWAIN and D.H. BALLARD. « Color Indexing ». *International Journal of Computer Vision*, 7(1) :11–32, 1991.
- [150] M.D. SWANSON and A.H. TEWFIK. « Affine-invariant multiresolution image retrieval using B-Splines ». In *IEEE proceedings of the International Conference on Image Processing*, Santa Barbara, CA, October 1997.

- [151] M. SZUMMER and R.W. PICARD. « Temporal Texture Modeling ». Technical Report 381, MIT Media Lab Perceptual Computing Section Technical Report, September 1996.
- [152] C. W. THERRIEN. *Discrete Random Signals and Statistical Signal Processing*. Prentice Hall Signal Processing Series, 1992.
- [153] M.E. TIPPING and C.M. BISHOP. « Probabilistic Principal Component Analysis ». Technical Report NCRG/97/010, Neural Computing Research Group, Aston University, Septembre 1997.
- [154] M. TURK and A. PENTLAND. « Eigenfaces for Recognition ». *Journal of Cognitive Neuroscience*, 3(1) :71–86, 1991.
- [155] M. UEHONARA and T. KANADE. « Optimal Approximation of Uniformly Rotated Images : Relationship Between Karhunen-Loeve Expansion and Discrete Cosine Transform ». *IEEE Transactions on Image Processing*, 7(1) :116–119, January 1998.
- [156] C. VERTAN and N. BOUJEMAA. « Spatially constrained color distribution for image indexing ». In *proceedings of International Conference on Color in Graphics and Image Processing (CGIP)*, Saint Etienne, France, October 2000.
- [157] J.S. WACHMAN and R.W. PICARD. « Tools for Browsing a TV Situation Comedy Based on Content Specific Attributes ». Technical Report 477, MIT Media Lab Perceptual Computing Section, December 1998.
- [158] C. WANG, H.-C. WU, and J.C. PRINCIPE. « A Cost Function for Robust Estimation of PCA ». In *SPIE Proceedings Applications and Science of Artificial Neural Networks II*, volume 2760, pages 120–127, 1996.
- [159] J.Y.A. WANG and E.H. ADELSON. « Layered Representation for motion analysis ». Technical Report 221, MIT Media Lab Perceptual Computing Section, 1993.
- [160] M.E.J. WOOD, N.W. CAMPBELL, and B.T. THOMAS. « Searching Large Image Databases using Radial-Basis Function Neural Networks ». In *Proceedings of the Sixth International Conference on Image Processing and its Applications*, pages 116–120, London, July 1997.
- [161] H.-H. YU and W. WOLF. « A Hierarchical, Multi-resolution Method for Dictionary-Driven Content-Based Image Retrieval ». In *IEEE proceedings of the International Conference on Image Processing*, Santa Barbara, CA, October 1997.
- [162] H.-H. YU and W. WOLF. « A Visual Search System for Video and Image Databases ». In *Proceedings of the International Conference on Multimedia Computing and Systems*, 1997.
- [163] R. ZABIH, J. MILLER, and K. MAI. « A feature-based algorithm for detecting and classifying production effects ». In *Multimedia System*, pages 119–128, 1999.
- [164] M.M. ZADEH, T. KASVAND, and C.Y. SUEN. « Localization and Recognition of Traffic Signs for Automated Vehicle Control Systems ». In *SPIE proceedings of Intelligent Transportation Systems (ITS)*, volume 3207, pages 272–282, Pittsburgh, PA, USA, 1997.

- [165] Li ZHAO and Yee-Hong YANG. « Mosaic image method : a local and global method ». *Pattern Recognition*, 32(8) :1421–1433, August 1999.



## **Troisième partie**

## **Annexes**



## Annexe A

# Influence du bruit gaussien sur les mesures d'angle

**Hypothèses.** On suppose que l'image  $I(x, y)$  se décompose comme la somme d'un signal déterministe  $d$  et un bruit centré gaussien :  $I(x, y) = d(x, y) + b(x, y)$ . La mesure d'angle du contour en  $(x, y)$  est calculée localement par :

$$\theta = \arctan\left(\frac{I_y}{I_x}\right)$$

avec  $I_x$  et  $I_y$  dérivées en x et y de  $I$ . On a :

$$\theta = \arctan\left(\frac{d_y + b_y}{d_x + b_x}\right)$$

$b_y$  et  $b_x$  sont des bruits gaussiens centrés de variance  $\sigma_b^2$ .

**Moyenne et variance de l'angle  $\theta$ .** Si on suppose  $\frac{b_x}{d_x} \ll 1$  alors :

$$\theta \simeq \arctan\left(\frac{d_y + b_y}{d_x} \left(1 - \frac{b_x}{d_x}\right)\right)$$

ou

$$\theta = \arctan\left(\frac{d_y}{d_x} + \frac{b_y}{d_x} - \frac{d_y b_x}{d_x^2} - \frac{b_y b_x}{d_x^2}\right)$$

ou

$$\theta \simeq \arctan\left(\frac{d_y}{d_x} + \frac{b_y}{d_x} - \frac{d_y b_x}{d_x^2}\right)$$

par développement limité en  $\frac{d_y}{d_x}$ , on obtient :

$$\theta \simeq \arctan\left(\frac{d_y}{d_x}\right) + \frac{1}{1 + \left(\frac{d_y}{d_x}\right)^2} \left(\frac{b_y}{d_x} - \frac{d_y b_x}{d_x^2}\right)$$

ou

$$\theta \simeq \arctan\left(\frac{d_y}{d_x}\right) + \frac{1}{d_y^2 + d_x^2} (b_y d_x - d_y b_x)$$

Donc

$$E[\theta] = \arctan\left(\frac{d_y}{d_x}\right) \quad (\text{A.1})$$

La variance se calcule par :

$$E[(\theta - E[\theta])^2] = E\left[\left(\frac{1}{d_x^2 + d_y^2} (b_y d_x - d_y b_x)\right)^2\right]$$

ou

$$E[(\theta - E[\theta])^2] = \frac{1}{(d_x^2 + d_y^2)^2} E[b_y^2 d_x^2 - 2 b_y b_x d_y d_x + d_y^2 b_x^2]$$

Donc

$$E[(\theta - E[\theta])^2] = \frac{1}{(d_x^2 + d_y^2)^2} (d_x^2 E[b_y^2] - 2 d_y d_x E[b_y b_x] + d_y^2 E[b_x^2])$$

avec  $b_x$  et  $b_y$  bruits gaussiens centrés tel que  $E[b_x^2] = \sigma_b^2$ ,  $E[b_y^2] = \sigma_b^2$  et  $E[b_x b_y] = 0$ . D'où

$$E[(\theta - E[\theta])^2] = \frac{\sigma_b^2}{(d_x^2 + d_y^2)} \quad (\text{A.2})$$

La moyenne de  $\theta$  est égale à la valeur attendue  $\arctan \frac{d_y}{d_x}$  et sa variance est inversement proportionnelle à la norme du gradient au carré. Ainsi sur une zone uniforme (norme du gradient nulle), la variance de  $\theta$  tend vers l'infini et sa distribution tend vers une distribution uniforme. Cette remarque est confirmée dans le paragraphe suivant.

**Cas d'une zone uniforme.** On pose  $z = \frac{I_y}{I_x}$ . Par définition, la densité de probabilité de  $z$  s'écrit :

$$f_z(z) = \int \int_{D_z} f_{I_x, I_y}(I_x, I_y) dI_x dI_y$$

où  $D_z = \{(I_x, I_y) \mid \frac{I_y}{I_x} \leq z\}$ . Après des changements de variables, la densité de probabilité de  $z$  se simplifie par [144] :

$$f_z(z) = \int_{-\infty}^{+\infty} |I_x| f_{I_x, I_y}(I_x, zI_x) dI_x$$

Les dérivées de l'image,  $I_x$  et  $I_y$ , sont indépendantes donc :

$$f_{I_x, I_y}(I_x, I_y) = f_{I_x}(I_x) f_{I_y}(I_y)$$

Les distributions  $f_{I_x}(I_x)$  et  $f_{I_y}(I_y)$  sont gaussiennes de variance  $\sigma_b^2$  et respectivement, de moyennes  $d_x$  et  $d_y$ . Sur les zones uniformes, on a  $d_x = 0$  et  $d_y = 0$ , d'où la densité de probabilité de  $z$  est :

$$f_z(z) = \frac{1}{2\pi\sigma_b^2} \int_{-\infty}^{+\infty} |I_x| \exp^{-\frac{(1+z^2)I_x^2}{2\sigma_b^2}} dI_x$$

ou

$$f_z(z) = \frac{1}{2\pi\sigma_b^2} 2 \left[ -2\sigma_b^2 \frac{\exp^{-\frac{(1+z^2)I_x^2}{2\sigma_b^2}}}{2(1+z^2)} \right]_0^{+\infty}$$

ou

$$f_z(z) = \frac{1}{\pi(1+z^2)}$$

L'angle  $\theta$  est défini par  $\theta = \arctan z$  d'où :

$$f_\theta(\theta) = f_z(\tan(\theta)) |\tan'(\theta)|$$

ou  $\forall \theta \in [-\frac{\pi}{2}; +\frac{\pi}{2}]$  :

$$f_\theta(\theta) = \frac{1}{\pi}$$

Sur une zone uniforme, la mesure  $\theta$  est distribuée uniformément.

**Conclusion.** Lorsque une image est bruitée par un bruit centré gaussien, on a montré que la variance de l'angle  $\theta$  calculé à la position  $(x, y)$ , est dépendante de la mesure de norme du gradient en  $(x, y)$ . De plus, lorsque l'on considère le cas particulier d'une zone uniforme de l'image, alors la distribution de la mesure d'angle est une distribution de probabilité uniforme.

Cela signifie que si l'on considère deux images  $I_1$  et  $I_2$  de la même scène :

$$\begin{aligned} I_1(x, y) &= d(x, y) + b_1(x, y) \\ I_2(x, y) &= d(x, y) + b_2(x, y) \end{aligned}$$

alors si  $(x, y)$  est une position sur une zone uniforme, alors les angles  $\theta_1(x, y)$  et  $\theta_2(x, y)$  calculés sur  $I_1$  et  $I_2$ , ont peu de chance de donner la même valeur. A contrario, si la position  $(x, y)$  est située sur un contour (i.e. telle que  $N(x, y) >> 0$ ), alors les angles  $\theta_1(x, y)$  et  $\theta_2(x, y)$  donneront sensiblement la même valeur.



## Annexe B

# Procédure d'extraction

## B.1 Notations et Hypothèses

On suppose que l'on a estimé les distributions  $\mathcal{P}(\mathbf{m}, t)$  et  $\mathcal{P}(\mathbf{m}, t + 1)$ , des mesures  $m$  calculées sur deux images  $I(t)$  et  $I(t + 1)$ .

### B.1.1 Hypothèses

$H_0$  : **Composition des images  $I(t)$ .** On suppose que chaque image  $I(t)$  de la séquence est composée d'un ensemble d'objets d'intérêt  $\mathcal{B}$  et d'un arrière plan  $\bar{\mathcal{B}}$ . La densité de probabilité  $\mathcal{P}(\mathbf{m}, t)$  des mesures locales  $\mathbf{m}$  dans l'image  $I(t)$  peut s'écrire :

$$\mathcal{P}(\mathbf{m}, t) = (1 - \epsilon(t)) \mathcal{P}(\mathbf{m}, t|\bar{\mathcal{B}}) + \epsilon(t) \mathcal{P}(\mathbf{m}, t|\mathcal{B})$$

$H_1$  : **Proportion de  $\mathcal{B}$ .** L'hypothèse  $H_1$  restreint le domaine de définition de la proportion  $\epsilon$  des objets et est liée au choix du paramètre  $\gamma$  dans les extracteurs.  $\epsilon$  doit appartenir à l'intervalle  $[0, \frac{\gamma-1}{\gamma}]$ .

$H_2$  : **Variations statistiques de  $\bar{\mathcal{B}}$ .** On suppose que les statistiques du fond  $\bar{\mathcal{B}}$  varient faiblement au cours du temps, soit :

$$\mathcal{P}(\mathbf{m}, t|\bar{\mathcal{B}}) \simeq \mathcal{P}(\mathbf{m}, t + 1|\bar{\mathcal{B}})$$

$H_3$  : **Différence statistique entre  $\mathcal{B}$  et  $\bar{\mathcal{B}}$ .** Cette hypothèse suppose que l'on peut choisir une mesure locale  $\mathbf{m}$  qui permet de distinguer les objets de  $\mathcal{B}$  du fond  $\bar{\mathcal{B}}$ , ou  $\forall \mathbf{m}$  :

$$\mathcal{P}(\mathbf{m}, t|\bar{\mathcal{B}}) \cdot \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \simeq 0$$

**$H_4$  : Variations statistiques de  $\mathcal{B}$ .** En pratique, nous sommes intéressés par les objets changeants qui apparaissent ou disparaissent de la scène, ou ceux qui ont des mesures locales qui changent significativement au cours du temps. Ainsi deux cas peuvent être envisagés. Le premier  $H_4$  considère que  $\mathcal{P}(\mathbf{m}, t|\mathcal{B})$  change fortement au cours du temps :

$$\forall \mathbf{m}, \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \cdot \mathcal{P}(\mathbf{m}, t+1|\mathcal{B}) \simeq 0$$

Le second cas  $H'_4$  suppose au contraire que :

$$\forall \mathbf{m}, \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \simeq \mathcal{P}(\mathbf{m}, t+1|\mathcal{B})$$

mais que c'est la proportion  $\epsilon(t)$  de l'objet qui varie *fortement*.

### B.1.2 Extracteurs

On pose :

$$\Delta_+(\mathbf{m}, t) = \max(\mathcal{P}(\mathbf{m}, t) - \gamma \mathcal{P}(\mathbf{m}, t+1), 0)$$

$$\Delta_-(\mathbf{m}, t) = \max(\mathcal{P}(\mathbf{m}, t) - \gamma \mathcal{P}(\mathbf{m}, t-1), 0)$$

avec  $\gamma > 0$ .

## B.2 Calcul de $\Delta_+(\mathbf{m}, t)$

Grâce à l'hypothèse  $H_0$ , on a :

$$\begin{aligned} \mathcal{P}(\mathbf{m}, t) - \gamma \mathcal{P}(\mathbf{m}, t+1) = & (1 - \epsilon(t)) \mathcal{P}(\mathbf{m}, t|\bar{\mathcal{B}}) - \gamma (1 - \epsilon(t+1)) \mathcal{P}(\mathbf{m}, t+1|\bar{\mathcal{B}}) \\ & + \epsilon(t) \mathcal{P}(\mathbf{m}, t|\mathcal{B}) - \gamma \epsilon(t+1) \mathcal{P}(\mathbf{m}, t+1|\mathcal{B}) \end{aligned} \tag{B.1}$$

Comme  $\mathcal{P}(\mathbf{m}, t+1|\bar{\mathcal{B}}) \simeq \mathcal{P}(\mathbf{m}, t|\bar{\mathcal{B}})$  par  $H_2$ , l'équation (B.1) devient :

$$\begin{aligned} \mathcal{P}(\mathbf{m}, t) - \gamma \mathcal{P}(\mathbf{m}, t+1) = & ((1 - \epsilon(t)) - \gamma (1 - \epsilon(t+1))) \mathcal{P}(\mathbf{m}, t|\bar{\mathcal{B}}) \\ & + \epsilon(t) \mathcal{P}(\mathbf{m}, t|\mathcal{B}) - \gamma \epsilon(t+1) \mathcal{P}(\mathbf{m}, t+1|\mathcal{B}) \end{aligned} \tag{B.2}$$

**Cas de  $H_4$ .** Les hypothèses  $H_3$  et  $H_4$  indiquent que si l'un des trois termes  $\mathcal{P}(\mathbf{m}, t|\bar{\mathcal{B}})$ ,  $\mathcal{P}(\mathbf{m}, t|\mathcal{B})$  et  $\mathcal{P}(\mathbf{m}, t+1|\mathcal{B})$ , est non nul, alors les autres sont nuls. Donc :

$$\begin{aligned} |\mathcal{P}(\mathbf{m}, t) - \gamma \mathcal{P}(\mathbf{m}, t+1)| &= |(1 - \epsilon(t)) - \gamma (1 - \epsilon(t+1))| \mathcal{P}(\mathbf{m}, t|\bar{\mathcal{B}}) \\ &\quad + \epsilon(t) \mathcal{P}(\mathbf{m}, t|\mathcal{B}) + \gamma \epsilon(t+1) \mathcal{P}(\mathbf{m}, t+1|\mathcal{B}) \end{aligned} \quad (\text{B.3})$$

D'après  $H_1$ ,  $(1 - \epsilon(t)) - \gamma (1 - \epsilon(t+1)) < 0$  d'où :

$$|1 - \epsilon(t) - \gamma (1 - \epsilon(t+1))| = \epsilon(t) - 1 + \gamma (1 - \epsilon(t+1)) \quad (\text{B.4})$$

Comme  $\max(a, 0) = \frac{a+|a|}{2}$ <sup>1</sup>, on a :

$$\Delta_+(\mathbf{m}, t) = \epsilon(t) \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \quad (\text{B.5})$$

**Cas de  $H'_4$ .** Dans ce cas  $\mathcal{P}(\mathbf{m}, t|\mathcal{B}) = \mathcal{P}(\mathbf{m}, t+1|\mathcal{B})$ , l'équation (B.2) devient :

$$\begin{aligned} \mathcal{P}(\mathbf{m}, t) - \gamma \mathcal{P}(\mathbf{m}, t+1) &= \\ &\quad ((1 - \epsilon(t)) - \gamma (1 - \epsilon(t+1))) \mathcal{P}(\mathbf{m}, t|\bar{\mathcal{B}}) \\ &\quad + (\epsilon(t) - \gamma \epsilon(t+1)) \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \end{aligned} \quad (\text{B.6})$$

et avec  $H_3$  :

$$\begin{aligned} |\mathcal{P}(\mathbf{m}, t) - \gamma \mathcal{P}(\mathbf{m}, t+1)| &= \\ &\quad |(1 - \epsilon(t)) - \gamma (1 - \epsilon(t+1))| \mathcal{P}(\mathbf{m}, t|\bar{\mathcal{B}}) \\ &\quad + |\epsilon(t) - \gamma \epsilon(t+1)| \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \end{aligned} \quad (\text{B.7})$$

Donc,

$$\Delta_+(\mathbf{m}, t) = \frac{\epsilon(t) - \gamma \epsilon(t+1) + |\epsilon(t) - \gamma \epsilon(t+1)|}{2} \mathcal{P}(\mathbf{m}, t|\mathcal{B}) \quad (\text{B.8})$$

- si  $\epsilon(t) > \gamma \epsilon(t+1)$  alors  $\Delta_+(\mathbf{m}, t) = (\epsilon(t) - \gamma \epsilon(t+1)) \mathcal{P}(\mathbf{m}, t|\mathcal{B})$ .
- si  $\epsilon(t) < \gamma \epsilon(t+1)$  alors  $\Delta_+(\mathbf{m}, t) = 0$ .

---

<sup>1</sup>  $\min(a, b) = \frac{a+b-|a-b|}{2}$



## Annexe C

# Expérience de pré-détection

Cette annexe présente les résultats de pré-détection obtenus sur les 100 premiers mètres (20 images) de la séquence Dep51 RN4PR 0 à 20. Cette séquence présente quelques variations d'intensité liées à l'instabilité du système d'acquisition. Les caractéristiques locales utilisées sont les mesures de couleurs et de formes définies au chapitre 3. Pour la première image de la séquence, seul le résultat de rétropénétration en arrière est disponible. Apparaissent en noir les positions des mesures dont la vraisemblance *a posteriori*,  $\mathcal{P}(\mathcal{B}|\mathbf{m})$  est supérieure à 0. Les pixels blancs indiquent des mesures locales de vraisemblance *a posteriori* nulle.

Les dernières lignes des images de la séquence sont instables et leurs variations statistiques sont perçues presque à chaque image. Dans une analyse ultérieure, ces lignes ne seront pas exploitées.

### C.1 Pré-détection avec des mesures de couleur

Cette section présente les résultats de pré-détection obtenus avec les mesures couleurs  $(r, g, L)$  et  $(r, g)$ . L'intérêt de l'utilisation d'histogrammes interpolés, ainsi que l'influence du paramètre  $\gamma$  y sont évalués.

#### C.1.1 Mesure $(r, g, L), \gamma = 2$

Dans cette première expérience, la mesure couleur  $(r, g, L)$  est utilisée et quantifiée en 20 classes pour chacune des composantes chromatiques  $r$  et  $g$ , et en 8 classes pour l'intensité  $L$ . Les histogrammes manipulés sont de dimension 3 et ont 3200 classes. Le tableau suivant présente les résultats de rétropénétration en avant et en arrière obtenus avec des histogrammes classiques et interpolés.

séquence	Rétroprojection par histogrammes classiques		Rétroprojection par histogrammes interpolés	
	<i>en arrière</i>	<i>en avant</i>	<i>en arrière</i>	<i>en avant</i>
0				
5				
10				
15				
20				
25				
30				

séquence	Rétroprojection par histogrammes classiques		Rétroprojection par histogrammes interpolés		
	<i>en arrière</i>	<i>en avant</i>	<i>en arrière</i>	<i>en avant</i>	
35					
40					
45					
50					
55					
60					
65					

séquence	Rétroprojection par histogrammes classiques		Rétroprojection par histogrammes interpolés		
	<i>en arrière</i>	<i>en avant</i>	<i>en arrière</i>	<i>en avant</i>	
70					
75					
80					
85					
90					

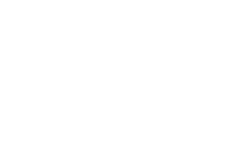
**Remarques :** Malgré un nombre de classes réduit sur la composante de luminance  $L$  par rapport aux chrominances, les changements d'intensité provoquent des fausses alarmes principalement sur la route

et le ciel (cf. images 65,70,75,...). On remarque que l'utilisation d'histogrammes interpolés permet de réduire sensiblement ces fausses détections sans affecter les performances de notre méthode. Aux images 40 et 75, le grossissement et la disparition de panneaux colorés sont parfaitement détectés. De même la disparition du véhicule à l'image 5 est détectée.

Afin de réduire l'influence des variations d'intensité du système d'acquisition, la composante de luminance n'est pas utilisée dans l'expérience suivante.

### C.1.2 Mesure $(r, g), \gamma = 2$

Dans cette expérience, seules les chromaticités  $(r, g)$  sont utilisées. Elles sont toujours quantifiées en 20 classes. Les histogrammes sont à 2 dimensions, et comportent 400 classes.

séquence	Rétroprojection par histogrammes classiques		Rétroprojection par histogrammes interpolés		
	<i>en arrière</i>	<i>en avant</i>	<i>en arrière</i>	<i>en avant</i>	
0					
5					
10					
15					

séquence	Rétroprojection par histogrammes classiques		Rétroprojection par histogrammes interpolés		
	<i>en arrière</i>	<i>en avant</i>	<i>en arrière</i>	<i>en avant</i>	
20					
25					
30					
35					
40					
45					
50					

séquence	Rétroprojection par histogrammes classiques		Rétroprojection par histogrammes interpolés		
	<i>en arrière</i>	<i>en avant</i>	<i>en arrière</i>	<i>en avant</i>	
55					
60					
65					
70					
75					
80					
85					

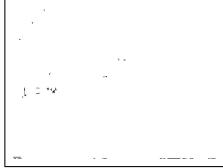
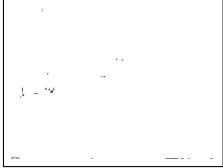
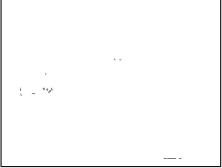
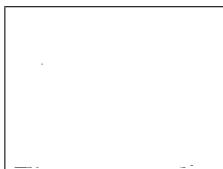
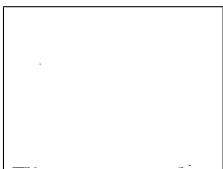
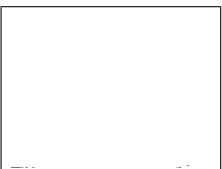
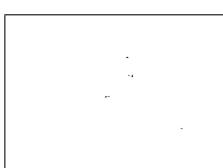
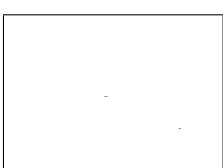
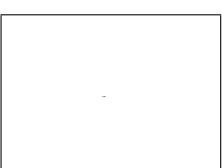
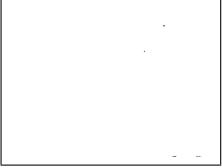
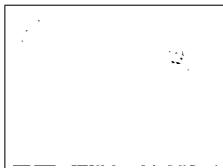
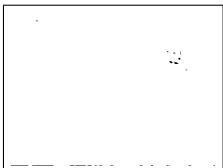
séquence	Rétroprojection par histogrammes classiques		Rétroprojection par histogrammes interpolés		
	<i>en arrière</i>	<i>en avant</i>	<i>en arrière</i>	<i>en avant</i>	
90					
95					

**Remarques :** Quelques fausses détections demeurent sur les cartes de rétroprojection 40 et 70 calculées avec des histogrammes classiques. Celles-ci disparaissent avec les histogrammes interpolés. On remarque que la disparition de la voiture blanche à l'image  $I_5$  n'est pas détectée avec les seules mesures de chrominances. Par contre, cet évènement est parfaitement détecté dans le cas de mesures couleur  $(r, g, L)$ .

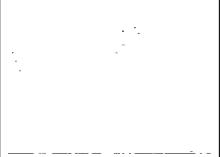
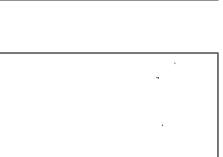
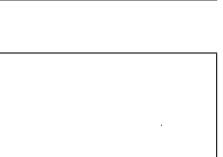
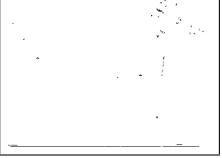
### C.1.3 Fusion d'information et effet de $\gamma$

Le tableau suivant présente les résultats de pré-détection obtenus en fusionnant les cartes de rétroprojection en avant et en arrière. Une analyse en composantes connexes est réalisée, limitant les régions détectées à celles dont la taille est supérieure à 10 pixels. L'expérience est effectuée pour trois valeurs différentes du paramètre  $\gamma$ . Ce paramètre n'a aucune incidence sur les résultats lorsque l'hypothèse  $H_4$  est vérifiée. Au contraire, dans le cas de l'hypothèse  $H'_4$  où la variation statistique est due à une variation de la proportion  $\epsilon$ , il est intéressant d'étudier l'influence du choix de  $\gamma$  sur les résultats.

séquence	$\gamma = 1.5$	$\gamma = 2$	$\gamma = 3$	
0				

séquence	$\gamma = 1.5$	$\gamma = 2$	$\gamma = 3$
5			
10			
15			
20			
25			
30			
35			

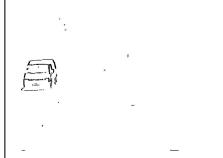
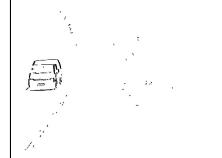
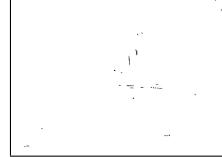
séquence	$\gamma = 1.5$	$\gamma = 2$	$\gamma = 3$	
40				
45				
50				
55				
60				
65				
70				

séquence	$\gamma = 1.5$	$\gamma = 2$	$\gamma = 3$	
75				
80				
85				
90				
95				

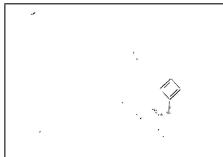
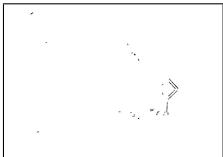
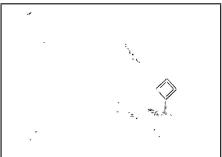
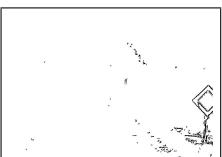
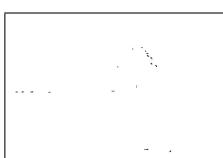
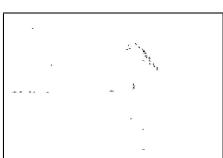
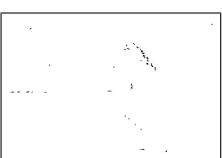
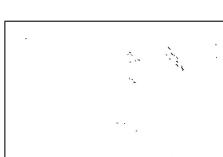
**Remarques :** le paramètre  $\gamma$  permet de contrôler l'amplitude des variations statistiques que l'on souhaite détecter. Pour  $\gamma = 3$ , le grossissement d'un objet peut être détecté seulement si le volume de ses mesures locales augmente d'un facteur supérieur à 3. Cela explique pourquoi le grossissement du panneau de l'image 70 n'est pas détecté. Prendre une valeur grande pour  $\gamma$  limite l'influence des fausses détections liées aux pixels mixtes.

## C.2 Pré-détection avec des mesures de forme

La mesure de forme utilisée est  $(\theta, \alpha, N)$ . L'angle  $\theta$  est quantifié en 21 classes, l'alignement  $\alpha$  sur 100 classes et la norme  $N$  en 10 classes. Les histogrammes de forme sont à 3 dimensions et comportent 21000 classes. Le tableau suivant présente les résultats de pré-détection obtenus avec des histogrammes interpolés. Les cartes de rétroposition en arrière et en avant sont fusionnées et seules les régions de taille supérieure à 10 pixels sont conservées.

séquence	<i>en arrière</i>	<i>en avant</i>	fusion
0			
5			
10			
15			
20			

séquence	<i>en arrière</i>	<i>en avant</i>	fusion
25			
30			
35			
40			
45			
50			
55			

séquence	<i>en arrière</i>	<i>en avant</i>	fusion	
60				
65				
70				
75				
80				
85				
90				

séquence	<i>en arrière</i>	<i>en avant</i>	fusion
			

95

**Remarques :** Les contours des objets d'intérêt comme les poteaux, les troncs et feuillage d'arbres, sont correctement détectés. Avec les mesures couleurs, la détection d'un objet présent dans deux images successives n'est due qu'à la variation de la proportion  $\epsilon$  de ses pixels (hypothèse  $H'_4$ ) induit par le changement d'échelle. La détection d'un objet présent dans deux images successives par les mesures de forme est due au changement de la mesure d'alignement  $\alpha$  qui dépend de la position de l'objet dans l'image (hypothèse  $H_4$ ).



## Annexe D

# Rotation dans le plan image

L'article de Uenohara et Kanade [155] propose des expressions analytiques pour les valeurs propres et les vecteurs propres calculés sur une base de  $K$  images centrées d'un même objet appris avec ses rotations dans le plan de l'image, notée  $\mathcal{B} = \{\tilde{x}^0, \dots, \tilde{x}^{K-1}\}$ . L'image  $\tilde{x}^k$ ,  $k \in \{0, \dots, K-1\}$ , est générée à partir de l'image  $\tilde{x}^0$  par une rotation d'angle  $\frac{2\pi k}{K}$  dans le plan de l'image. Dans le premier paragraphe, les résultats principaux de cet article sont rappelés avec nos notations. La seconde section établit la distribution particulière des projections des images d'apprentissage dans l'espace propre.

## D.1 Valeurs propres et vecteurs propres

En utilisant la décomposition en valeurs singulières (cf. paragraphe 5.1.1.2), la matrice  $Q = \tilde{X}^T \tilde{X}$  de dimension  $K \times K$  est diagonalisée. C'est une matrice Toeplitz et symétrique d'où  $q_{K-i} = q_i$  :

$$Q = \begin{pmatrix} q_0 & q_1 & \cdots & q_{K-1} \\ q_{K-1} & q_0 & \cdots & q_{K-2} \\ \cdots & \cdots & q_0 & \cdots \\ q_1 & q_2 & \cdots & q_0 \end{pmatrix} \quad (\text{D.1})$$

La valeur propre  $\lambda_k$  s'écrit [155] :

$$\lambda_k = \sum_{l=0}^{K-1} q_l e^{j \frac{k2\pi l}{K}} \quad (\text{D.2})$$

$Q$  est symétrique d'où  $q_{K-i} = q_i$  et :

- Si  $K$  est pair :

$$\lambda_k = q_0 + q_{\frac{K}{2}} (-1)^k + \sum_{l=1}^{\frac{K}{2}-1} q_l \left( e^{j \frac{k2\pi l}{K}} + e^{-j \frac{k2\pi l}{K}} \right)$$

- Si  $K$  impair :

$$\lambda_k = q_0 + \sum_{l=1}^{\frac{K-1}{2}} q_l \left( e^{j\frac{k2\pi l}{K}} + e^{-j\frac{k2\pi l}{K}} \right)$$

Comme  $e^{j\frac{k2\pi l}{K}} + e^{-j\frac{k2\pi l}{K}} = 2 \cos(\frac{k2\pi l}{K})$ , on a  $\lambda_k \in \mathbb{R}$  et :

$$\lambda_k = \Re(\lambda_k) = \sum_{l=0}^{K-1} q_l \cos\left(\frac{k2\pi l}{K}\right) \quad (\text{D.3})$$

On rappelle que les vecteurs propres d'une matrice Toeplitz sont des vecteurs complexes de Fourier [12], dont la composante  $m$  peut être définie par :  $\mathbf{v}_{km} = e^{j(\frac{k2\pi m}{K})}$ . Comme sa valeur propre  $\lambda_k$  est réelle, alors  $\lambda_k$  est une valeur propre double associée à deux vecteurs propres réels orthogonaux  $\mathbf{v}_k^a = \Re(\mathbf{v}_k)$  et  $\mathbf{v}_k^b = \Im(\mathbf{v}_k)$ , dont la composante  $m$  s'écrit :

$$\begin{cases} v_{km}^a = \cos\left(\frac{k2\pi m}{K}\right) \\ v_{km}^b = \sin\left(\frac{k2\pi m}{K}\right) \end{cases}$$

Soit  $\mathbf{u}_k^a$  et  $\mathbf{u}_k^b$  deux vecteurs propres de la matrice de covariance  $\Sigma = \tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$  associés à la valeur propre double  $\lambda_k$ , on a :

$$\begin{cases} \mathbf{u}_k^a = \frac{1}{\sqrt{\lambda_k}} \tilde{\mathbf{X}} \mathbf{v}_k^a = \frac{1}{\sqrt{\lambda_k}} \sum_{m=0}^{K-1} v_{km}^a \tilde{\mathbf{x}}^m \\ \mathbf{u}_k^b = \frac{1}{\sqrt{\lambda_k}} \tilde{\mathbf{X}} \mathbf{v}_k^b = \frac{1}{\sqrt{\lambda_k}} \sum_{m=0}^{K-1} v_{km}^b \tilde{\mathbf{x}}^m \end{cases}$$

## D.2 Distribution des projections des modèles dans l'espace propre

Nous montrons dans cette section que les projections des modèles d'apprentissage sont répartis suivant un cercle dans chaque plan propre associé à une valeur propre double. Soit  $\mathbf{y}$  une image quelconque, on pose :

$$z_k(\mathbf{y}) = \langle \mathbf{y} | \mathbf{u}_k^a \rangle + j \langle \mathbf{y} | \mathbf{u}_k^b \rangle$$

$z_k(\mathbf{y})$  est l'affixe complexe de la projection de l'image  $\mathbf{y}$  sur le plan associé à la valeur propre double  $\lambda_k$ , dans le repère défini par les vecteurs propres  $(\mathbf{u}_k^a, \mathbf{u}_k^b)$ . On peut réécrire  $z_k(\mathbf{y})$  :

$$z_k(\mathbf{y}) = \frac{1}{\sqrt{\lambda_k}} \sum_{m=0}^{K-1} \langle \mathbf{y} | \tilde{\mathbf{x}}^m \rangle e^{j\frac{2\pi km}{K}}$$

Pour les images d'apprentissage centrées  $\tilde{\mathbf{x}}$ , nous montrons d'abord que :

$$z_k(\tilde{\mathbf{x}}^{l+1}) = e^{j\frac{2\pi k}{K}} z(\tilde{\mathbf{x}}^l) \quad (\text{D.4})$$

On a :

$$z_k(\tilde{\mathbf{x}}^{l+1}) = \frac{1}{\sqrt{\lambda_k}} \sum_{m=0}^{K-1} \langle \tilde{\mathbf{x}}^{l+1} | \tilde{\mathbf{x}}^m \rangle e^{j\frac{2\pi km}{K}}$$

Or les images d'apprentissage ont la propriété [155] :

$$\left| \begin{array}{l} \forall m > 0, \langle \tilde{\mathbf{x}}^{l+1} | \tilde{\mathbf{x}}^m \rangle = \langle \tilde{\mathbf{x}}^l | \tilde{\mathbf{x}}^{m-1} \rangle \\ \text{et, } \langle \tilde{\mathbf{x}}^{l+1} | \tilde{\mathbf{x}}^0 \rangle = \langle \tilde{\mathbf{x}}^l | \tilde{\mathbf{x}}^{K-1} \rangle \end{array} \right.$$

d'où

$$\begin{aligned} z_k(\tilde{\mathbf{x}}^{l+1}) &= \frac{1}{\sqrt{\lambda_k}} \left( \sum_{m=1}^{K-1} \langle \tilde{\mathbf{x}}^l | \tilde{\mathbf{x}}^{m-1} \rangle e^{j\frac{2\pi km}{K}} + \langle \tilde{\mathbf{x}}^l | \tilde{\mathbf{x}}^{K-1} \rangle \right) \\ &= \frac{e^{j\frac{2\pi k}{K}}}{\sqrt{\lambda_k}} \left( \sum_{m=1}^{K-1} \langle \tilde{\mathbf{x}}^l | \tilde{\mathbf{x}}^{m-1} \rangle e^{j\frac{2\pi k(m-1)}{K}} + \langle \tilde{\mathbf{x}}^l | \tilde{\mathbf{x}}^{K-1} \rangle e^{j\frac{2\pi k(K-1)}{K}} \right) \\ &= \frac{e^{j\frac{2\pi k}{K}}}{\sqrt{\lambda_k}} \sum_{m=0}^{K-1} \langle \tilde{\mathbf{x}}^l | \tilde{\mathbf{x}}^m \rangle e^{j\frac{2\pi km}{K}} \\ &= e^{j\frac{2\pi k}{K}} z(\tilde{\mathbf{x}}^l) \end{aligned}$$

L'équation (D.4) est vraie et peut être généralisée par :

$$z_k(\tilde{\mathbf{x}}^l) = \left( e^{j\frac{2\pi k}{P}} \right)^l z_k(\tilde{\mathbf{x}}^0) \quad (\text{D.5})$$

L'affixe des projections des images d'apprentissage dans les plans propres associés à une valeur propre double est liée par une rotation à l'affixe de l'image  $\tilde{\mathbf{x}}^0$ . Ces projections sont donc réparties suivant un cercle dans les plans propres associés à une valeur propre double.

**Rayon  $R_k$ .** Posons  $c_k^a(l)$  et  $c_k^b(l)$  les coordonnées du modèle  $\tilde{\mathbf{x}}^l$  suivant les vecteurs propres  $(\mathbf{u}_k^a, \mathbf{u}_k^b)$ .

On a :

$$\begin{aligned} E[(c_k^a)^2] &= \frac{1}{K} \sum_{l=0}^{K-1} (c_k^a(l))^2 = \lambda_k \\ E[(c_k^b)^2] &= \frac{1}{K} \sum_{l=0}^{K-1} (c_k^b(l))^2 = \lambda_k \end{aligned}$$

Soit  $R_k$  le rayon du cercle suivant lequel les projections des modèles sont distribués dans le plan propre associé à la valeur propre double  $\lambda_k$ . On a :

$$R_k^2 = E[(c_k^a)^2 + (c_k^b)^2] = 2\lambda_k \quad (\text{D.6})$$



## Annexe E

# Preuve de la convexité de $J_{HU}(\mathbf{c}, \sigma_\rho)$

On propose de démontrer que l'énergie définie par Huber pour les fonctions convexes  $\rho$ , a un seul minimum en  $(\mathbf{c}, \sigma_\rho)$

$$J_{HU}(\mathbf{c}, \sigma_\rho) = \frac{1}{n} \sum_{n=1}^N \left[ \rho\left(\frac{\epsilon_n}{\sigma_\rho}\right) + a \right] \cdot \sigma_\rho$$

On rappelle que le paramètre d'échelle  $\sigma_\rho$  est positif. De la même façon que Huber dans [62] p.178, pour démontrer que  $J_{HU}$  est convexe en  $(\mathbf{c}, \sigma_\rho)$ , nous supposons que les variables  $(\mathbf{c}, \sigma_\rho)$  dépendent linéairement d'un paramètre  $t$  et nous calculons la dérivée seconde de  $J_{HU}$  par rapport à  $t$ . En omettant l'indice  $n$ , on pose :

$$q(t) = \sigma_\rho \cdot \rho(\epsilon/\sigma_\rho) + a \cdot \sigma_\rho$$

En dérivant par rapport à  $t$ , on obtient :

$$\dot{q}(t) = \dot{\sigma}_\rho \cdot \rho\left(\frac{\epsilon}{\sigma_\rho}\right) + \rho'\left(\frac{\epsilon}{\sigma_\rho}\right) \cdot \left(\dot{\epsilon} - \frac{\epsilon \dot{\sigma}_\rho}{\sigma_\rho}\right) + a \cdot \dot{\sigma}_\rho$$

et

$$\ddot{q}(t) = \ddot{\epsilon} \cdot \rho'\left(\frac{\epsilon}{\sigma_\rho}\right) + \frac{1}{\sigma_\rho} \rho''\left(\frac{\epsilon}{\sigma_\rho}\right) \cdot \left(\dot{\epsilon} - \frac{\epsilon \dot{\sigma}_\rho}{\sigma_\rho}\right)^2$$

La fonction  $\rho$  est une fonction croissante et convexe sur  $\mathbb{R}^+$  d'où :

$$\begin{cases} \rho'(\epsilon/\sigma) \geq 0 \\ \rho''(\epsilon/\sigma) \geq 0 \end{cases}$$

**Signe de  $\ddot{\epsilon}$  dans le cas des images en niveaux de gris.** Dans ce cas, le résidu  $\epsilon$  est linéaire en  $\mathbf{c}$  et donc linéaire en  $t$ . Cela implique :  $\ddot{\epsilon} = 0$ .

**Signe de  $\ddot{\epsilon}$  dans le cas des images multibandes.** Dans ce cas,  $\epsilon = \|w(\mathbf{c})\|$  avec  $w(\mathbf{c})$  linéaire en  $\mathbf{c}$ , lui-même linéaire en  $t$ . On note  $f(t) = w \circ \mathbf{c}(t)$  l'application linéaire de  $\mathbb{R}$  dans  $\mathbb{R}^P$ . On a :

$$\epsilon = \sqrt{< f | f >}$$

Il vient :

$$\dot{\epsilon} = \frac{<\dot{f}|f>}{\epsilon}$$

et  $\ddot{\epsilon}$  s'exprime :

$$\ddot{\epsilon} = \frac{\|f\|^2 \|\dot{f}\|^2 - <\dot{f}|f>^2}{\epsilon^3}$$

car  $\ddot{f}$  est un vecteur nul. On a :

$$| <\dot{f}|f> | \leq \|\dot{f}\| \|f\|$$

d'où  $\ddot{\epsilon} \geq 0$ .

**Conclusion.** On a  $\ddot{q} \leq 0$  d'où  $J_{HU}$  est convexe en  $(\mathbf{c}, \sigma_\rho)$ .

## Annexe F

# Liens entre la théorie semi-quadratique et l'approche de Huber

On considère que la fonction  $\rho$  vérifie les conditions suivantes :

- $\rho$  symétrique, croissante sur  $\mathbb{R}^+$ .
- $\beta^\omega(t) = \frac{\rho'(t)}{2t}$  décroissante sur  $\mathbb{R}^+$  et bornée :

$$\begin{cases} \lim_{t \rightarrow \infty} \beta^\omega(t) = 0 \\ \lim_{t \rightarrow 0} \beta^\omega(t) = 1 \end{cases}$$

Cette annexe montre que la fonction  $w \rightarrow \rho(\|w\|)$  définie de  $\mathbb{R}^P$  dans  $\mathbb{R}$  peut s'écrire :

✓ ARTUR

$$\rho(\|w\|) = \inf_{u \in \mathbb{R}^P} \left\{ Q^\omega(w, u) = \frac{\rho'(\|u\|)}{2\|u\|} \|w\|^2 + \rho(\|u\|) - \frac{\rho'(\|u\|)}{2} \|u\| \right\}$$

et que l'inf est atteint pour la valeur :  $u = w$ .

✓ LEGEND

$$\rho(\|w\|) = \inf_{u \in \mathbb{R}^P} \left\{ Q^r(w, u) = \left\| w - u \left( 1 - \frac{\rho'(\|u\|)}{2\|u\|} \right) \right\|^2 + \rho(\|u\|) - \left( \frac{\rho'(\|u\|)}{2} \right)^2 \right\}$$

et que l'inf est atteint pour la valeur  $u = w$ .

---

Il est facile de vérifier que  $Q^\omega(w, w) = \rho(\|w\|)$  et  $Q^r(w, w) = \rho(\|w\|)$ . Il reste à montrer que les énergies  $Q^\omega(w, u)$  et  $Q^r(w, u)$  sont supérieures à  $\rho(\|w\|)$ ,  $\forall (w, u) \in \mathbb{R}^P \times \mathbb{R}^P$ . La démonstration de Huber est rappelée à la section F.1 dans le cas mono dimensionnel  $P = 1$ . L'extension à des variables multidimensionnelles  $P > 1$ , est réalisée dans la section F.2.

## F.1 Cas des images en niveaux de gris

On rappelle que Huber définit deux polynômes,  $q^\omega(t)$  et  $q^r(t)$ , qui vérifient les conditions en  $v \in \mathbb{R}^+$  :

$$\left| \begin{array}{l} q^*(v) = \rho(v) \\ q^{*\prime}(v) = \rho'(v) \end{array} \right. \quad (\text{F.1})$$

$q^*$  désignant les polynômes  $q^\omega$  ou  $q^r$ .

### F.1.1 Montrons que $\rho(t) \leq q^\omega(t) \leq q^r(t)$

On montre que les polynômes proposés par Huber, sont situés au-dessus de la courbe de la fonction  $\rho$  (cf. figure 6.2).

#### F.1.1.1 Montrons que $\rho(t) \leq q^\omega(t)$

Le polynôme  $q^\omega$  est défini par :  $q^\omega(t) = \beta^\omega t^2 + \alpha^\omega$  avec :

$$\left| \begin{array}{l} \alpha^\omega = \rho(v) - \frac{\rho'(v)}{2} v \\ \beta^\omega = \frac{\rho'(v)}{2 v} \end{array} \right.$$

$q^\omega$  vérifie les conditions F.1 en  $v$  et est une fonction paire. On pose  $Z_1(t) = q^\omega(t) - \rho(t)$ . On voit immédiatement que :  $Z_1(v) = 0$ . D'autre part :

$$\begin{aligned} Z'_1(t) &= q^{\omega'}(t) - \rho'(t) \\ &= \frac{\rho'(v)}{v} t - \rho'(t) \\ &= t \left( \frac{\rho'(v)}{v} - \frac{\rho'(t)}{t} \right) \end{aligned} \quad (\text{F.2})$$

D'après les propriétés de  $\rho$ , la fonction  $\frac{\rho'(t)}{t}$  est symétrique et décroissante sur  $\mathbb{R}^+$ . Cela implique :

- o  $\forall t \in [0; v], Z'_1(t) \leq 0$ . Donc  $Z_1$  est décroissante sur  $[0; v]$ ,
- o  $\forall t \in [v; +\infty], Z'_1(t) \geq 0$ . Donc  $Z_1$  est croissante sur  $[v; +\infty]$ .

d'où  $\forall t \in \mathbb{R}^+, Z_1(t) \geq Z_1(v) = 0$ . La fonction  $Z_1$  est paire, donc  $\forall t \in \mathbb{R}$  :

$$\rho(t) \leq q^\omega(t)$$

et l'inf est atteint en  $v$  :  $\rho(v) = q^\omega(v)$ .

### F.1.1.2 Montrons que $q^\omega(t) \leq q^r(t)$

Le polynôme  $q^r$  est défini par  $q^r(t) = (t - \beta^r)^2 + \alpha^r$  avec :

$$\left| \begin{array}{l} \alpha^r = \rho(v) - \left( \frac{\rho'(v)}{2} \right)^2 \\ \beta^r = v \left( 1 - \frac{\rho'(v)}{2v} \right) \end{array} \right.$$

$q^r$  vérifie les conditions F.1 en  $v$ . On pose :

$$Z_2(t) = q^r(t) - q^\omega(t)$$

Comme  $q^r(v) = \rho(v)$ , alors  $Z_2(v) = 0$ . La dérivée de  $Z_2$  en  $t$  s'écrit :

$$Z'_2(t) = 2t - 2v + \rho'(v) - \frac{\rho'(v)}{v} t \quad (\text{F.3})$$

On remarque que  $Z'_2(v) = 0$ . La dérivée seconde de  $Z_2$  est :

$$Z''_2(t) = 2 - \frac{\rho'(v)}{v} \quad (\text{F.4})$$

D'après les propriétés de  $\rho$ , la fonction  $\beta^\omega(t) = \frac{\rho'(t)}{2t}$  a ses valeurs dans l'intervalle  $[0; 1]$  d'où  $Z''_2(t) \geq 0$ .

Alors la fonction  $Z'_2$  est croissante sur  $\mathbb{R}$ . Or  $Z'_2(v) = 0$  d'où :

- ✓  $\forall t \leq v, Z'_2(t) \leq 0$ . Donc  $Z_2$  est décroissante sur  $]-\infty; v[$ .
- ✓  $\forall t \geq v, Z'_2(t) \geq 0$ . Donc  $Z_2$  est croissante sur  $]v; +\infty[$ .

Cela implique :  $\forall t \in \mathbb{R}, Z_2(t) \geq Z_2(v) = 0$ . Donc  $\forall t \in \mathbb{R}$  :

$$q^\omega(t) \leq q^r(t)$$

## F.1.2 Remarques

Les polynômes  $q^\omega$  et  $q^r$  ont été définis pour vérifier des conditions en  $v \in \mathbb{R}$ . On propose de définir deux fonctions à deux variables  $t$  et  $v$ , notées  $Q^\omega$  et  $Q^r$ , définies par :

$$\left| \begin{array}{l} Q^\omega(t, v) = \rho(v) - \frac{\rho'(v)}{2} v + \frac{\rho'(v)}{2v} t^2 \\ Q^r(t, v) = \rho(v) - \left( \frac{\rho'(v)}{2} \right)^2 - \left( t - v + \frac{\rho'(v)}{2} \right)^2 \end{array} \right.$$

Alors on a montré à la section précédente que  $\forall (t, v) \in \mathbb{R} \times \mathbb{R}, \rho(t) \leq Q^\omega(t, v) \leq Q^r(t, v)$  et  $\rho(t) = Q^\omega(t, t) = Q^r(t, t)$ . Donc, on peut écrire :

$$\left| \begin{array}{l} \rho(t) = \inf_v \{Q^\omega(t, v)\} \\ \rho(t) = \inf_v \{Q^r(t, v)\} \end{array} \right. \quad (\text{F.5})$$

et l'inf est atteint pour  $v = t$ .

**Lien avec la théorie semi-quadratique.** En introduisant le changement de variables :

$$\left| \begin{array}{l} b^A = \frac{\rho'(v)}{2v} \\ b^L = v \left( 1 - \frac{\rho'(v)}{2v} \right) \end{array} \right.$$

l'équation F.5 devient :

$$\left| \begin{array}{l} \rho(t) = \inf_{b^A} \{ \rho^A(t, b^A) = b^A t^2 + \Psi(b^A) \} \\ \rho(t) = \inf_{b^L} \{ \rho^L(t, b^L) = (t - b^L)^2 + \xi(b^L) \} \end{array} \right.$$

avec les fonctions  $\Psi$  et  $\xi$  définies paramétriquement (via  $v$ ) par :

$$\left| \begin{array}{l} b^A = \frac{\rho'(v)}{2v} \\ \Psi(b^A) = \rho(v) - \frac{\rho'(v)}{2} v \end{array} \right.$$

et

$$\left| \begin{array}{l} b^L = v \left( 1 - \frac{\rho'(v)}{2v} \right) \\ \xi(b^L) = \rho(v) - \left( \frac{\rho'(v)}{2} \right)^2 \end{array} \right. \quad (\text{F.6})$$

Obtenir des expressions explicites de  $\Psi(b^A)$  et  $\xi(b^L)$ , en fonction des variables  $b^A$  et  $b^L$  nécessite le calcul des fonctions inverses  $(b^A(v))^{-1}$  et  $(b^L(v))^{-1}$ . Pour la fonction  $\Psi$ , la fonction de pondération  $b^A(v) = \frac{\rho'(v)}{2v}$  n'est inversible que si on a la condition supplémentaire :  $b^A(v)$  strictement décroissante sur  $\mathbb{R}^+$ . Cette condition est présente dans les propriétés 1 (page 107) utilisées par la théorie semi-quadratique. Cette condition est vérifiée pour les fonctions HS, HL, et GM utilisées dans cette thèse. Le tableau F.1 présente les fonctions HS, HL, et GM avec l'expression littérale de leur fonction  $\Psi$ . Le calcul de la fonction inverse de  $\xi$  peut être complexe, aussi on utilise sa définition paramétrique (cf. équation F.6).

La figure F.1 présente les variables auxiliaires introduites par la régularisation semi-quadratique,  $b^A$  et  $b^L$  calculées en fonction de l'erreur  $w \in \mathbb{R}$ . Les fonctions  $\Psi(b^A)$  et  $\xi(b^A)$ , calculées pour la fonction de Geman et Mc Clure, sont également présentées.

Une fonction couramment utilisée pour la M-estimation, pour laquelle la condition d'inversibilité de la fonction de pondération n'est pas vérifiée, est la fonction de Huber [62] :

$$\rho(t) = \min(t^2, 2t - 1)$$

Sa fonction de pondération est  $\beta^\omega$  est définie par :

$$\beta^\omega(t) = \begin{cases} 1 & \text{si } t \leq 1 \\ \frac{1}{t} & \text{si } t > 1 \end{cases}$$

Le rapprochement entre les méthodes de Huber et la théorie semi-quadratique permet de proposer des expressions paramétriques des fonctions  $\Psi$  et  $\xi$ . On rappelle que la seule définition de la fonction  $\xi$  proposée dans le cadre de la régularisation semi quadratique, est  $\xi(b^L) = \sup_t [\rho(t) - (t - b^L)^2]$  [25].

	$\rho(t)$	$b^A(t)$	$\Psi(b^A)$	$\xi(b^L)$	convexité de $\rho$
HS	$2\sqrt{1+t^2} - 2$	$\frac{1}{\sqrt{1+t^2}}$	$b^A + \frac{1}{b^A} - 2$	$\xi(b^L) = \rho(t) - b^A(t) t^2$ $b^L = t (1 - b^A(t))$	convexe
HL	$\ln(1+t^2)$	$\frac{1}{1+t^2}$	$b^A - \ln(b^A) - 1$	"	non convexe
GM	$\frac{t^2}{1+t^2}$	$\frac{1}{(1+t^2)^2}$	$b^A - 2\sqrt{b^A} + 1$	"	non convexe

TAB. F.1: Expressions des fonctions robustes  $\rho$  utilisées, et des fonctions des variables auxiliaires  $\Psi$  et  $\xi$ .

## F.2 Extension aux images multibandes

Dans le cas des images multi-bandes, on cherche deux expressions permettant d'approcher la fonction  $w \rightarrow \rho(\|w\|)$  définie de  $\mathbb{R}^P$  dans  $\mathbb{R}$ . Sur le même modèle que la section précédente, deux fonctions sont définies :

$$q^\omega(w) = \beta^\omega \|w\|^2 + \alpha^\omega$$

et

$$q^r(w) = \|w - \beta^r\|^2 + \alpha^r$$

On suppose que  $w$  dépend linéairement d'un paramètre  $x$ . La dérivée en  $x$  de  $w$  est notée  $\dot{w}$ . Les fonctions  $q^\omega$  et  $q^r$  sont définies de  $\mathbb{R}^P$  dans  $\mathbb{R}$  et vérifient en  $u \in \mathbb{R}^P$  les conditions suivantes :

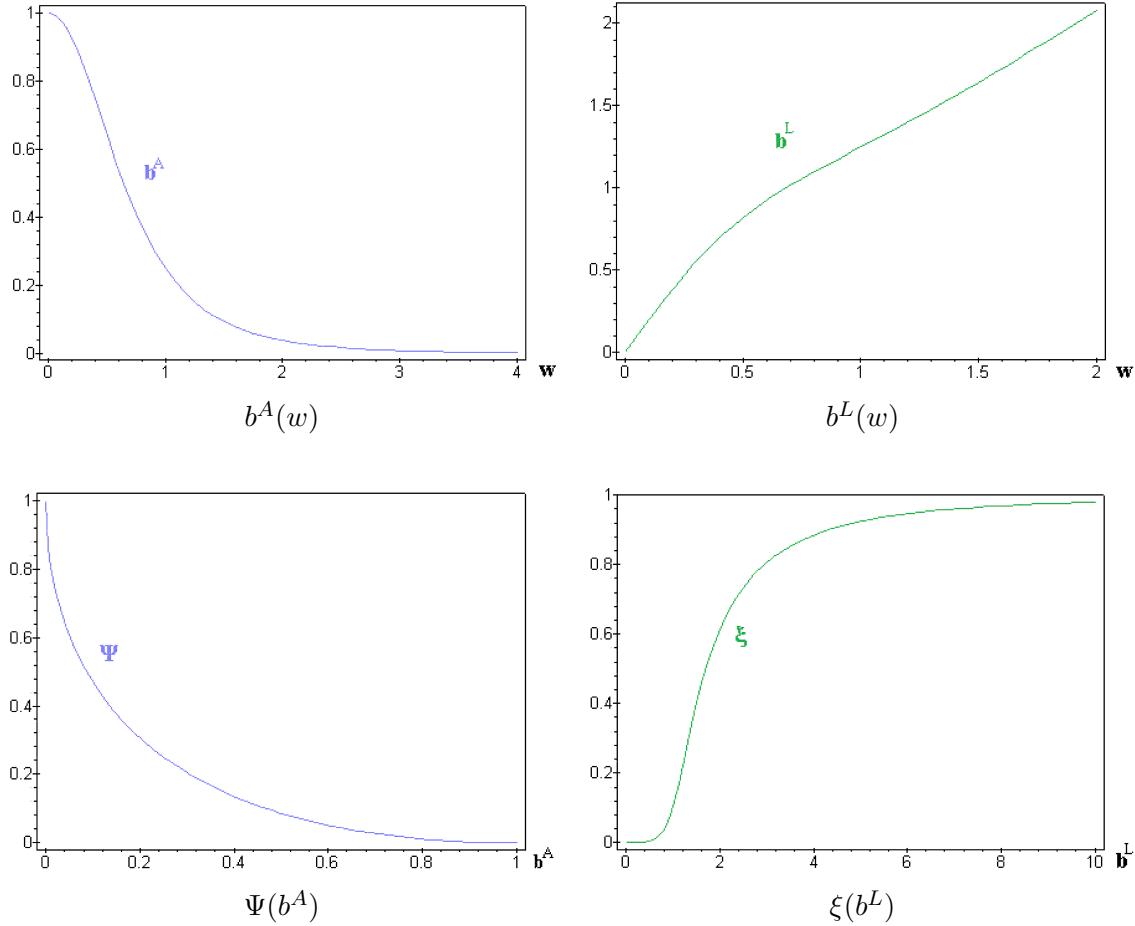
$$\begin{aligned} q^\omega(u) &= q^r(u) = \rho(\|u\|) \\ \dot{q}^\omega(u) &= \dot{q}^r(u) = \dot{\rho}(\|u\|) = \frac{\langle \dot{u} | u \rangle}{\|u\|} \rho'(\|u\|) \end{aligned} \tag{F.7}$$

où  $\dot{u}$ ,  $\dot{q}^\omega$ ,  $\dot{q}^r$  et  $\dot{\rho}$  sont les dérivées en  $x$ .

### F.2.1 Montrons que $\rho(\|w\|) \leq q^\omega(w)$

La fonction  $q^\omega$  est définie de  $\mathbb{R}^P$  dans  $\mathbb{R}$  par :

$$q^\omega(w) = \beta^\omega \|w\|^2 + \alpha^\omega$$

FIG. F.1:Fonctions  $b^A(w)$ ,  $b^L(w)$ ,  $\Psi(b^A)$  et  $\xi(b^L)$  de la fonction de Geman et Mc Clure (cas  $P = 1$ ).

Les conditions (F.7) en  $u$  impliquent :

$$\left| \begin{array}{l} \alpha^\omega = \rho(\|u\|) - \frac{\rho'(\|u\|)}{2} \|u\| \\ \beta^\omega = \frac{\rho'(\|u\|)}{2 \|u\|} \end{array} \right.$$

On a montré à la section F.1.1.1 que  $\rho(t) \leq \beta^\omega t^2 + \alpha^\omega$  quel que soit  $(t, v) \in \mathbb{R} \times \mathbb{R}$  avec :

$$\left| \begin{array}{l} \alpha^\omega = \rho(v) - \frac{\rho'(v)}{2} v \\ \beta^\omega = \frac{\rho'(v)}{2 v} \end{array} \right.$$

Par changement de variable  $t = \|w\|$  et  $v = \|u\|$ , on a immédiatement :

$$q^\omega(w) \geq \rho(\|w\|)$$

### F.2.2 Montrons que $q^\omega(w) \leq q^r(w)$ .

Le polynôme  $q^r$  est défini de  $\mathbb{R}^P$  dans  $\mathbb{R}$  par :

$$q^r(w) = \|w - \beta^r\|^2 + \alpha^r$$

Les conditions (F.7) en  $u$  impliquent :

$$\left| \begin{array}{l} \beta^r = u \left( 1 - \frac{\rho'(\|u\|)}{2\|u\|} \right) \\ \alpha^r = \rho(\|u\|) - \left( \frac{\rho'(\|u\|)}{2} \right)^2 \end{array} \right.$$

On pose  $\forall w \in \mathbb{R}^P$  :

$$Z_2(w) = q^r(w) - q^\omega(w)$$

On suppose que  $w$  dépend linéairement d'un paramètre  $x$ . La dérivée de  $Z_2(w)$  par rapport à  $x$  s'écrit :

$$\begin{aligned} \dot{Z}_2(w) &= 2 \langle \dot{w} | w \rangle - 2 \left\langle \dot{w} \left| u \left( 1 - \frac{\rho'(\|u\|)}{2\|u\|} \right) \right. \right\rangle - 2 \frac{\rho'(\|u\|)}{2\|u\|} \langle \dot{w} | w \rangle \\ &= 2 \langle \dot{w} | (w - u) \rangle \left( 1 - \frac{\rho'(\|u\|)}{2\|u\|} \right) \end{aligned} \tag{F.8}$$

$w$  dépend linéairement de  $x$  ce qui implique  $\ddot{w} = 0$ . La dérivée seconde de  $Z_2$  en  $x$  s'écrit :

$$\ddot{Z}_2(w) = 2 \|\dot{w}\|^2 \left( 1 - \frac{\rho'(\|u\|)}{2\|u\|} \right) \tag{F.9}$$

La fonction de pondération  $\beta^w$ , a ses valeurs entre 0 et 1, d'où  $\ddot{Z}_2(w) \geq 0$ . La dérivée  $\dot{Z}_2$  est donc une fonction croissante en  $x$ . On remarque que  $\dot{Z}_2(u) = 0$  ce qui implique que le minimum de la fonction  $Z_2$  est atteint en  $w = u$ , ou

$$\forall w \in \mathbb{R}^P, Z_2(w) \geq Z_2(u) = 0$$

Donc, on a montré que :

$$q^\omega(w) \leq q^r(w)$$

### F.2.3 Conclusion

Les fonctions  $q^\omega(w)$  et  $q^r(w)$  ont été défini pour vérifier les conditions F.7 en  $u \in \mathbb{R}^P$ . On note  $Q^\omega$  et  $Q^r$  les fonctions à deux variables  $(w, u) \in \mathbb{R}^P \times \mathbb{R}^P$ , définies par :

$$Q^\omega(w, u) = \frac{\rho'(\|u\|)}{2\|u\|} \|w\|^2 + \rho(\|u\|) - \frac{\rho'(\|u\|)}{2} \|u\|$$

$$Q^r(w, u) = \left\| w - u \left( 1 - \frac{\rho'(\|u\|)}{2\|u\|} \right) \right\|^2 + \rho(\|u\|) - \left( \frac{\rho'(\|u\|)}{2} \right)^2$$

Alors, on a montré que la fonction  $w \rightarrow \rho(\|w\|)$  peut s'exprimer par :

$$\rho(\|w\|) = \inf_{u \in \mathbb{R}^P} \{Q^\omega(w, u)\}$$

$$\rho(\|w\|) = \inf_{u \in \mathbb{R}^P} \{Q^r(w, u)\}$$

et l'inf est atteint pour  $u = w$ .

**Extension de la théorie semi-quadratique aux images multi-bandes** Pour une fonction  $\rho$  vérifiant les propriétés 3 (page 112), on peut écrire :

$$\begin{cases} \rho(\|w\|) = \inf_{b^A \in \mathbb{R}} \{\rho^A(w, b^A) = b^A \|w\|^2 + \Psi(b^A)\} \\ \rho(\|w\|) = \inf_{b^L \in \mathbb{R}^P} \{\rho^L(w, b^L) = \|w - b^L\|^2 + \xi(b^L)\} \end{cases} \quad (\text{F.10})$$

avec la variable  $b^A \in \mathbb{R}$  définie par :

$$\begin{cases} b^A = \frac{\rho'(\|u\|)}{2\|u\|} \\ \Psi(b^A) = \rho(\|u\|) - \frac{\rho'(\|u\|)}{2}u \end{cases}$$

et  $b^L \in \mathbb{R}^P$  définie par :

$$\begin{cases} b^L = \|u\| \left(1 - \frac{\rho'(\|u\|)}{2\|u\|}\right) \\ \xi(b^L) = \rho(\|u\|) - \left(\frac{\rho'(\|u\|)}{2}\right)^2 \end{cases}$$

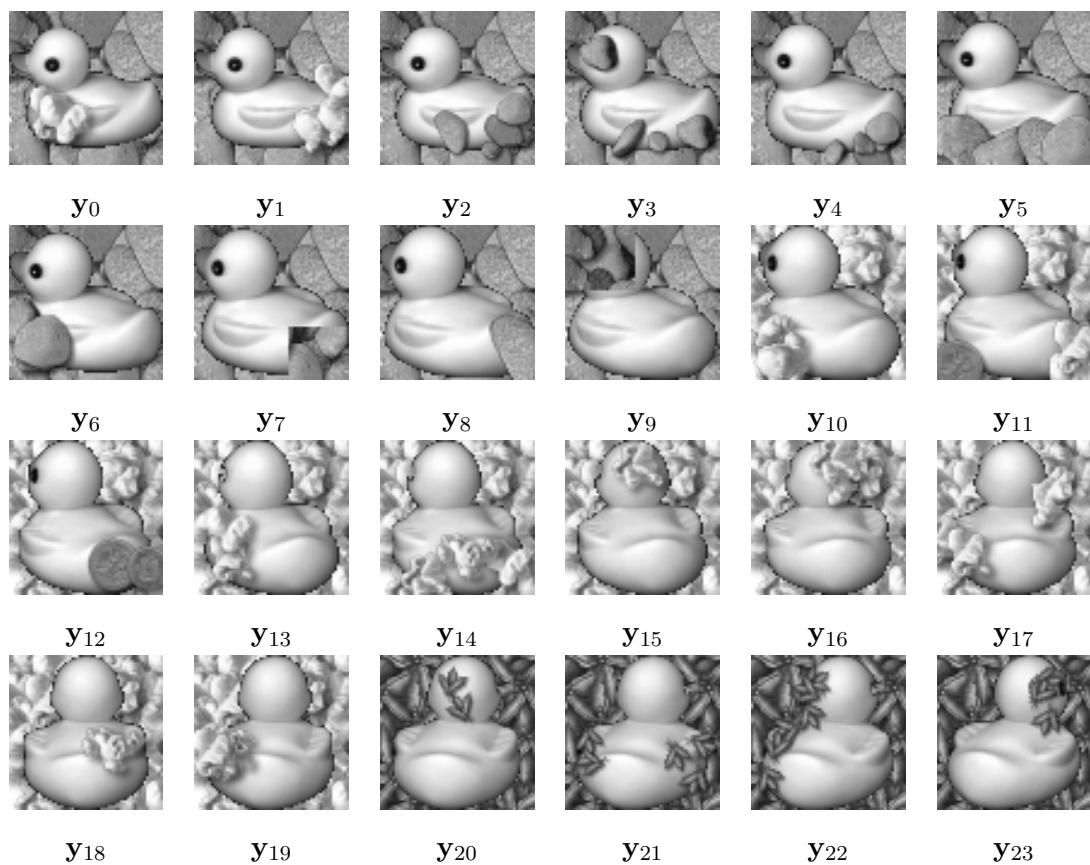
L'inf est atteint lorsque  $u = w$ .

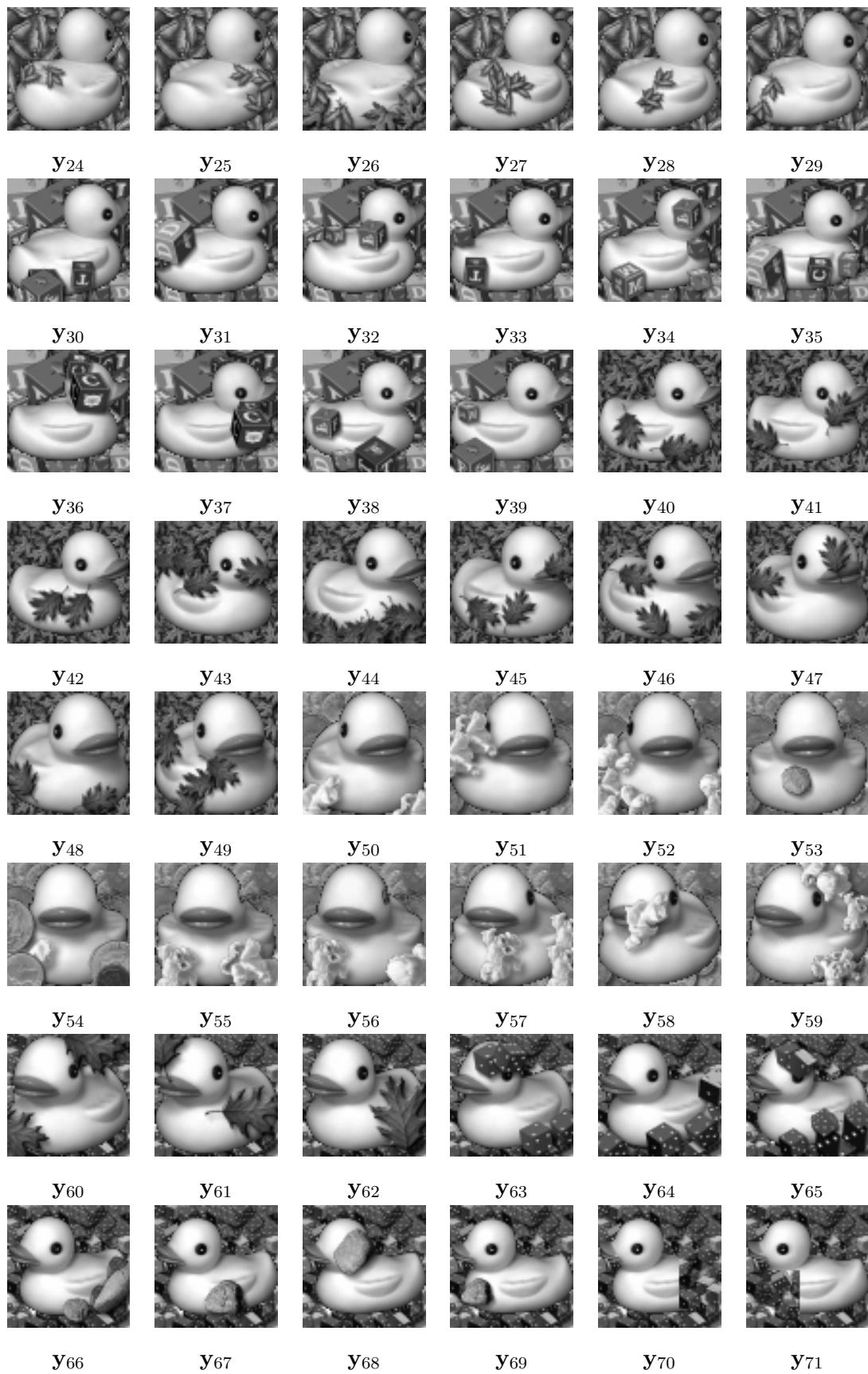
## Annexe G

# Reconnaissance : expérience avec la base COIL

Le tableau G.1 présente 72 observations des objets de la base COIL, partiellement occultées et apparaissant sur des fonds texturés.

TAB. G.1: Reconnaissance : Observations partiellement occultées.



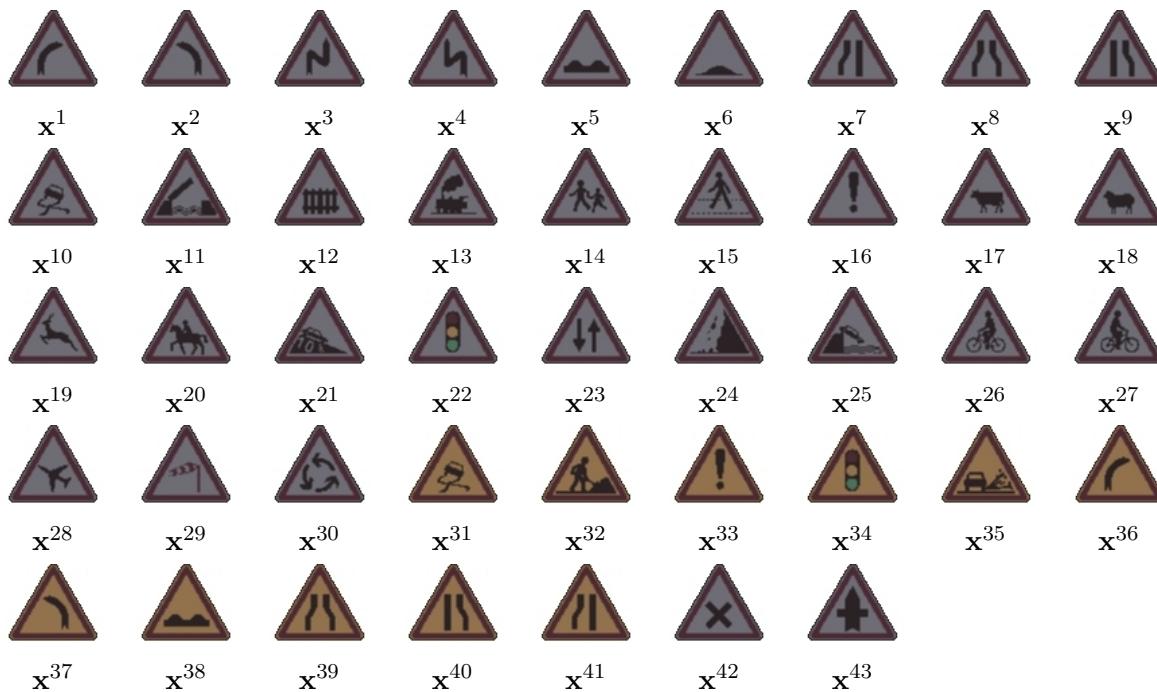


## Annexe H

# Reconnaissance : expérience avec la base A43

## H.1 Images d'apprentissage

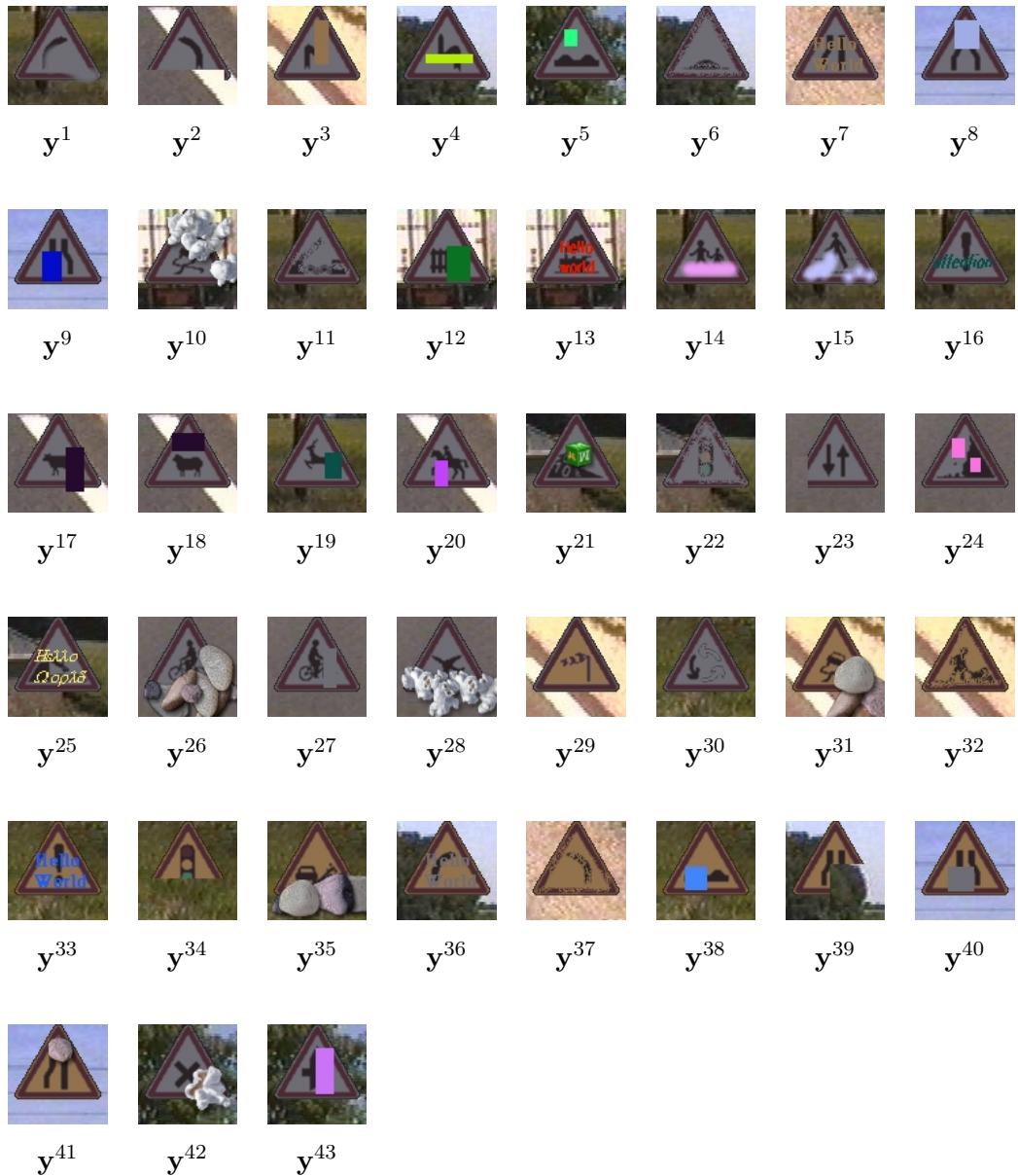
La base d'images d'apprentissage, appelée A43, contient 43 panneaux triangulaires appris avec leur rotation dans le plan de l'image (une image tous les 10 degrés), soit 1548 images d'apprentissage. Les 43 panneaux sont représentés au tableau H.1. Les images sont de dimension  $76 \times 76$  pixels, d'où les vecteurs sont de dimension  $76 \times 76 \times 3 = 17328$ .



TAB. H.1: Panneaux triangulaires de la base d'apprentissage A43.

## H.2 Test de reconnaissance sur des images présentant des occultations

L'image observée,  $\mathbf{y}^k$ , correspond à une occurrence  $\mathbf{x}^k$  dégradée. Les 43 images  $\{\mathbf{y}^k\}_{k \in \{1, \dots, 43\}}$ , présentant des panneaux partiellement occultés, sont présentées au tableau H.2.



TAB. H.2:*Observations.*

Le tableau H.3 présente les résultats de reconnaissance obtenus avec les différents estimateurs ro-

bustes selon les différents apprentissages sur les images présentées au tableau H.2.

$(\hat{\mu}^c, \hat{\Sigma}^c)$		$(\hat{\mu}^n, \hat{\Sigma}^c)$		$(\hat{\mu}^c, \hat{\Sigma}^c)$		$(\hat{\mu}^n, \hat{\Sigma}^c)$	
$\hat{\mathbf{c}}_{RMV}$	$\hat{\mathbf{c}}_{RMAP}$	$\hat{\mathbf{c}}_{RMV}$	$\hat{\mathbf{c}}_{RMAP}$	$\hat{\mathbf{c}}_{RMV}$	$\hat{\mathbf{c}}_{RMAP}$	$\hat{\mathbf{c}}_{RMV}$	$\hat{\mathbf{c}}_{RMAP}$
$\mathbf{y}^1$	$\mathbf{x}^{29}$	$\mathbf{x}^{29}$	$\mathbf{x}^1$	$\mathbf{x}^1$			
$\mathbf{y}^3$	$\mathbf{x}^1$	$\mathbf{x}^1$	$\mathbf{x}^1$	$\mathbf{x}^1$			
$\mathbf{y}^5$	$\mathbf{x}^5$	$\mathbf{x}^5$	$\mathbf{x}^5$	$\mathbf{x}^5$			
$\mathbf{y}^7$	$\mathbf{x}^7$	$\mathbf{x}^7$	$\mathbf{x}^7$	$\mathbf{x}^7$			
$\mathbf{y}^9$	$\mathbf{x}^9$	$\mathbf{x}^9$	$\mathbf{x}^9$	$\mathbf{x}^9$			
$\mathbf{y}^{11}$	$\mathbf{x}^6$	$\mathbf{x}^6$	$\mathbf{x}^6$	$\mathbf{x}^6$			
$\mathbf{y}^{13}$	$\mathbf{x}^{13}$	$\mathbf{x}^{13}$	$\mathbf{x}^{13}$	$\mathbf{x}^{13}$			
$\mathbf{y}^{15}$	$\mathbf{x}^2$	$\mathbf{x}^2$	$\mathbf{x}^2$	$\mathbf{x}^2$			
$\mathbf{y}^{17}$	$\mathbf{x}^{17}$	$\mathbf{x}^{17}$	$\mathbf{x}^{17}$	$\mathbf{x}^{17}$			
$\mathbf{y}^{19}$	$\mathbf{x}^{19}$	$\mathbf{x}^{19}$	$\mathbf{x}^{19}$	$\mathbf{x}^{19}$			
$\mathbf{y}^{21}$	$\mathbf{x}^{12}$	$\mathbf{x}^{12}$	$\mathbf{x}^{21}$	$\mathbf{x}^{21}$			
$\mathbf{y}^{23}$	$\mathbf{x}^{23}$	$\mathbf{x}^{23}$	$\mathbf{x}^{23}$	$\mathbf{x}^{23}$			
$\mathbf{y}^{25}$	$\mathbf{x}^{25}$	$\mathbf{x}^{25}$	$\mathbf{x}^{25}$	$\mathbf{x}^{25}$			
$\mathbf{y}^{27}$	$\mathbf{x}^{27}$	$\mathbf{x}^{27}$	$\mathbf{x}^{27}$	$\mathbf{x}^{27}$			
$\mathbf{y}^{29}$	$\mathbf{x}^{31}$	$\mathbf{x}^{31}$	$\mathbf{x}^{31}$	$\mathbf{x}^{31}$			
$\mathbf{y}^{31}$	$\mathbf{x}^{31}$	$\mathbf{x}^{31}$	$\mathbf{x}^{31}$	$\mathbf{x}^{31}$			
$\mathbf{y}^{33}$	$\mathbf{x}^{33}$	$\mathbf{x}^{33}$	$\mathbf{x}^{33}$	$\mathbf{x}^{33}$			
$\mathbf{y}^{35}$	$\mathbf{x}^{38}$	$\mathbf{x}^{38}$	$\mathbf{x}^{38}$	$\mathbf{x}^{38}$			
$\mathbf{y}^{37}$	$\mathbf{x}^{37}$	$\mathbf{x}^{37}$	$\mathbf{x}^{37}$	$\mathbf{x}^{37}$			
$\mathbf{y}^{39}$	$\mathbf{x}^{32}$	$\mathbf{x}^{32}$	$\mathbf{x}^{40}$	$\mathbf{x}^{32}$			
$\mathbf{y}^{41}$	$\mathbf{x}^{41}$	$\mathbf{x}^{41}$	$\mathbf{x}^{41}$	$\mathbf{x}^{41}$			
$\mathbf{y}^{43}$	$\mathbf{x}^1$	$\mathbf{x}^{27}$	$\mathbf{x}^1$	$\mathbf{x}^1$			

TAB. H.3: Modèles reconnus par le système de reconnaissance robuste, sur un espace propre  $F$  de dimension  $J = 60$ , pour les observations du tableau H.2.



## Annexe I

# Courbes COR

La courbe COR ou caractéristique opérationnelle du récepteur, présente le pourcentage de bonnes détection en ordonnée, en fonction du pourcentage de fausses alarmes en abscisse. Elle permet d'évaluer les performances de nos détecteurs. Nos tests de détection comportent  $K$  images présentant  $N_o$  occurrences d'objets à détecter. Le passage d'un détecteur sur ces images permet de calculer  $K$  cartes de ressemblance (carte de similarité ou de vraisemblance ...). On note :

- le nombre total de fausses alarmes possibles  $N_F$ , qui correspond au nombre de positions dans les images où n'apparaît pas d'objets,
- le nombre d'objets à détecter  $N_o$ .

Pour un seuil  $s$  donné, on calcule :

- le nombre de fausses alarmes  $n_F(s)$ , i.e. le nombre de positions où la mesure de ressemblance est supérieure au seuil  $s$  bien qu'il n'y ait pas d'objet.
- le nombre de bonnes détection  $n_o(s)$ , i.e. le nombre de positions où il y a un objet et dont la mesure de ressemblance est au dessus du seuil  $s$ .

Alors la courbe COR se trace en fonction de  $s$ , suivant le modèle paramétrique :

$$\begin{cases} x = \frac{n_F(s)}{N_F} \\ y = \frac{n_o(s)}{N_o} \end{cases}$$

On connaît les positions exactes de nos objets dans les images, notée  $(i, j, k)$  avec  $i$  le numéro de colonnes,  $j$  le numéro de ligne, et  $k$  le numéro de l'image. On tolère une précision de un pixel dans la détection c'est-à-dire, si pour un seuil donné  $s$ , le maximum de ressemblance dans un 8-voisinage de  $(i, j)$ , noté  $v(i, j)$ , est supérieure à  $s$ , alors on considère que l'objet présent en  $(i, j)$  est détecté. Toutes les positions n'appartenant pas à la liste des positions objets et leurs voisinages, sont de fausses alarmes

potentielles. Ceci est illustrée à la figure I.1 : dans cet exemple, trois mesures de ressemblance sont au-dessus du seuil  $s$  dans la carte calculée sur l'image  $I_k$ . L'une correspond à une bonne détection car elle est située sur le 8-voisinage de l'objet  $O_1$ , et les deux autres sont des fausses alarmes. L'objet  $O_2$  n'est quant à lui pas détecter pour ce seuil  $s$ .

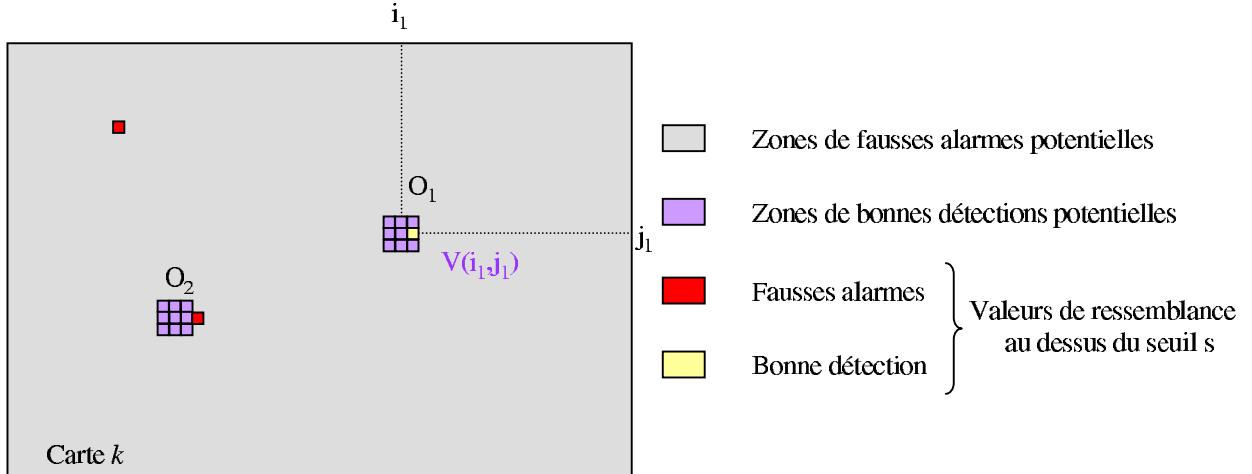


FIG. I.1: Analyse d'une carte de ressemblance de l'image  $I_k$  à un seuil  $s$  donnée.

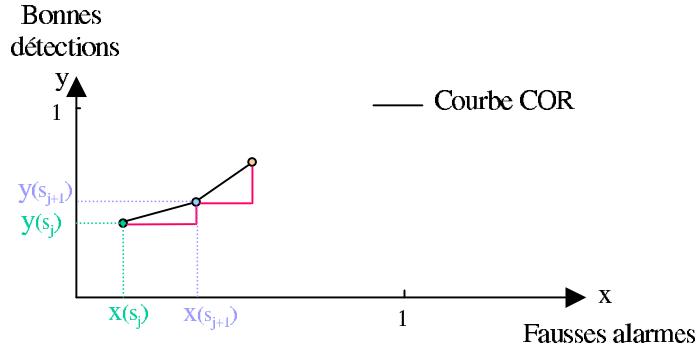
**Calcul des courbes COR.** Les valeurs de ressemblance correspondant aux bonnes détections, sont utilisées pour définir les seuils  $s$  pour tracer la courbe COR. Celles-ci sont classées en ordre décroissant : la première valeur correspond alors à la meilleure mesure de ressemblance obtenue pour la détection d'un objet, et définit le seuil le plus grand  $s_1$ . On définit ainsi une série de seuils  $s_1 \geq s_2 \geq \dots \geq s_{N_o}$ .

Pour simplifier, on considère le cas des inégalités strictes  $s_1 > s_2 > \dots > s_{N_o}$ . Dans ce cas, le pourcentage de bonnes détections est directement déterminé :  $n_o(s_1) = 1, n_o(s_2) = 2, \dots, n_o(s_{N_o}) = N_o$ . Il ne reste qu'à compter pour chaque seuil le nombre de fausses alarmes  $n_F(s)$ . On dispose alors de  $N_o$  points pour tracer la courbe COR (cf. tableau I.1).

$s$	$n_F(s)$	$x = \frac{n_F(s)}{N_F}$	$n_o(s)$	$y = \frac{n_o(s)}{N_o}$
$s_1$	$n_F(s_1)$	$x = \frac{n_F(s_1)}{N_F}$	$n_o(s_1) = 1$	$y = \frac{1}{N_o}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$s_{N_o}$	$n_F(s_{N_o})$	$x = \frac{n_F(s_{N_o})}{N_F}$	$n_o(s_{N_o}) = N_o$	$y = \frac{N_o}{N_o} = 1$

TAB. I.1: Calcul des  $N_o$  points  $(x(s), y(s))$  de la courbe COR.

Il est inutile de calculer des points supplémentaires de la courbe COR, avec une autre valeur de seuil  $s$  car si  $s_j < s < s_{j+1}$ , alors l'ordonnée  $y(s)$  est équivalente à  $y(s_j)$  ( le nombre d'objets détecté reste

FIG. I.2: *Tracé de la courbe COR.*

constant). Cette remarque indique que la courbe COR a une allure de courbe en escalier (cf. courbe rouge à la figure I.2) dû au nombre fini d'objet à détecter dans l'ensemble des images  $I_k$  considérés. Pour obtenir une courbe lissée, on a choisi de joindre les points  $(x(s_j), y(s_j))$  et  $(x(s_{j+1}), y(s_{j+1}))$  par un segment de droite (cf. courbe noire à la figure I.2).

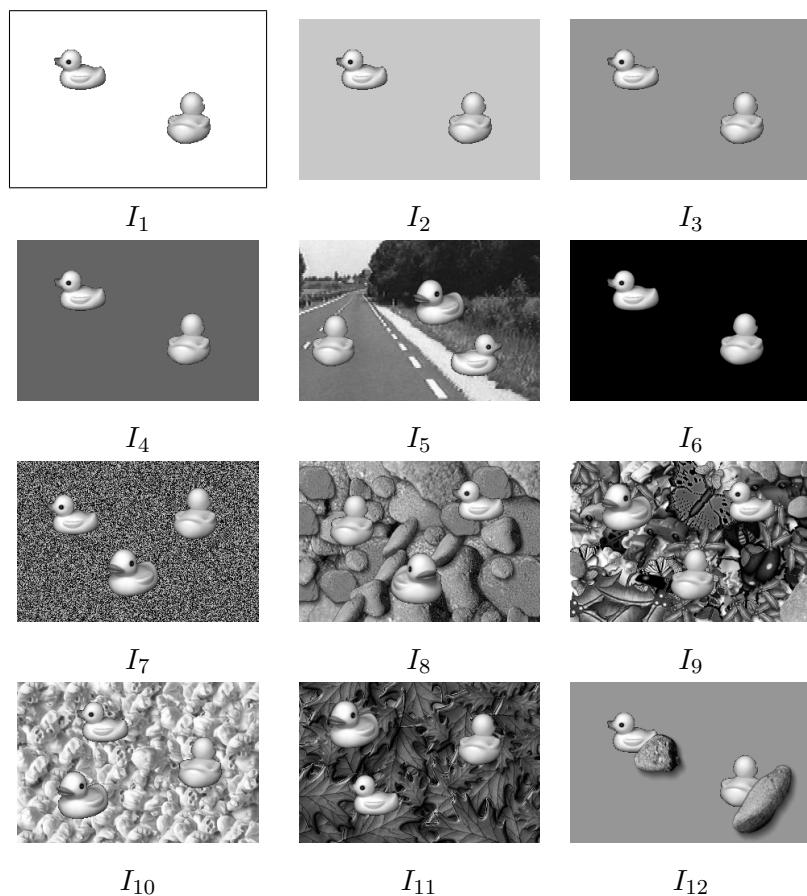
**Remarque.** Le nombre total de fausses alarmes possibles  $N_F$  correspond au nombre total d'observations  $\mathbf{y}$  que l'on peut extraire des images  $\{I_k\}_{k \in \{1, \dots, K\}}$ , et qui ne sont pas des occurrences des objets de la base  $\mathcal{B}$ , soit  $\mathbf{y} \in \overline{\mathcal{B}}$ . Pour un seuil  $s$  donné, l'abscisse  $x = \frac{n_F(s)}{N_F}$  correspond donc à la proportion d'observations  $\mathbf{y} \in \overline{\mathcal{B}}$  qui sont classées par erreur comme étant des observations de  $\mathcal{B}$ .

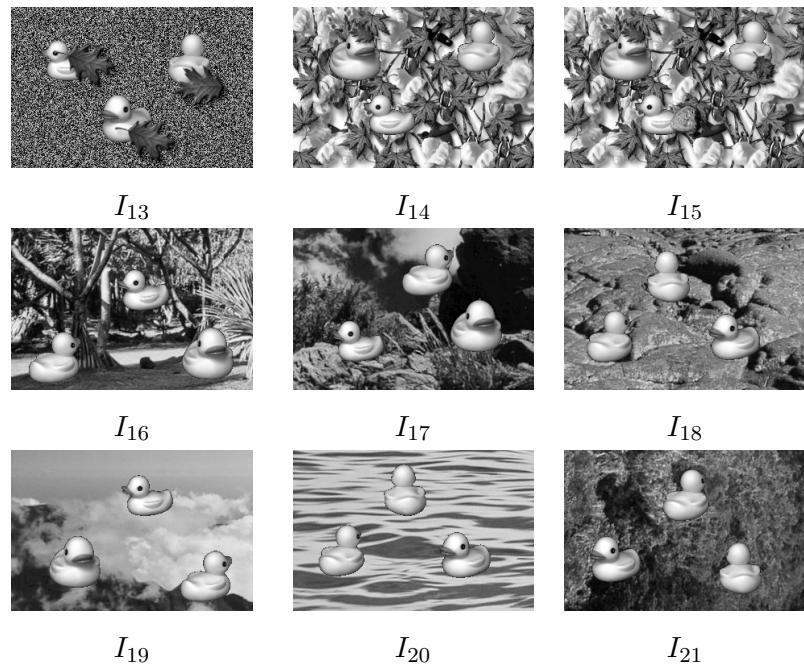


## Annexe J

# Détection : expérience sur des images en niveaux de gris

Les images du test de détection avec la base d'apprentissage COIL sont présentées au tableau J.1. 57 occurrences de l'objet y apparaissent dans 21 images.





TAB. J.1: *Images du test de détection avec la base d'apprentissage COIL.*

## Annexe K

# Détection : expérience sur des images en couleur

### K.1 Test avec la base A43

Les images utilisées pour tester les méthodes de détection avec la base d'apprentissage A43, sont présentées au tableau K.1. Celles-ci proposent 26 occurrences de panneaux dans 18 images.



*I<sub>1</sub>*



*I<sub>2</sub>*



*I<sub>3</sub>*



*I<sub>4</sub>*



*I<sub>5</sub>*



*I<sub>6</sub>*



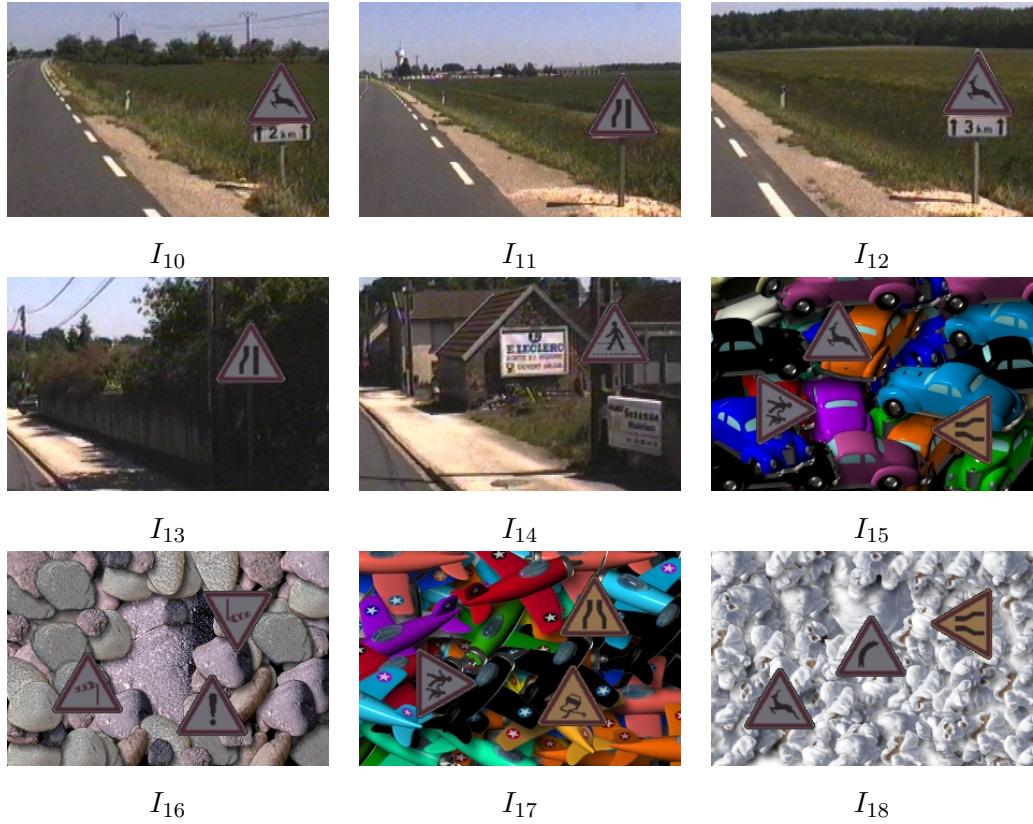
*I<sub>7</sub>*



*I<sub>8</sub>*



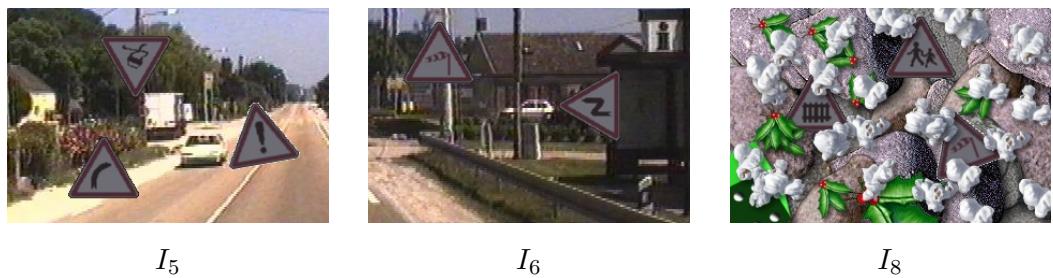
*I<sub>9</sub>*

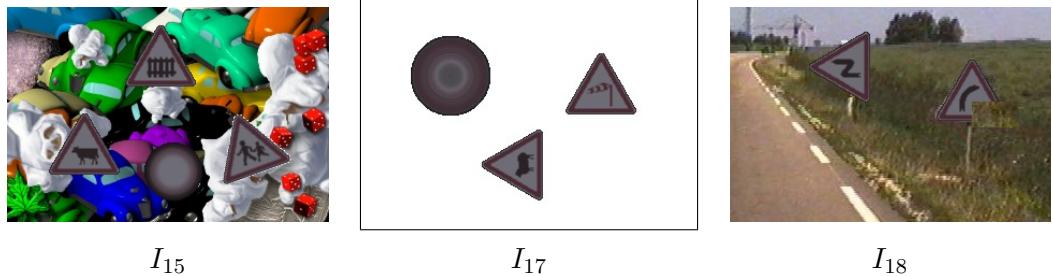


TAB. K.1: Détection : Images de test avec la base d'apprentissage A43.

## K.2 Test avec la base AVG

Les images de ce test sont les images ( $I_1, I_2, I_3, I_4, I_7, I_9, I_{10}, I_{11}, I_{12}, I_{13}, I_{14}, I_{16}$ ) présentées au tableau K.1, complétées par les images du tableau K.2. Ce test contient 29 occurrences de panneaux dans 18 images. On remarque que l'image  $I_{17}$  présente unurre qui correspond à l'image moyenne  $\hat{\mu}^n$  de la base AVG. Des occultations apparaissent sur les images en  $I_8$  et  $I_{18}$ .





TAB. K.2: Détection : Images de test avec la base d'apprentissage AVG.



## **Annexe L**

# **Publications de l'auteur**

### **Conférences internationales**

- R. Dahyot, P. Charbonnier et F. Heitz, *Unsupervised Statistical Change Detection in Camera-in-Motion Video*, in proceedings of IEEE International Conference on Image Processing (ICIP), October 2001, GREECE.
- R. Dahyot, P. Charbonnier et F. Heitz, *Robust Visual Recognition of Colour Images*, in proceedings of IEEE conference on Computer Vision and Pattern Recognition (CVPR), volume 1, pages 685-690, June 2000, Hilton Head Island USA.

### **Workshops internationales**

- R. Dahyot, P. Charbonnier et F. Heitz, *Non-Supervised Robust Visual Recognition of Colour Images using Half-Quadratic Theory*, in proceedings of European Workshop on Content-Based Multimedia Indexing, October 1999, Toulouse FRANCE.

### **Conférences nationales**

- R. Dahyot, P. Charbonnier et F. Heitz, *Détection d'évènements dans les séquences d'images avec caméra en mouvement*, accepté au colloque GRETSI'01, septembre 2001, Toulouse, FRANCE.
- R. Dahyot, P. Charbonnier et F. Heitz, *Détection robuste d'objets : une approche par modèle d'apparence*, accepté au colloque GRETSI'01, septembre 2001, Toulouse, FRANCE.
- R. Dahyot, P. Charbonnier et F. Heitz, *Reconnaissance robuste non supervisée d'images en couleur utilisant la théorie semi-quadratique*, 17<sup>ème</sup> colloque GRETSI, volume 2, pages 295-298, septembre 1999, Vannes FRANCE.



## Résumé

**Titre :** Analyse d'images séquentielles de scènes routières par modèles d'apparence pour la gestion du réseau routier

Nous abordons dans cette thèse, le problème de l'indexation par le contenu de bases d'images séquentielles de scènes routières. On s'intéresse particulièrement à la détection et la reconnaissance des objets fixes sur l'accotement (arbres, panneaux,...). Deux méthodes de détection sont proposées. La première repose sur une analyse de l'apparence locale et sur une hypothèse de comportements statistiques temporels différents entre la classe d'objets d'intérêt et les autres éléments de la séquence vidéo. L'analyse des changements statistiques temporels permet d'extraire les densités de probabilités des mesures locales associées aux objets d'intérêt dont la localisation dans la séquence est ensuite réalisée par rétro-projection. Cette technique originale est non supervisée et détecte les événements tels que l'apparition, la disparition ou le changement d'apparence des objets dans la séquence vidéo. La seconde méthode de détection proposée repose sur un apprentissage d'un ensemble de caractéristiques globales de l'apparence représentatives de la classe d'objets d'intérêt. Les techniques classiques de détection et de reconnaissance basées sur une représentation par espace propre, ne sont pas robustes aux perturbations locales telles que les occultations partielles. Pour palier ce problème, nous proposons de nouvelles procédures reposant sur des estimateurs statistiques robustes. Elles sont comparées aux méthodes classiques de détection et de reconnaissance par espace propre et montrent leur nette supériorité. Toutes les techniques proposées ici sont développées dans le cadre statistique bayésien permettant de prendre en compte un *a priori* sur la distribution dans l'espace propre.

**Mots clés :** Détection et reconnaissance d'objets, apprentissage statistique, estimation robuste, M-estimateurs, analyse en composantes principales, histogramme multidimensionnel.

## Abstract

**Title :** Appearance based road scene video analysis for the management of the road network.

This thesis deals with the detection and the recognition of static objects (poles, traffic signs,...) on the road side in road scene video databases. The first part focuses on statistical event detection using local appearance based features. Statistics of the objects of interest and of the cluttered background have a different temporal behaviour which is used to extract relevant probability density functions of local measures on objects. Objects are localized thanks to a backprojection algorithm. This method is unsupervised and allows to detect events like entering, exiting or changing objects in the scene.

The second part on this thesis focuses on robustly detecting and recognizing possibly occluded objects using global appearance-based features. Appearances representative of an object class, are collected. In order to reduce their dimension, we use an eigenspace representation. Standard detection and recognition methods rely on Gaussian noise models. Therefore they cannot deal with structured noise due to cluttered background or partial occlusion. We propose new techniques based on robust estimators able to deal with outliers. They are developed in a bayesian framework allowing to take into account an *a priori* knowledge on the distribution in the eigenspace. The proposed bayesian robust method significantly outperforms the standard robust or non-robust approaches.

**Key Words :** Detection and recognition of objects, statistical learning, robust estimation, M-estimators, Principal Component Analysis, multidimensional histogram.