

DESCRIPTEURS

Comme expliqué précédemment, la phase de description de la zone locale (le voisinage) autour des PIs est une étape cruciale pour un système de reconnaissance. La représentation locale donnée par ces descripteurs constitue l'entrée de la phase de mise en correspondance des modèles. Dans ce chapitre, nous présentons quelques descripteurs 3D de la littérature et nous introduisons le principe de nos deux nouveaux descripteurs, inspirés de ces derniers. L'évaluation comparative de ces descripteurs est effectuée sur les aspects suivants : invariance aux transformations géométriques de rotation et l'échelle, pouvoir descriptif, robustesse au bruit et à l'échantillonnage et temps de calcul.

4.1. Présentation des descripteurs testés

Les descripteurs 3D peuvent être divisés en deux catégories selon Tombari et al. (Tombari, et al., 2010): Signature et Histogramme. Une Signature décrit chaque point du voisinage d'un point donné par une ou plusieurs mesures géométriques exprimées en coordonnées locales relatives à un système local de référence invariant (Reference Frame, RF). L'utilisation de ces informations spatiales bien localisées augmente le pouvoir descriptif des descripteurs à base de signature.

Les méthodes basées sur les histogrammes décrivent la région du support par un cumul de mesures géométriques ou topologiques locales (ex. nombre de points, superficie d'un maillage) dans des histogrammes par rapport à un domaine spécifique (ex. coordonnées du point, courbures) et qui nécessite la définition soit d'un axe de référence (Reference Axis, RA), soit d'une RF locale. En général, les descripteurs à base d'histogrammes sont robustes au bruit et à l'échantillonnage grâce au processus de regroupement de l'information de la forme dans des cases d'intervalles. Les approches que nous allons présenter appartiennent à l'une de ces types de catégories: histogramme ou signature ou les deux ensembles.

4.1.1 Descripteur Spin Image

Comme mentionné précédemment dans le chapitre 2, le Spin Image (Johnson, et al., 1999) a été proposé pour décrire des points d'intérêt par Li et Guskov (Li, et al., 2007). Ce descripteur traduit les propriétés locales de la surface dans un système de coordonnées orienté, fixe et lié à l'objet. Ce système est indépendant du changement de vue, contrairement à un système de coordonnées lié à la caméra qui lui dépend de l'angle de vue. Le spin est définie en un point orienté désigné par sa position 3D (p) et sa direction associée (n la normale de la surface locale). Une base 2D de coordonnées locales est formée en utilisant le plan tangent P en p , orienté perpendiculairement à la normale n , et la ligne L passant par p parallèlement à n . Un système de coordonnées cylindriques (α, β) du point p est alors déduit, où α est la coordonnée radiale définissant la distance (non négative) perpendiculaire à L et β est la coordonnée d'élévation qui définit la distance signée perpendiculaire (positive ou négative) à P . La carte des mesures du Spin peut être exprimée comme une fonction de projection des points 3D

de l'objet en coordonnées 2D associés dans la base du point orienté. La projection est comme suit (Johnson, et al., 1998) :

$$S_O: R^3 \rightarrow R^2$$

$$S_O(x) \rightarrow (\alpha, \beta) = (\sqrt{\|x - p\|^2 - (n \cdot (x - p))^2}, n \cdot (x - p))$$

Équation 4.1

La Figure 4-1 explique la formation de ce système de coordonnées. Une fenêtre locale de voisinage appelée support est définie pour chaque PI à décrire, et tout point de ce support est exprimé dans la base du PI selon la fonction (Équation 4.1). Les coordonnées cylindriques (α, β) de ces points sont cumulées par la suite dans des cases discrètes d'un tableau 2D selon les équations (Équation 4.3). Pour chaque couple de coordonnées est incrémenté la case correspondante, ainsi que les cases qui l'entourent dans la table pour réduire l'effet du bruit (Équation 4.4). Pour ce faire, la contribution du point 2D est dispersée sur les quatre cases voisines du tableau 2D en utilisant une interpolation bilinéaire. La création de la représentation sous forme de tableau 2D de la Spin Image est illustrée dans la Figure 4-2. Un exemple de spins images générés pour trois points orientés est donné dans la Figure 4-3.

Les paramètres : taille du tableau et taille maximale de l'objet agissent sur le pouvoir descriptif de ce descripteur. La taille maximale de l'intervalle du tableau correspond aux maximums des valeurs de coordonnées $(\alpha_{max}, \beta_{max})$ dans toutes les bases générées des points orientés de l'objet. La taille du bin noté b doit être ajustée de façon à ce que l'image du spin ne soit pas très grande tout en gardant élevé le pouvoir descriptif. La taille de l'image finale (i_{max}, j_{max}) dépend de la valeur de la taille du bin et est calculée comme suit (Johnson & Hebert, 1998):

$$i_{max} = \frac{2\beta_{max}}{b} + 1 \qquad j_{max} = \frac{\alpha_{max}}{b} + 1$$

Équation 4.2

D'après Johnson et Hebert (Johnson, et al., 1999), un meilleur résultat de l'appariement des descripteurs Spin, indépendant de la taille et de la résolution du modèle (densité des points), est obtenu pour une taille du bin multiple de la résolution du maillage.

Le cumul des valeurs 2D (α, β) des points dans des valeurs discrètes correspondant aux bins de l'histogramme se fait selon les équations suivantes :

$$i = \left\lfloor \frac{\beta_{max} - \beta}{b} \right\rfloor \qquad j = \left\lfloor \frac{\alpha}{b} \right\rfloor$$

Équation 4.3

Les poids bilinéaires des attributs servant à l'incrémentation des bins sont :

$$a = \alpha - ib \qquad b = \beta - jb$$

Équation 4.4

Avec cette définition, le pouvoir descriptif des Spins Images permet d'apparier convenablement deux surfaces différentes d'un même objet. Pour une variation de l'échantillonnage des deux surfaces ou pour une présence de bruit, les deux spins images sont légèrement différents. Cependant avec un échantillonnage uniforme de la surface (les arêtes des facettes ont presque la même longueur) les spins images des points des deux surfaces sont linéairement corrélés.

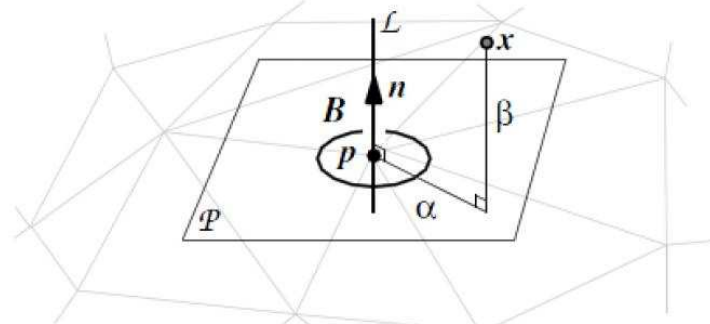


Figure 4-1 Formation du système de coordonnées cylindriques pour un point orienté (Johnson, et al., 1998)

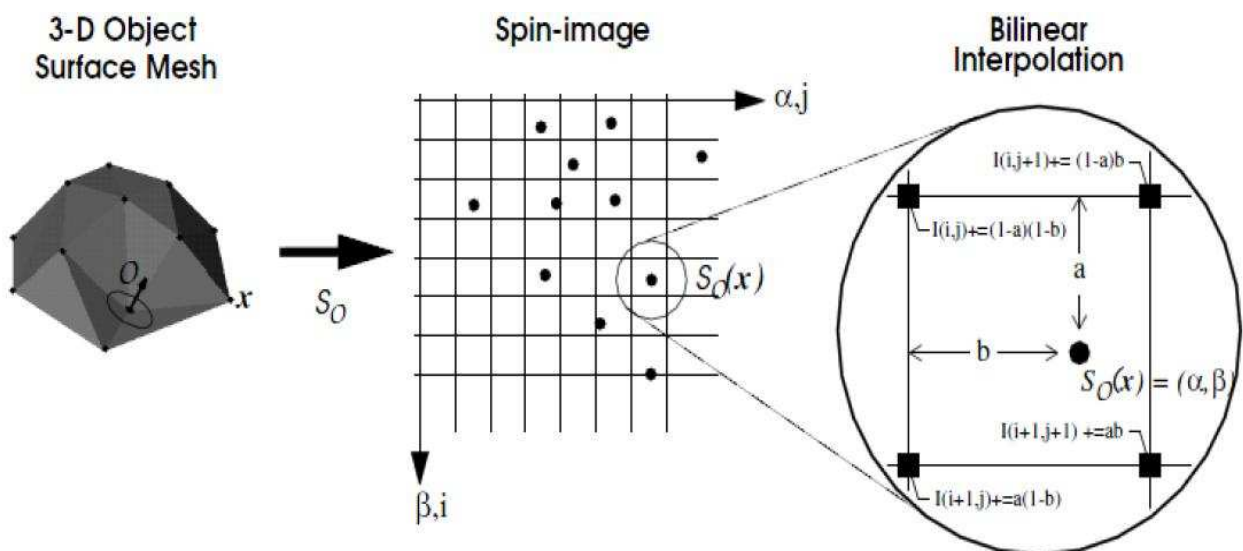


Figure 4-2 Processus de création de la représentation 2D du spin image (Johnson, et al., 1998)

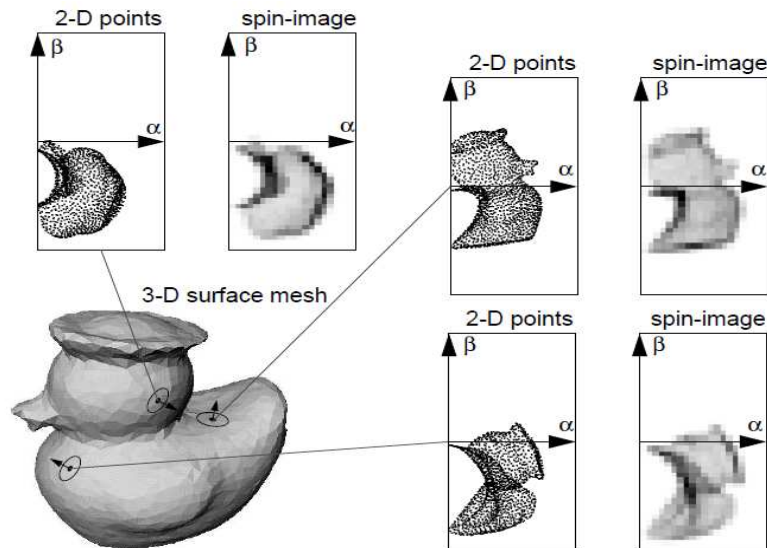


Figure 4-3 Représentation des Spin-images pour deux valeurs de support pour trois points orientés de la surface du modèle de Duck. Dans l'accumulateur, les zones foncées correspondent à plus de points projetés dans les cases (Johnson, et al., 1998)

4.1.2 Descripteur Thrift

Le deuxième descripteur considéré dans notre étude comparative est le Thrift, proposé par A. Flint et al. (Flint, et al., 2007) comme extension du SIFT 2D. Ce dernier se compose d'un histogramme cumulant, pour un voisinage du PI, le cosinus de l'angle entre deux normales calculées sur deux supports. En effet, pour chaque point d'intérêt, un support des points voisins se situant à une distance inférieure à une distance seuil σ du point référence, et est défini par :

$$Support(z) = \{y \in \mathcal{X} : \|y - x\| \leq \sigma\}$$

Équation 4.5

Pour chaque point y du support deux fenêtres W_1 et W_2 de points correspondant à deux valeurs de distances w_{small} et w_{large} , et sont définies par :

$$\begin{aligned} W_1 &= \{p \in \mathcal{X} : \|p - y\| \leq \omega_{small}\} \\ W_2 &= \{p \in \mathcal{X} : \|p - y\| \leq \omega_{large}\} \end{aligned}$$

Équation 4.6

En pratique, les normales de surface sont approximées avec des plans de moindre carrées. Soit, deux plans P_1 et P_2 correspondent respectivement aux fenêtres W_1 et W_2 , et n_{small} et n_{large} leurs normales respectives (Figure 4-4), qui peuvent être interprétées comme les courbures principales de la carte de densité (présentée dans le détecteur Thrift, section 2.3.1.2). Les distances w_{small} et w_{large} sont constantes pour tout les points d'intérêt et proportionnelles au rayon de référence du point. Rappelons qu'une échelle σ est donnée en sortie du détecteur, les valeurs utilisées, dans les expérimentations de (Flint, et al., 2007), sont: $w_{small} = 0.3 * \sigma$ et $w_{large} = 0.8 * \sigma$.

Le descripteur au point z est donné par l'histogramme des valeurs des cosinus de l'angle entre les n_{small} et n_{large} des points y appartenant au support :

$$\cos(\theta) = \frac{\mathbf{n}_{small} \cdot \mathbf{n}_{large}}{\|\mathbf{n}_{small}\| \|\mathbf{n}_{large}\|}$$

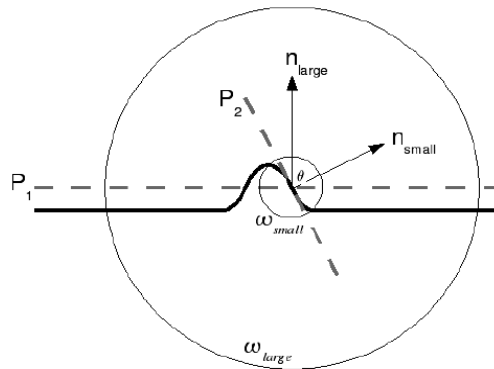


Figure 4-4 Deux plans par moindres carrés P_1 et P_2 et leurs normales (Flint, et al., 2007)

L'intervalle des valeurs du cosinus est $[0, 1]$ pour une variation d'angle entre 0° et 90° et le cumul de ces valeurs dans les cases de l'histogramme est normalisé. La dimension du descripteur final dépend du nombre de cases donné en paramètre.

Le descripteur Thrift se caractérise par son invariance aux rotations 3D, due à son principe de comparer des normales estimées sur deux échelles. De plus, l'utilisation des plans des moindres carrés pour l'estimation des normales assure l'invariance à la densité des points du maillage.

4.1.3 Local surface patches (LSP)

Le descripteur LSP "Local Surface Patches", introduit par Chen et Bhanu (Chen, et al., 2004), est formé d'un histogramme 2D qui cumule, à la fois, l'information des cosinus de l'angle entre les normales des points du support et la normale référence du PIs, et la valeur de l'indice de forme de ces points. La région des voisins N désigne le support et est définie par l'Équation 4.7 :

$$N = \{pixels \ N, \|N - P\| \leq \epsilon_1\} \text{ and } \text{acos}(n_p \bullet n_n) < A.$$

Équation 4.7

Dans cet équation, le \bullet désigne le produit scalaire entre la normale n_p au point référence p , et la normale n_n d'un voisin appartenant à N . Le cosinus est égal au produit scalaire entre les deux normales et acos est la fonction de cosinus inverse. Les paramètres ϵ_1 et A contribuent à la performance du descripteur en terme de pouvoir discriminant de la représentation locale du patch. Pour chaque point N_i du voisinage N , l'indice de forme S_i et le cosinus de l'angle entre sa normale et celle du point référence sont calculés. Un histogramme 2D est formé en cumulant les points dans des cases de deux axes relatifs à la valeur de l'indice de forme et à l'angle des normales selon l'Équation 4.8. L'axe des indices de forme varie entre 0 et 1 et l'axe du cosinus prend ses valeurs dans $[-1, 1]$. Dans l'Équation 4.8, l'opérateur de crochet correspond à l'opérateur d'arrondissement des valeurs réelles (float) en la valeur de l'entier le plus proche et de plus petite valeur. Dans cette équation, h_x est l'indice sur l'axe

horizontal et v_y est l'indice sur l'axe vertical. b_h et b_y correspondent aux bins des intervalles sur les deux axes. Une interpolation bilinéaire est utilisée pour réduire l'effet du bruit dans le calcul de l'histogramme 2D. L'exemple de la Figure 4-5 montre un histogramme LSP sous forme d'image en niveau de gris. Les pixels clairs correspondent aux bins dont le nombre de points cumulés est grand.

$$h_x = \left\lfloor \frac{S_i}{b_h} \right\rfloor, \quad v_y = \left\lfloor \frac{\cos \theta + 1}{b_y} \right\rfloor$$

Équation 4.8

En plus de l'histogramme 2D, la représentation LSP inclut le type T_p obtenu par les mesures des courbures gaussiennes et moyennes de la surface locale (une valeur de 1 à 9) et le centroïde du support. L'ensemble est utilisé pour la phase de mise en correspondance des PIs.

Dans ce chapitre, lors de notre implémentation et évaluation de ce descripteur, nous limitons la représentation LSP à l'histogramme 2D uniquement. Notons que dans la formulation de ce descripteur, la notion de RA unique et non ambiguë est utilisée par le biais du vecteur normal du point référence.

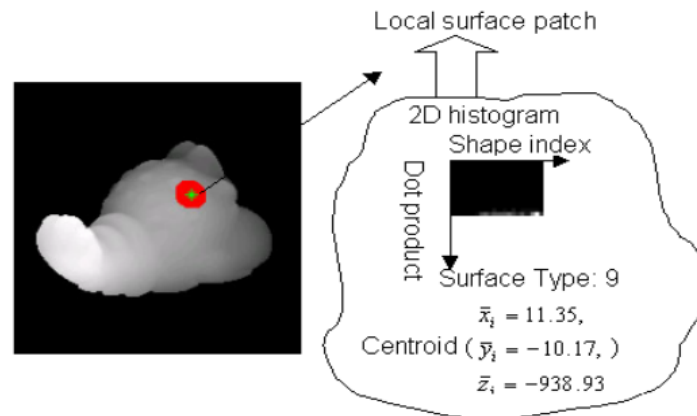


Figure 4-5 Histogramme 2D des indices de formes vs le produit scalaire entre la normale du point référence et ceux de ses voisins dans la représentation LSP (Chen, et al., 2007)

4.1.4 IndThrift

A l'issue de notre étude de ces deux derniers descripteurs, nous avons proposé un nouveau descripteur qui mixe la formulation du descripteur LSP avec celle du Thrift : une variante du descripteur Thrift sous forme d'histogramme 2D. Le nouveau design s'inspire du formalisme de l'histogramme 2D du LSP et remplace la mesure du cosinus entre la normales référence et les normales du voisinage par celle proposée dans le Thrift. Nous appelons ce descripteur: IndThrift (Indice de forme + Thrift).

Plus clairement, pour chaque point du support du PI, deux rayons sont définis et deux normales sont associées à chacune des deux fenêtres correspondantes à ces deux rayons. La formation de l'histogramme 2D se fait comme suit: Pour chaque point v du voisinage du PI, s'incrémente une case de l'histogramme correspondant au couple (indice de forme de v , angle entre les deux normales de v). L'axe du cosinus cumule les valeurs des cosinus de l'angle entre les deux normales de chaque point du support du PI. L'axe de l'indice de forme cumule les valeurs de l'indice de forme des points appartenant au voisinage. Un passage par une interpolation bilinéaire sur les deux axes apporte plus de

robustesse et de stabilité à ce nouveau descripteur.). Notons que ce nouveau descripteur ne définit pas une base locale référence associée au PI (une propriété héritée du descripteur Thrift). Dans notre étude comparative, nous suggérons d'implémenter, puis d'évaluer ce descripteur en termes de pouvoir descriptif, invariance aux rotations et à l'échelle, et robustesse au bruit et à la densité des points.

4.1.5 Descripteur SHOT

Un troisième descripteur qui a inspiré notre démarche de description est le SHOT (Signature of Histograms of Orientations) proposé par Tombari et al. (Tombari, et al., 2010). Les atouts de ce descripteur de forme regroupent, d'un côté, la définition d'un système de référence 3D unique et non ambiguë (RF), et d'un autre côté l'utilisation d'une information géométrique sur la position des points dans le support. Le pouvoir descriptif et la robustesse au bruit, occultation, translations et rotations, ont été démontré par (Tombari, et al., 2010) dans une étude comparative.

L'estimation des RFs locales se fait par le calcul des directions des vecteurs propres (EVD : Eigen Vector Directions) d'une nouvelle matrice de covariance inspirée de la méthode du Total Least Squares (TLS) pour l'estimation des directions des normales. Cette matrice M est définie comme combinaison linéaire des distances des points appartenant à un support sphérique du PI et donnée par :

$$M = \frac{1}{\sum_{i:d_i \leq R} (R-d_i)} \sum_{i:d_i \leq R} (R-d_i)(\mathbf{p}_i - \mathbf{p})(\mathbf{p}_i - \mathbf{p})^T$$

$$\text{avec } d_i = \|\mathbf{p}_i - \mathbf{p}\|_2$$

Équation 4.9

Dans cette expression, un poids faible est attribué aux points les plus éloignés et tous les points contenus dans un support sphérique de rayon R sont utilisés pour le calcul de M . Le centroïde du support est remplacé par le point référence \mathbf{p} . Les vecteurs propres de cette matrice forment des directions orthogonales, répétables et robustes au bruit. La normale correspond au vecteur propre associé à la plus faible valeur propre. A ce niveau pour avoir une RF répétable, les vecteurs propres doivent être non ambigus. L'idée est de réorienter le signe de chaque vecteur propre d'une façon à ce qu'il soit cohérent avec le signe de la majorité des vecteurs qu'il représente. Le signe des axes x et z est déterminé avec ce principe et l'axe y est obtenu par le produit vectoriel $z \times x$.

Le descripteur est formé par des histogrammes locaux définis sur des volumes 3D dans une grille sphérique 3D superposée sur le support des points à la manière du SIFT. Cette grille partitionne l'espace selon les axes radial, azimut et élévation. Dans la formulation du descripteur SHOT, le nombre de bins spatial est fixé à 32 qui est le résultat de 8 divisions en azimut, 2 divisions en élévation et 2 divisions en radial, représentées dans la Figure 4-6.

Pour chaque histogramme local, les points sont cumulés en fonction de l'angle, θ_i , entre la normale de chaque point \mathbf{n}_{vi} du support et la normale \mathbf{n}_u du PI. La fonction $\cos\theta_i = \mathbf{n}_u \cdot \mathbf{n}_{vi}$ est déduite pour chaque couple de normales. La répartition de la taille des cases ne se fait pas uniformément sur l'intervalle des valeurs du cosinus. Un pas grossier est donné pour les cases correspondant aux directions proches de la normale référence et un pas plus fin est utilisé pour les cases où les directions sont orthogonales. En effet, les directions les plus orthogonales à la direction de la normale référence coïncident avec des zones de forte variation locale de la forme (une zone de forte courbure) et sont alors plus riche en

informations. Ce choix permet également de réduire l'impact du bruit qui peut subsister sur des régions quasi planes et qui sont moins descriptives. Sur ces régions, le cumul des cosinus se fait sur un petit nombre de cases de l'histogramme, ce qui permet de réduire la différence entre les histogrammes bruités lors de la mise en correspondance.

La généralisation du design du SHOT est présentée dans ce qui suit.

Notons par $SH_{G,f}(P)$ la signature générique des histogrammes sur le support sphérique autour d'un point P . L'entité G est le vecteur des valeurs des propriétés du sommet, et l'entité f est la métrique utilisée pour comparer ces propriétés. Pour calculer un histogramme de signature, la fonction f est appliquée sur toutes les paires (G_P, G_Q) , avec Q représentant un sommet générique appartenant au support sphérique autour de P . Dans la formulation du SHOT, G correspond à l'estimation de la normale, N , et $f(\cdot)$ est le produit scalaire noté par $p(\cdot)$;

$$f(G_P, G_Q) = p(N_P, N_Q) = N_P \cdot N_Q$$

Équation 4.10

Après le calcul de m signatures d'histogrammes relatives à différentes paires (propriété, métrique) sur la sphère, ces histogrammes sont juxtaposés (Équation 4.11) pour former le descripteur final $D(P)$ au point P .

$$D(P) = \bigcup_{i=1}^m SH_{(G,f)}^i(P)$$

Équation 4.11

Dans la structure du SHOT, un effet de bord peut apparaître pour une perturbation dans le calcul de la RF locale. Pour réduire l'impact d'une telle perturbation, une interpolation quadra-linéaire est faite, premièrement, sur les cases voisines dans l'histogramme local, et deuxièmement, sur les cases ayant le même indice dans l'histogramme et qui correspondent aux volumes voisins dans la grille spatiale. Plus précisément, pour chaque nouvelle entrée l'incrémentation se fait par une multiplication d'un poids égale à $1 - d$ dans chaque dimension. Au niveau de l'interpolation dans l'histogramme, d est la distance de l'entrée actuelle par rapport à la valeur centrale de la case. Dans la grille spatiale, d est la distance angulaire de l'entrée par rapport à la valeur du centre du volume, pour les deux dimensions relatives à l'élévation et à l'azimut. Et concernant la dimension radiale, d est la distance euclidienne de l'entrée par rapport à la valeur du centre du volume. Selon la dimension en question, la distance d est normalisée par la distance entre les voisins des cases de l'histogramme ou par la distance entre les volumes de la répartition spatiale. La robustesse aux variations de l'échantillonnage est réalisée par une normalisation du descripteur. En effet, les différences locales sont la source de l'information discriminante du descripteur. De plus, cette normalisation assume que la densité de l'échantillonnage des points varie d'une façon globale, contrairement à la méthode de Frome et al. (Frome, et al., 2004). L'approche pointée par (Johnson, et al., 1999) pour éliminer l'effet de bord est également utilisée. Ce formalisme a permis d'augmenter la répétabilité et la robustesse du descripteur tout en optimisant le temps de calcul. Les paramètres influant sur la performance du SHOT sont essentiellement le rayon du support et le nombre de bins dans l'histogramme.

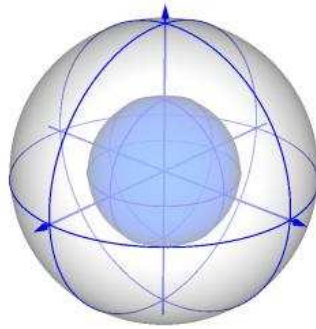


Figure 4-6 Structure spatiale du descripteur SHOT: 8 divisions en azimut, 2 divisions en élévation et 2 divisions en radial (par soucis de clarté seulement 4 divisions azimut sont représentées) (Tombari, et al., 2010)

4.1.6 CSHOT

La formulation du descripteur SHOT a été étendue pour ajouter l'information de la texture à la représentation de la forme et a donné le descripteur Color-SHOT (CSHOT) (Tombari, et al., 2011). Ce descripteur est composé d'une juxtaposition des histogrammes des normales et des histogrammes de couleurs sur le support de la grille sphérique. Une meilleure performance que celle du SHOT est réalisée avec cette version combinée sur des bases de données publiques.

D'après la formulation généralisée (Équation 4.11) du SHOT, le paramètre m prend la valeur 2 dans le cas particulier du CSHOT pour exprimer la combinaison de deux signatures : celle des histogrammes de mesures de forme et celle des mesures de texture. La première signature correspond exactement à la formulation originale du SHOT. Pour la définition de la deuxième signature, un vecteur des intensités des triplets RGB désigné par R représente les propriétés de la texture et la métrique L_1 (Équation 4.12) est utilisée pour comparer les triplets RGB entre eux. Au final, l'histogramme de la texture est formé par les valeurs de la somme absolue des différences entre les valeurs des triplets RGB du point référence et ceux des points voisins dans le support.

$$l(R_P, R_Q) = \sum_{i=1}^3 |R_P(i) - R_Q(i)|$$

Équation 4.12

Le descripteur CSHOT hérite des paramètres du SHOT à savoir le rayon du support et le nombre de bins dans chaque histogramme. Un nombre de bin différent pour chaque type d'histogramme est utilisé étant donné la différence dans la nature des deux histogrammes de signatures.

Certes le pouvoir descriptif du CSHOT a amélioré la performance du SHOT, néanmoins, l'information de la texture n'étant pas donnée par tous les capteurs 3D, l'utilisation de ce descripteur est restreinte.

4.1.7 IndSHOT

Inspiré des formulations des deux descripteurs décrits plus haut le LSP et CSHOT, nous avons proposé de définir un nouveau descripteur qui combine les principes de ces deux derniers pour les raisons suivantes :

- Le succès du descripteur CSHOT revient, d'une part, à la définition d'une RF robuste et invariante aux transformations géométriques et, d'une autre part, à la façon de juxtaposer et normaliser les signatures de deux histogrammes. En comparaison à l'état de l'art, l'hybridité signature-histogramme de ce descripteur a permis, grâce à la propriété des signatures, de hausser le pouvoir descriptif et d'augmenter la robustesse au bruit et à l'échantillonnage, une caractéristique héritée de celles des histogrammes. Comme l'attestent les résultats dans (Tombari, et al., 2011), le CSHOT a surpassé la performance de la forme originale SHOT avec l'ajout de l'information de texture (couleur). Cependant, cet apport n'est exploitable qu'en présence de données de texture (couleur). L'idée, à ce niveau, est de proposer une autre mesure qui caractérise la forme au lieu de la texture, et donc de substituer à la couleur cette mesure dans la formulation du CSHOT.
- Concernant le descripteur LSP, l'histogramme 2D proposé combine deux mesures de la forme locale sur le support: la première est le cosinus des normales qui est invariant aux transformations géométriques 3D et la deuxième est la valeur de l'indice de forme qui est invariante à l'échelle et à la résolution.

Suite à ces constatations, notre suggestion est d'utiliser l'histogramme de mesures de l'indice de forme comme deuxième signature d'histogramme à la place de l'histogramme des mesures de la texture utilisé dans le CSHOT. Dans le design généralisé du SHOT, nous aurons, pareillement que pour le CSHOT, une première signature d'histogrammes des mesures du cosinus de l'angle entre la normale référence et les normales des points du support, une deuxième signature d'histogrammes de mesures des différences entre la valeur de l'indice de forme du point référence et celle de son voisinage. Pour cette deuxième signature, G est l'indice de forme noté par IF et f est la norme L_1 notée par $t(.)$ dans l'Équation 4.13.

$$t(IF_P, IF_Q) = |IF_P - IF_Q|$$

Équation 4.13

Dans notre implémentation de ce descripteur, la normale d'un sommet d'une facette est calculée par l'ajustement d'un plan aux points qui lui sont connectés dans la facette et la normale du point référence (le PI) correspond à la RF. Nous notons ce nouveau descripteur par IndSHOT (Indice de forme + SHOT). Une illustration du design du IndSHOT est donnée dans la Figure 4-7.

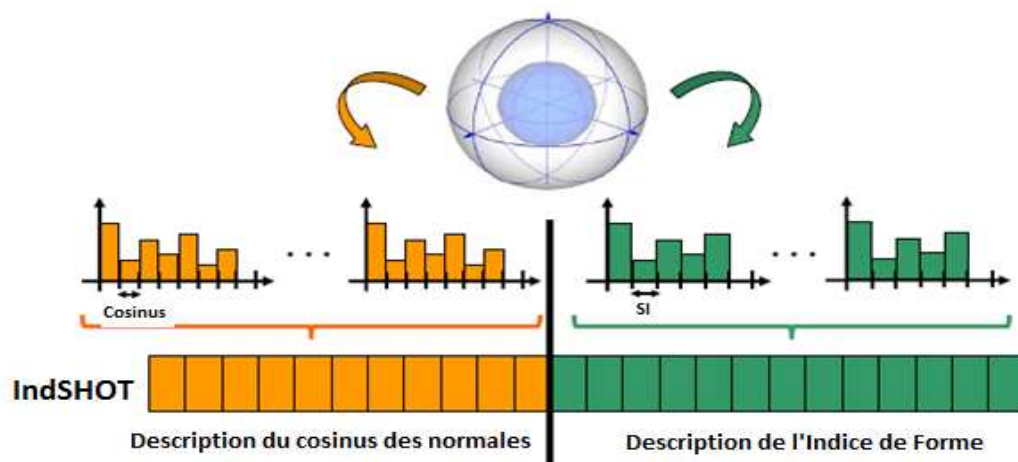


Figure 4-7 Représentation du descripteur IndSHOT

Un résumé des descripteurs présentés concernant leur catégorie et leur définition d'une RF locale unique et non ambiguë est donné dans le Tableau 4-1. Dans la deuxième section de ce chapitre, nous allons évaluer la performance de ces descripteurs.

Méthode	Catégorie	Définition d'une référence
Spin Image (Johnson, et al., 1999)	Histogramme 2D des coordonnées radiale vs les coordonnées d'élévation des points voisins au PI	Axe de référence
LSP (Chen, et al., 2007)	Histogramme 2D des cosinus entre la normale du PI et celle des points voisins vs leurs indices de forme	Axe de référence
Thrift (Flint, et al., 2007)	Histogramme 1D des cosinus entre les normales des points voisins, calculées sur deux rayons	–
IndThrift	Histogramme 2D des cosinus entre les normales des points voisins calculées sur deux rayons vs leurs indices de forme	–
SHOT (Tombari, et al., 2010)	Histogramme 1D des cosinus entre la normale du PI et celle des points voisins	RF Locale
IndSHOT	Histogramme 1D joignant les cosinus entre la normale du PI et celle des points voisins, et les différences de leurs indices de forme	RF Locale

Tableau 4-1 Catégorisation des cinq descripteurs

4.2. Evaluation Expérimentale

Dans cette partie, la distance choisie pour mesurer la similarité entre les histogrammes des descripteurs est la distance euclidienne comme utilisée par (Tombari, et al., 2010) dans leur étude comparative.

4.2.1 Paramètres

Une analyse quantitative des descripteurs est menée pour déterminer les valeurs des paramètres des descripteurs d'une façon qui augmente leur performance et optimise le temps de calcul pour la phase de mise en correspondance.

Dans les expérimentations de cette partie, la base utilisée correspond aux 9 modèles de la base Minolta introduite dans le chapitre Détection. L'évaluation de l'influence de chaque paramètre sur le résultat de la mise en correspondance est donnée par un graphe généré de la façon suivante:

- Pour une valeur du paramètre étudié, le descripteur est calculé sur m positions de points sélectionnés aléatoirement sur le modèle original. Puis, un lissage Laplacien de facteur de relaxation $f = 0.15 \times MR$ (résolution du maillage), est appliqué sur les données du modèle original et le descripteur est recalculé sur les m mêmes positions du modèle transformé (lissé).
- Par la suite, chaque descripteur obtenu du modèle transformé est comparé aux autres descripteurs du modèle d'origine pour trouver les deux premiers plus proches voisins en termes de distance euclidienne entre les histogrammes des descripteurs.
- Nous avons une mise en correspondance entre une paire de descripteurs si le ratio entre la distance euclidienne du premier plus proche voisin et la distance euclidienne du deuxième plus proche voisin est inférieure à un seuil fixé à 0.75. Une paire de descripteurs appariés est dite correcte si elle décrit la même position spatiale dans les modèles où sont calculés les deux descripteurs.
- Ce processus est répété pour tous les descripteurs calculés sur les vues de chaque modèle de la base Minolta. Pour chaque type de descripteur, la moyenne des rappels et des précisions est déduite des vues de chaque objet. Les résultats correspondent à la médiane des taux de rappels et de précisions obtenus sur tous les 9 objets.

Le rappel calcule la fraction du nombre des correspondances correctes sur le nombre total des descripteurs en test. La précision est la fraction du nombre des correspondances correctes sur le nombre total des correspondances. Chaque courbe du graphe correspond à un type de descripteur et indique l'évolution du rappel/précision en fonction du paramètre étudié.

En général, selon le type du descripteur, des traitements spécifiques peuvent améliorer le résultat de la mise en correspondance. Par exemple, dans le cas d'un histogramme avec des « creux » (sparse), la distance euclidienne se calcule sur l'ensemble des paires des cases non vides et de même indice. Nous ne tenons pas compte de ce genre de traitements pour cette partie du manuscrit. Le paramètre m varie entre 50 et 150 points, selon la densité de points des vues, pour toutes les expérimentations de cette section.

Le paramétrage se fait pour les descripteurs Spin, SHOT, LSP, Thrift, IndSHOT et IndThrift. Notons que pour les descripteurs (LSP, IndSHOT et IndThrift), le calcul de l'indice de forme n'est pas effectué dans l'étape de description mais est récupéré de la phase de détection. Le temps de calcul présenté dans cette partie n'inclue pas ce calcul.

4.2.1.1 Choix du nombre de Bins

❖ Description

La taille des Bins est un paramètre important dans la génération des descripteurs à base d'histogrammes et correspond à la largeur géométrique des cases dans les histogrammes. En effet, ce paramètre détermine la marge de l'espacement dans les cases de l'histogramme nécessaire pour répartir et moyenniser les valeurs de la variable cumulée et éviter le cas des points isolés. Un espacement grossier ou plus fin impacte directement sur le pouvoir de description.

La taille du bin dépend du nombre de cases (ou dimension) choisi pour l'histogramme et vice-versa. Le produit de la taille du bin avec le nombre de bins donne la taille de l'intervalle de variation des valeurs, qui correspond à la différence entre la valeur maximale et la valeur minimale de la mesure cumulée.

Une analyse quantitative de l'impact du nombre de bins sur la similarité entre descripteurs appariés est donnée dans la Figure 4-8. Pour simplifier cette étude, nous prenons n comme nombre de bins pour la description du cosinus des normales et $n-2$ comme nombre de bins pour la description de l'indice de forme et des coordonnées cylindriques dans le Spin. Ce qui nous donne une dimension de $(n-2)^2$ pour le descripteur Spin, $n(n-2)$ pour les descripteurs LSP et IndThrift, n pour le Thrift, $32(n+1)$ pour le descripteur SHOT et $32((n-1)+(n+1))$ pour le descripteur IndSHOT. Le 32 correspond au nombre de bins spatial. En variant la valeur du nombre de bins, nous traçons la courbe précision et la courbe rappel. La performance des descripteurs est déduite du comportement des ces courbes: plus la valeur du rappel et de la précision est grande plus le descripteur est performant.

❖ Résultats

A ce niveau, nous cherchons la valeur n des bins qui puisse maximiser le rappel et précision pour la majorité des descripteurs. D'après ces courbes, d'une manière générale, en augmentant le nombre de bin, la précision et le rappel augmentent légèrement et se stabilisent pour le SHOT et IndSHOT, de même pour le comportement de la courbe de précision du LSP, Spin et IndThrift. Quant au rappel, il augmente puis diminue pour le LSP, Spin et IndThrift. Le Thrift continue à augmenter pour les deux courbes.

Par ailleurs, nous remarquons que les valeurs du rappel et de précision pour le IndSHOT sont les plus élevés et celle du Thrift s'écarte largement des autres courbes lui inférant une faible performance.

En outre, pour une valeur fixe du nombre de bin, ces courbes renseignent aussi sur le degré de robustesse des descripteurs pour un lissage de la structure du maillage. Nous allons voir cette propriété plus tard dans ce chapitre.

Pour le cas des petites valeurs du nombre de bin, plusieurs valeurs sont moyennées pour correspondre à une seule case et le pouvoir discriminant du descripteur diminue. En effet, une structure très condensée de l'information ne traduit plus la distribution réelle de la forme. Le résultat est que plusieurs histogrammes auront la même forme sans pour autant correspondre à des régions similaires de la surface. Par ailleurs, pour des grandes valeurs du nombre de bins, la distribution devient éparse (avec beaucoup de cases vides) et l'information locale est dispersée sur plusieurs cases en donnant des points singuliers (isolés). Certes cette description est la plus proche de la distribution réelle de la mesure cumulée, mais nous perdons la compacité et la robustesse au bruit et à la densité des points. Comme l'illustre la courbe de la Figure 4-8, plus le nombre de bins augmente, plus le rappel décroît traduisant la diminution des correspondances correctes. Pour des valeurs du bin dans l'intervalle 10 à 20, et mis à part le Thrift, la précision des descripteurs est au dessus de 0.9 et le rappel est quasiment à son maximum.

Le temps de calcul des descripteurs n'est pas affecté par le choix du nombre de bins. Cependant, ce paramètre influe sur le temps de calcul pour la phase de mise en correspondance entre les descripteurs.

- Le choix du nombre de bins adéquat doit se faire sur la valeur qui satisfait le compromis entre, d'un côté, la compacité et la robustesse (une faible valeur du bin) du descripteur et d'un autre côté, son pouvoir descriptif (une grande valeur du bin). Notre choix est porté sur la valeur $n = 12$.

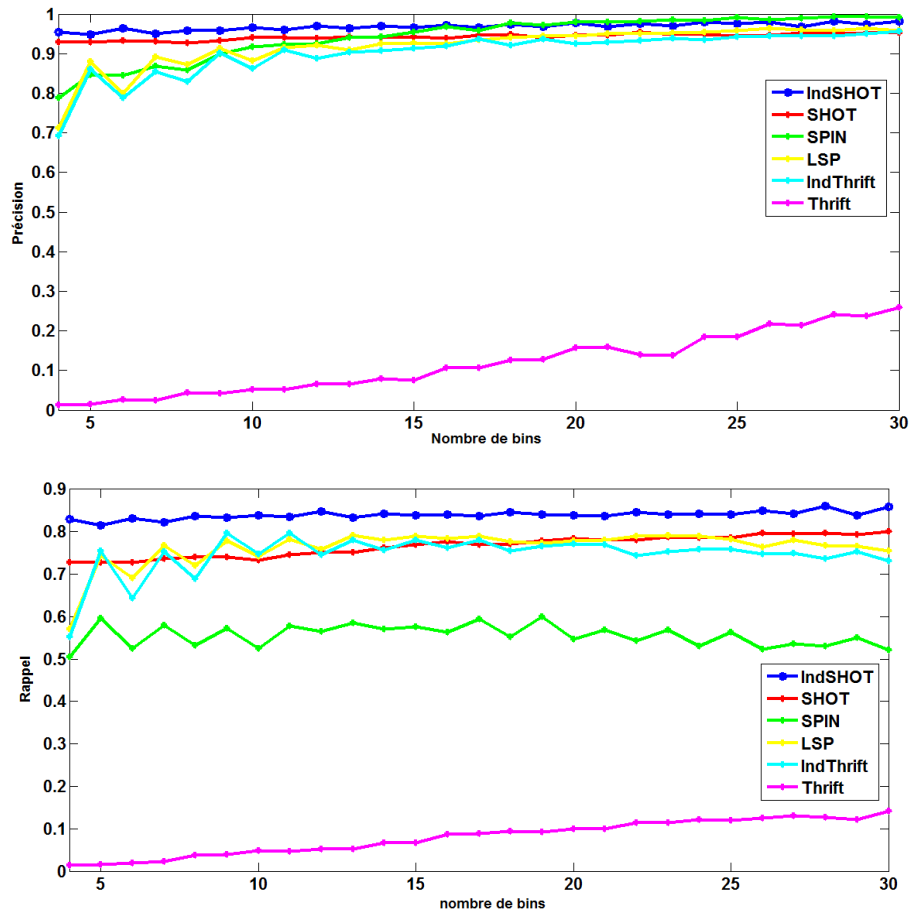


Figure 4-8 Courbe rappel et courbe précision en fonction du nombre de bins

4.2.1.2 Choix du rayon du support

❖ Description

Dans ce paragraphe, nous nous intéressons à la valeur du deuxième paramètre: rayon du support r utilisé dans le calcul des descripteurs des PIs. Du fait que nous travaillons à une échelle fixe, la phase de détection ne nous fournit pas une échelle de description. Le rayon du support utilisé pour le calcul des descripteurs est le même pour tous les PIs et est fixé en paramètre par l'utilisateur. Sa valeur est proportionnelle à la taille de la diagonale de la boîte englobante de l'objet pour assurer l'invariance à la densité des points et à l'échelle. Notons que le rayon du support dans le Spin Image est le produit de la taille du bin par le nombre de bins dans chaque dimension de la Spin Image.

Dans la Figure 4-9, nous traçons la courbe rappel et la courbe précision en fonction du rayon du support. Et dans la Figure 4-10, nous montrons l'évolution du temps de calcul des descripteurs en fonction de la taille du rayon, pour une vue de l'objet Duck.

❖ Résultats

Le rayon du support du PI influe sur la distance de mise en correspondance de la façon suivante: D'une part, plus le rayon est grand plus le voisinage est grand et plus les mesures calculées (indice de forme ou cosinus entre les angles de normale ou les coordonnées cylindriques) sur ce support sont lissées et stables. D'autre part, pour un rayon de support plus grand, des nouveaux voisins plus éloignés sont ajoutés et des nouvelles mesures sont cumulées dans l'histogramme. L'allure de

l'histogramme obtenu est modifiée par rapport à celle d'un histogramme résultant d'un rayon de plus petite taille. Le type de la forme locale autour du PI intervient aussi sur la stabilité du descripteur : une zone plane sur un rayon r est plus stable pour un changement de rayon qu'une zone sous forme de dôme, par exemple.

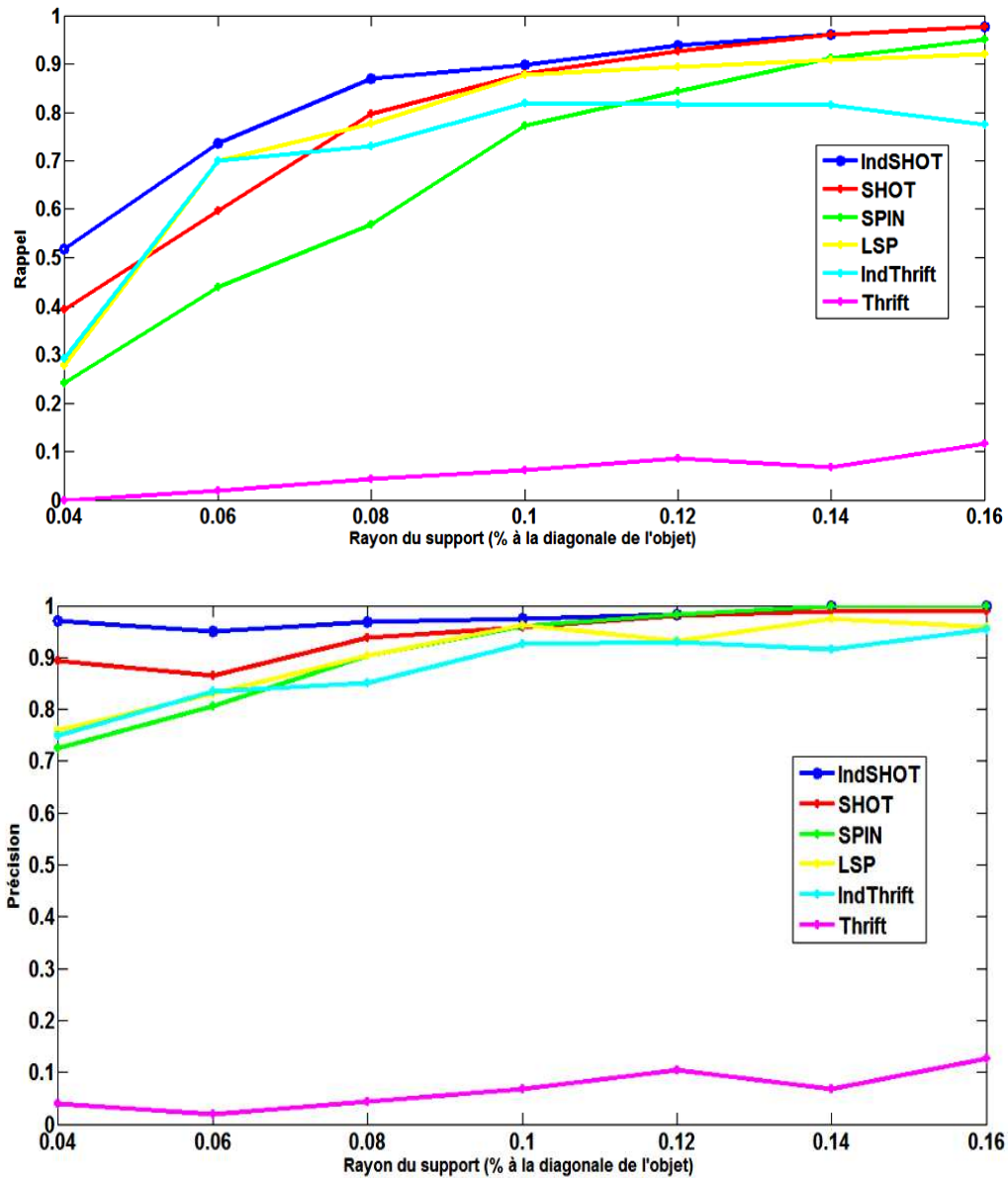


Figure 4-9 Courbe précision et courbe rappel en fonction du rayon du support

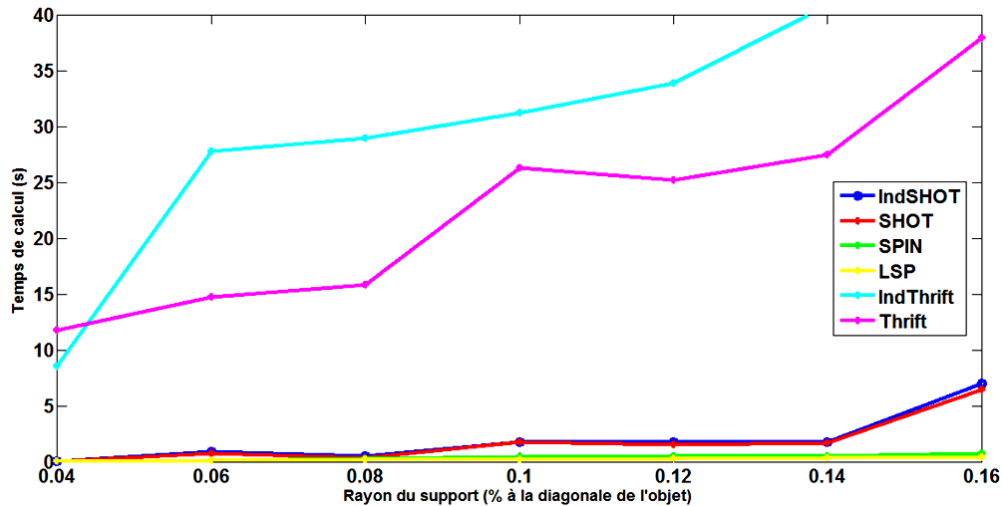


Figure 4-10 courbe temps de calcul en fonction du rayon du support

Ces remarques sont consolidées par le résultat de la Figure 4-9 où l'appariement est plus correct pour des plus grandes valeurs du rayon (entre 0.08 et 0.16) pour tous les descripteurs. Le rappel augmente puis décroît, pour le IndThrift entre les valeurs 0.08 et 0.12 du rayon. Nous pensons, qu'à partir d'une certaine valeur du support, le descripteur devient stable puis sa performance décroît (tel le cas pour le IndThrift). Cette valeur représente la taille limite et maximale du support pour laquelle la structure locale décrite change de forme et par conséquent, le descripteur change d'allure. D'après les deux courbes rappel et précision, cette limite pourrait être atteinte à partir de la valeur 0.08 x diagonale du support. Or, pour obtenir un meilleur pouvoir descriptif, le support doit avoir une taille proportionnelle à une structure de petite taille pour pouvoir moyenner suffisamment de détails localement sur la surface. Les structures de grandes tailles nous ramènent à une description de l'ordre des propriétés globales des modèles et nous sortons d'une approche locale.

Par ailleurs, la contrainte majeure est le temps de calcul qui clairement croît avec l'augmentation de la taille du support. Nous notons que le temps de description du IndThrift et du Thrift, pour les faibles tailles du rayon est d'environ 10 fois plus important que celui des autres descripteurs.

- Le choix de la valeur du rayon du support doit alors répondre au compromis efficacité-performance des descripteurs. D'après les courbes obtenues, une proportion de 0.08 de la taille de la diagonale de l'objet paraît convenable pour satisfaire ces deux contraintes.

4.2.1.3 Choix de l'angle du support

Une autre condition à vérifier pour choisir les points du voisinage du point référence est l'angle du support de ce point. Cet angle correspond à l'angle maximum formé par la normale du point référence et la normale d'un point du voisin. Cet angle est introduit pour minimiser l'impact des occultations et du bruit. Pour un point référence **A** de position et normale $(\mathbf{p}_a, \mathbf{n}_a)$, un point **B** de position et normale $(\mathbf{p}_b, \mathbf{n}_b)$ est dans le support de **A** si :

$$\text{Acos}(\mathbf{n}_a, \mathbf{n}_b) < \text{Angle_support}$$

Le choix d'une faible valeur de cet angle peut ne pas être suffisant pour décrire la forme locale du voisinage du point référence (sauf dans le cas des zones planes). Néanmoins, un angle faible de support permet d'assurer une robustesse au bruit et aux occultations. En effet, en général, la direction d'une normale d'un point aberrant s'écarte largement de celle des normales des points de la surface locale du support. Par conséquent, un très grand angle de support ($>150^\circ$), inclue un grand nombre de voisins qui peuvent correspondre à des points bruités et faussent les valeurs cumulées dans les histogrammes. D'un autres coté, plus l'angle est grand, plus la direction des normales voisines est orthogonale à la direction de la normale du point référence, ce qui permet de représenter une information plus riche sur la forme locale de la surface autour du point référence (cas de zone sous forme dôme ou selle). En outre, les angles faibles correspondent à des points situés à peu près sur la même zone plane que le point référence et ne représente pas forcément une information très descriptive de la forme locale.

Face à ce compromis, notre choix final de la valeur de l'angle se fait sur la valeur qui réalise la balance entre le pouvoir descriptif et la robustesse du descripteur. Pour les tests de ce chapitre nous avons choisi la valeur 90° .

4.2.2 Comparaison des descripteurs

Cette évaluation suit un principe similaire à celui utilisé pour les paramètres dans le paragraphe précédent. Pour ce faire, les descripteurs sont calculés sur le modèle d'origine et sur le modèle transformé (modèle test) pour les mêmes positions spatiales des PIs. Pour avoir une comparaison la plus juste possible entre les descripteurs et pour que la performance des descripteurs ne soit pas affectée par les erreurs apportées par les détecteurs, nous utilisons le même « détecteur » de PIs pour tous les algorithmes : nous extrayons aléatoirement un ensemble PIs sur chaque modèle (un nombre entre 50 et 150), puis nous extrayons les points correspondants physiquement sur le modèle transformé. La mesure de distance euclidienne est utilisée pour la phase de mise en correspondance entre chaque descripteur du modèle test et tous les descripteurs du modèle original. Pour décider de l'appariement, le ratio entre la distance du plus proche voisin et le deuxième plus proche voisin, est utilisé : si le ratio est inférieur à un seuil th_d , la correspondance est établie entre le descripteur test (du modèle transformé) et le descripteur du premier plus proche voisin du modèle d'origine. La correspondance des descripteurs est validée si les positions spatiales des deux points de calcul se correspondent.

Comme méthodologie d'évaluation des descripteurs 2D, la courbe ROC, représentant le taux des vrais positifs versus le taux des faux positifs, est généralement utilisée pour comparer les descripteurs. Néanmoins, l'ambiguïté dans le calcul du taux des faux positifs défavorise cette méthode. Nous adoptons alors dans notre évaluation l'approche proposée par Mikolajczyk et Schmid dans (Mikolajczyk, et al., 2005) et nous présentons le résultat en termes de courbes Rappel versus 1-Précision. Pour obtenir ces courbes, nous varions la valeur du seuil th_d . Nous commençons par l'évaluation de l'invariance des descripteurs pour des transformations géométriques rigides (rotation et translation).

4.2.2.1 Invariance aux rotations

❖ Description

Parmi les objectifs de notre travail, la caractérisation de la forme de l'objet d'une façon invariante à un changement d'angles de vue est très importante. En effet, en test, un même objet doit être reconnu pour une prise de vue différente de celles stockées dans la base de données. L'étape de description doit satisfaire la propriété d'invariance aux transformations géométriques.

Un changement de l'angle de vue peut se présenter soit par une rotation du capteur autour de l'objet soit par la rotation de l'objet autour d'un capteur fixe. Dans notre expérimentation, nous nous situons dans le deuxième cas. Dans la base Minolta, nous sélectionnons 5 vues, avec un écart d'angle de 20°, pour chacun des 9 objets. L'idée, à ce niveau, est de considérer pour chaque objet 4 transformations de rotation autour de l'axe y entre les vues adjacentes (exemple une transformation de la vue 0° vers la vue 20°). Le processus de calcul des descripteurs et de mise en correspondance, décrit plus haut, est appliqué pour l'ensemble des vues des 9 objets.

Une illustration du résultat du processus de mise en correspondance des descripteurs IndSHOT, sur un modèle d'Ange, entre la vue initiale située à 100° et la vue située à 120°, est donnée dans la Figure 4-11. Une évaluation quantitative est illustrée, sous forme de courbes: rappel vs 1-précision, dans la Figure 4-12. Nous illustrons, dans la Figure 4-13, les images des histogrammes des six descripteurs calculés au même point physique (le point n°1 sur les deux vues de la Figure 4-11) entre deux vues.

❖ Résultats

- La Figure 4-11 montre que le descripteur IndSHOT a correctement apparié 34 points parmi 65 points mis en correspondance sur les 71 points de calcul sur les deux vues, ce qui donne un rappel de 47% et une précision de 52%. Un résultat qui annonce une forte stabilité de ce descripteur pour une variation d'angles de vue.
- Sur les histogrammes de la Figure 4-13, nous remarquons la forte similarité entre les deux images des histogrammes des descripteurs IndSHOT et IndThrift. D'ailleurs, lors du processus de mise en correspondance, ceux deux points sont correctement appariés entre eux pour ces deux descripteurs. Bien que les deux histogrammes du Thrift se ressemblent, la remarque est que ce descripteur a la même allure quasiment sur tous les points de la surface de l'objet, ce qui augmente le nombre faux homologues. En outre, les images du Spin et du LSP, représentent des matrices avec plusieurs cases vides qui n'ont pas la même position sur les deux images des deux vues. Or, la mesure de similarité utilisée dans notre cas est la distance euclidienne qui compare les histogrammes bin à bin. Il en résulte que la distance de mise en correspondance est grande pour ces descripteurs et les points ne sont pas appariés.
- D'une manière générale, en augmentant le seuil de distance pour la mise en correspondance des descripteurs, le nombre total des correspondances augmente, et nous nous attendons à ce que la précision ($TP / (TP + FP)$) diminue et le rappel ($TP / (TP + FN)$) augmente.
 - Les résultats de la Figure 4-12 montrent que le IndSHOT est de très loin plus robuste à une rotation des données initiales. En effet, la courbe des valeurs du rappel vs la précision de ce descripteur est nettement au-dessus des autres courbes. Ce descripteur a amélioré la stabilité du descripteur d'origine SHOT.
 - Le pouvoir d'invariance porté par le IndThrift est inférieur aux deux précédents mais finit par rattraper le SHOT à partir d'une certaine valeur du seuil du ratio ($th_d > 0.65$) des distances.

- Cette invariance aux rotations est due essentiellement à l'utilisation des bases locales (RF) dans la formulation de ces descripteurs. La description de la variation locale de la forme autour du PI relativement à une RF (SHOT et IndSHOT) est invariante par rotation. De plus, une normale calculée en un point est la même quel que soit l'angle de la prise de vue. Par ailleurs, une similitude conserve les angles géométriques (les mesures d'angles non orientés). En outre, l'indice de forme est constitué des courbures principales qui correspondent à un calcul différentiel de second ordre de la surface (présenté dans le chapitre 1) et sont invariants aux orientations. Avec ces caractéristiques, les descripteurs SHOT, IndThrift et IndSHOT cumulant les cosinus des normales et/ou les valeurs de l'indice de forme, sont discriminants et invariants aux rotations.

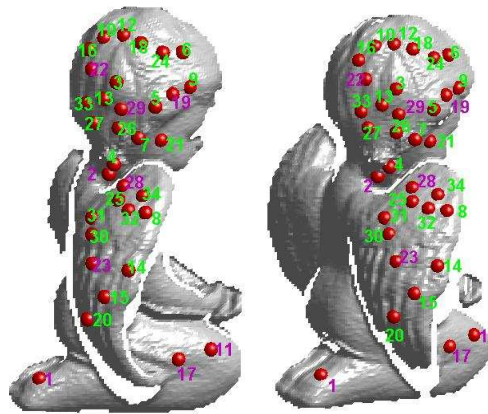


Figure 4-11 Résultat de la mise en correspondance après rotation des PIs entre la vue initiale à 100° et la vue à 120° des descripteurs IndSHOT

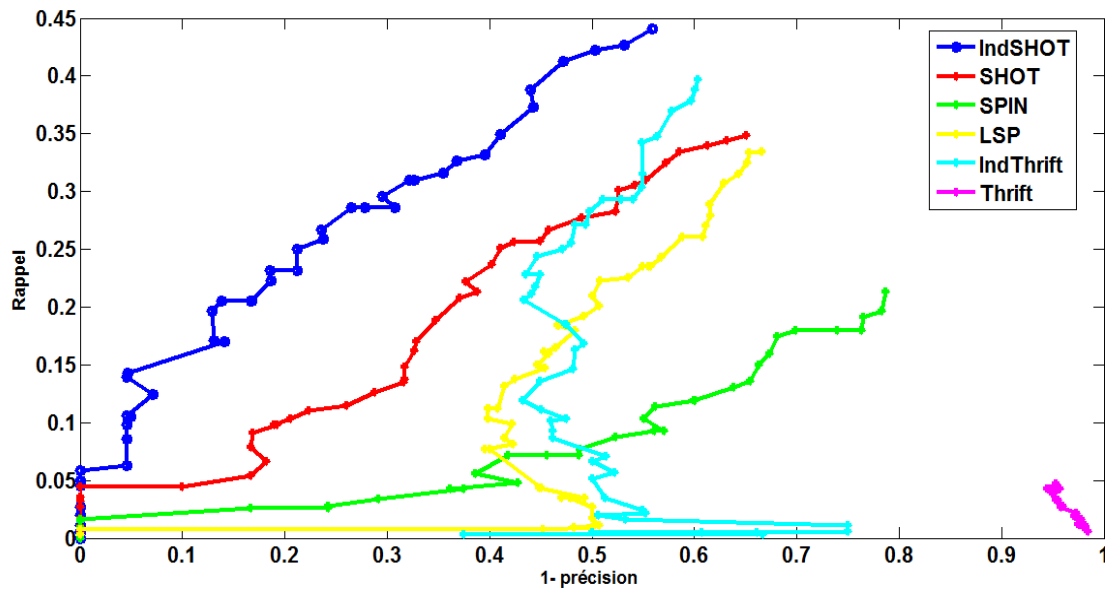


Figure 4-12 Courbes Rappel vs 1-Précision pour un changement d'angles entre les modèles Minolta

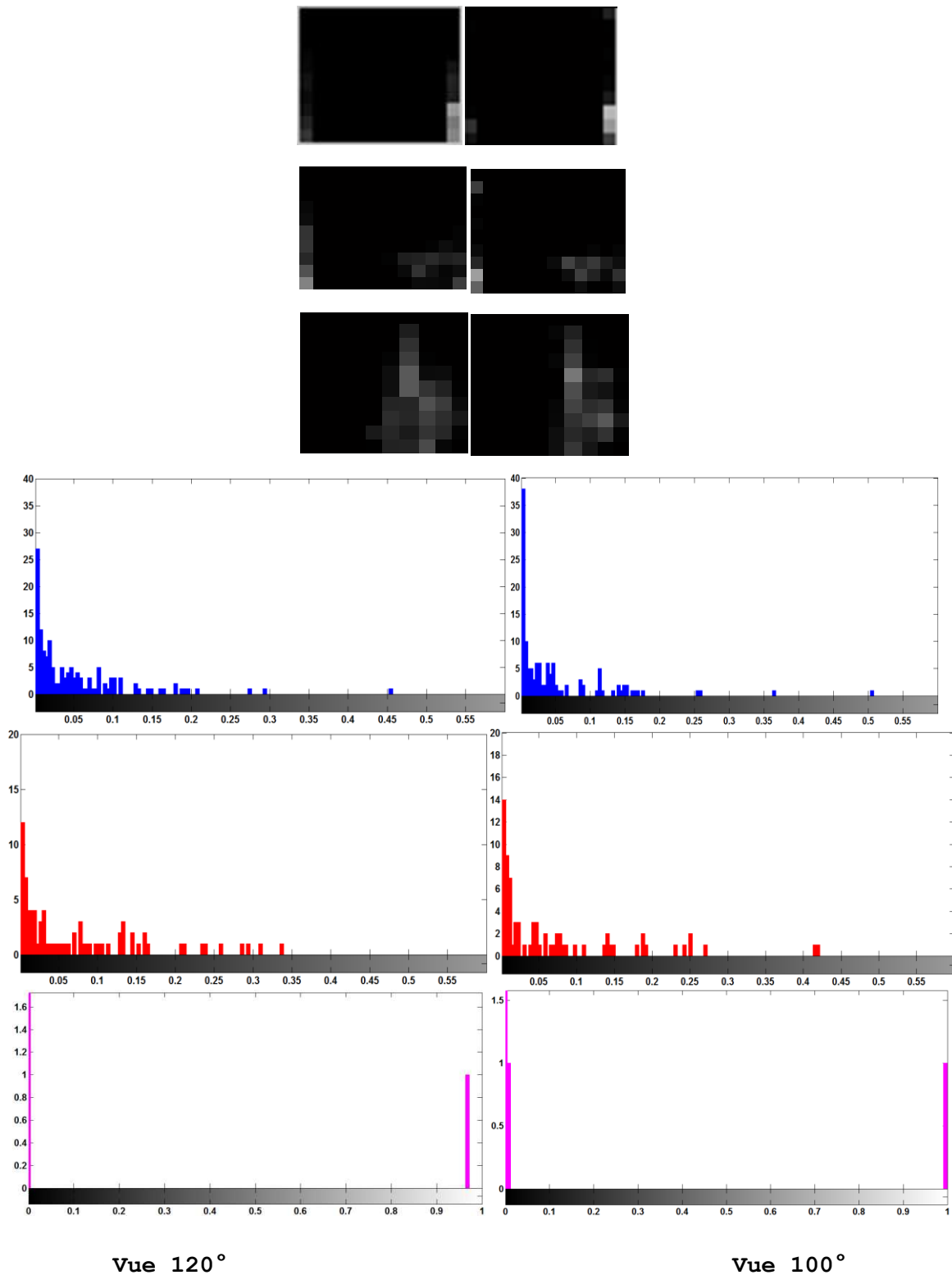


Figure 4-13 Représentations, au même point physique situés sur deux vues (100° et 120°), des descripteurs: IndThrift, LSP, Spin, IndSHOT, SHOT et Thrift, dans l'ordre de haut vers le bas

- Le LSP a un comportement assez proche de celui du IndThrift, puisqu'ils encodent la forme locale avec la même information (normales et indice de forme).
- Le Spin, en dépit de sa propriété d'invariance aux rotations, montre une très faible stabilité. Rappelons que pour un point orienté sur une surface 3D, le Spin image est

décrite par deux paramètres: distance du plan tangent du point orienté à ses voisins et la distance du point orienté à son vecteur normale. Sachant qu'une rotation est une similitude, les distances sont conservées. Les deux mesures de distance du Spin sont alors invariantes par rotation. Or, dans notre évaluation, nous considérons un changement d'angles dans des conditions d'acquisition réelle. Ce qui implique que pour une même région de la surface vue sous deux angles différents, la densité des points et l'amplitude du bruit varient. La sensibilité du Spin à ces variations justifie le résultat obtenu.

- Le Thrift a la plus faible performance. En effet, bien que les mesures cumulées correspondent à des mesures de cosinus de normales, la première lacune de ce descripteur est l'absence d'une RF ou d'une RA. De plus, les cosinus correspondent aux cosinus de l'angle entre deux normales du même point calculées sur deux supports de tailles différentes. Ces angles sont très faibles et varient sur un petit intervalle de l'histogramme (le cosinus est entre 0.8 et 1) et ne renseignent pas suffisamment sur la forme locale. Avec la formulation du Thrift, plusieurs descripteurs ont la même allure et se confondent lors de la mise en correspondance, ce qui baisse le taux de rappel. Par ailleurs, l'ajout de l'indice de forme à ce descripteur, avec la définition du IndThrift, a nettement haussé sa performance.

La conclusion pour cette évaluation est l'importance de l'utilisation d'une base de référence pour la description locale du support. Le IndSHOT l'emporte pour l'invariance à la rotation sur les autres descripteurs. Pour notre évaluation le changement d'angles se fait sur des données réelles. Ce qui nous rapproche d'un scénario de mise en correspondance de vues partielles d'objets pour une tâche de reconnaissance. D'après ces résultats, nous estimons que ce descripteur réussira à correctement apparier les différentes vues des objets.

4.2.2.2 Invariance à l'échelle et à la densité de points

❖ Description

Dans ce paragraphe, nous traitons la problématique de l'invariance des descripteurs pour un changement d'échelle et pour une variation de la densité du point du maillage. Géométriquement, une transformation de translation ou un changement d'échelle correspond à une homothétie dans l'espace.

Dans un premier temps, nous considérons une décimation du maillage qui consiste à réduire le nombre de facettes et de points dans le maillage initial de chaque objet. Dans la Figure 4-14, Nous traçons les courbes rappel vs précision des 6 descripteurs, correspondant au processus de décimation du maillage initial avec un facteur de 20% de réduction sur les 9 modèles de Minolta.

Dans un deuxième temps, nous considérons la base des 7 objets Minolta présentant une variation d'échelles. Dans notre évaluation, nous avons choisi deux échelles pour chaque objet avec 11 vues sur chaque échelle. Le même processus de génération des courbes rappel-précision est appliqué pour la transformation de changement d'échelle. Ces courbes sont données pour seulement les descripteurs SHOT, IndSHOT et IndThrift dans la Figure 4-15.

❖ Résultats

Variation de la densité des points :

- Après décimation, le maillage obtenu a une valeur de résolution (moyenne des distances entre les sommets du maillage) plus grand et les détails de la forme sont moins visibles. Par exemple, si nous prenons un maillage de 100 facettes, avec une résolution de 0.63 et une décimation de facteur 20%, le maillage transformé aura 80 facettes et une résolution du maillage de 0.72. En augmentant le niveau de décimation, le nombre de points apparié diminue. Nous remarquons nettement qu'avec la décimation, la performance des descripteurs a chuté.
- En variation de densité de points, le IndSHOT a une meilleure performance pour les premières valeurs du seuil sur le ratio des distances des deux premiers plus proche voisins. Les descripteurs LSP et IndThrift ont à peu près le même comportement. A partir du seuil (0.6), la performance de ces deux derniers devient meilleure que celle du IndSHOT.

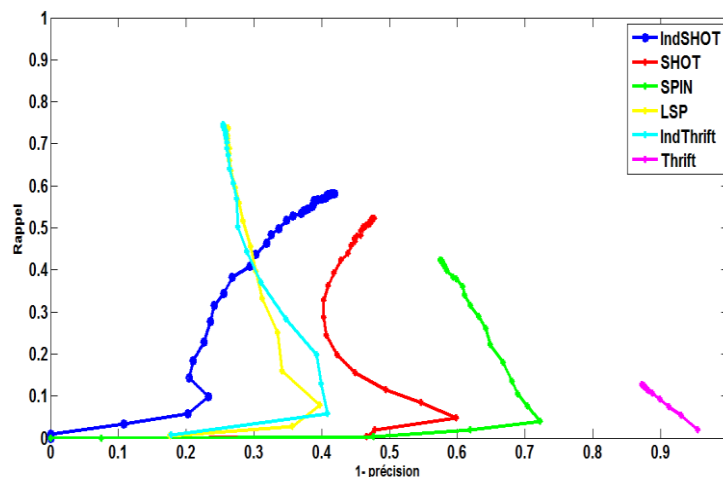


Figure 4-14 Courbe rappel-précision pour une décimation de 20% du maillage initial

Le descripteur Thrift a une très faible invariance à la densité des points avec un rappel < 0.2 . Les mêmes explications données pour l'invariance à la rotation sont valides ici.

La performance du Spin chute considérablement pour une densité plus faible du nuage de point. Pareillemnt que pour le LSP, la base de référence (RA) utilisée pour le calcul des coordonnées cylindrique dépend de la méthode de calcul de la normale en un sommet. Cependant, et sachant qu'une homothétie de rapport k modifie les distances par un facteur $|k|$, les spins images du même point sur deux surfaces échantillonnées uniformément avec deux échelles différentes, sont linéairement corrélés. Une méthode standard pour la mise en correspondance de données corrélées linéairement est le coefficient de corrélation linéaire. L'utilisation de cette mesure de similarité, à la place de la distance euclidienne, augmenterait la performance de l'appariement pour le descripteur Spin. Cependant, pour une comparaison plus égale entre les descripteurs, nous conservons la même mesure de similarité entre les différents descripteurs.

- Il est à noter que ces descripteurs sont sensibles aux perturbations dans le calcul des normales qui est la moyenne des normales des facettes directement adjacentes au point. Une normale

d'une facette est calculée par une méthode basique de produit vectoriel entre les vecteurs des sommets de la facette. Or, en variant la résolution du maillage par une décimation, le même point n'aura plus les mêmes voisins sur les deux maillages et seront plus éloigné spatialement dans le maillage décimé. La normale résultante est, de ce fait, plus au moins différente de la normale initiale.

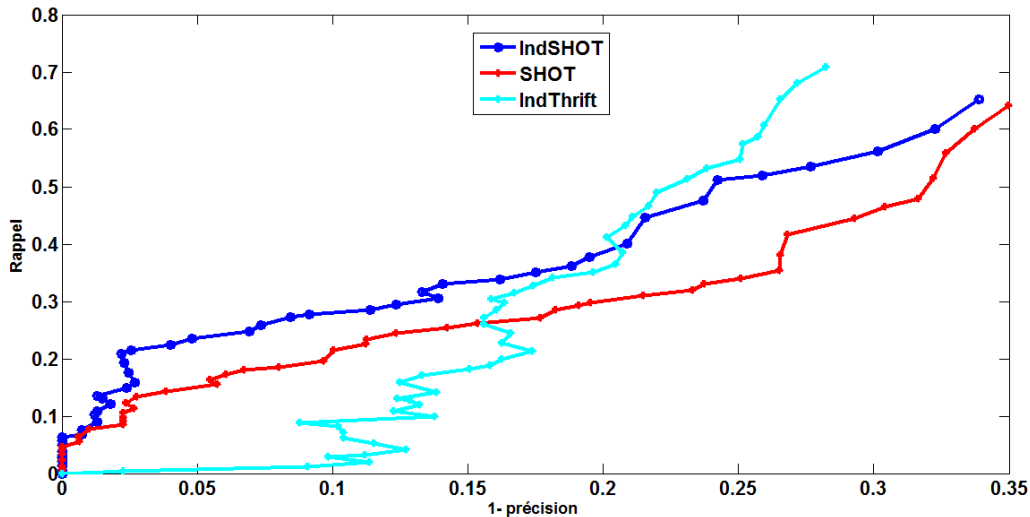


Figure 4-15 Courbe rappel-précision pour un changement réel d'échelle pour les descripteurs SHOT, IndSHOT et IndThrift

Variation d'échelles :

Nous mettons l'accent sur le fait que l'évaluation de la variation d'échelle s'est faite sur des données réelles en prenant des acquisitions des objets à différentes distances. En déplaçant l'objet, sa posture et sa pose (par rapport à l'angle de vue) n'est pas exactement la même. Un décalage dans les positions spatiales de calcul des descripteurs peut fausser notre processus de mise en correspondance. Pour ce fait, nous avons introduit une marge dans la distance seuil (distance moyenne) servant à valider l'appariement des points physiquement.

Nous rappelons que la normalisation dans le formalisme des descripteurs permet d'assurer l'invariance à la densité des points et à l'échelle. Pour une variation d'échelle, d'après les courbes de la Figure 4-15, les descripteurs IndSHOT et SHOT donnent un meilleur taux de précision-rappel pour les faibles valeurs du ratio de comparaison des distances. Nous savons que les angles sont conservés par une homothétie, et que le SHOT utilise une RF invariante pour l'expression des cosinus des angles, la stabilité des descripteurs SHOT et IndSHOT s'explique alors. D'autre part, la caractéristique d'invariance à l'échelle de l'indice de forme consolide la robustesse du descripteur IndSHOT et IndThrift. Pour les grandes valeurs du seuil (>0.75), le IndThrift devient meilleure que le IndSHOT. Néanmoins, pour les grandes valeurs de précision, qui caractérise typiquement les applications temps réel, IndSHOT a un meilleur degré de Rappel.

En variation d'échelle, un rappel $>60\%$ n'est atteint que vers les dernières valeurs du seuil.

- Comme évoqué dans la partie évaluation du chapitre Détecteurs, la décimation des données peut être assimilée à une variation artificielle d'échelle. Avec les résultats présentés en décimation et en changement d'échelles, le IndSHOT encore une fois a une meilleure performance que les autres descripteurs.
- L'invariance à un changement d'échelles semble être le plus grand défi pour les méthodes de description proposées si le support est fixe (mono-échelle). Dans notre choix de la taille du support, nous avons pris un rayon proportionnel à la diagonale de la taille de l'objet. Pour deux échelles différentes, la taille de l'objet ne varie pas et la zone locale prise en compte dans la description est la même. Si le descripteur est suffisamment discriminant et stable, son invariance à une variation de la densité des points est réalisable pour ce choix.
- Nous prévoyons que l'utilisation d'une échelle appropriée (entre autre le résultat de la version multi-échelle) pour la description de chaque point améliorera cette performance.

4.2.2.3 Robustesse au bruit

❖ Description

Pour mesurer l'impact du bruit sur l'étape de description et de mise en correspondance, nous ajoutons un bruit Gaussien sur les données initiales. Nous menons deux expérimentations avec deux valeurs de variance gaussienne σ_1 et σ_2 qui correspondent aux pourcentages respectifs: 10% et 30% de la résolution du maillage de chaque modèle.

Les courbes rappel-précision pour ces deux perturbations du maillage initiales des 9 modèles Minolta (5 vues par objets) sont données dans la Figure 4-16.

❖ Résultats

- Sur les deux perturbations de données, le descripteur IndSHOT est le moins sensible au bruit avec des taux de rappel-précision plus élevés que les autres descripteurs. Cette robustesse au bruit est due essentiellement à la robustesse de la RF et au processus de normalisation du descripteur. Cette robustesse est confirmée par l'ajout de l'information de l'indice de forme puisque une meilleure performance, que celle du SHOT, est réalisée avec cette nouvelle combinaison.
- Le Spin paraît très sensible au bruit dès la faible perturbation de variance ($\sigma_1 = 0.1 \times MR$). En augmentant l'écart type du bruit gaussien, la précision et le rappel décroît et tend vers 0. Ceci est dû au fait que le descripteur est très sensible aux petites variations dans l'estimation de la normale. Ce descripteur est basé sur le cumul des coordonnées cylindrique (radiale et élévation) des points du support, exprimés par rapport à la normale du PI (RA). Or, pour une perturbation des coordonnées des points du maillage, le calcul de la normale référence qui se fait d'une façon non robuste au bruit (par une simple approximation de moindre carrée) est perturbé. Ce résultat est cohérent avec la conclusion donnée par (Tombari, et al., 2010).
- La performance du LSP décroît rapidement en augmentant le degré du bruit. La défaillance de la robustesse du LSP est, similairement au Spin, liée à la non robustesse de l'estimation de la RA et des quantités différentielles (normales) au bruit.

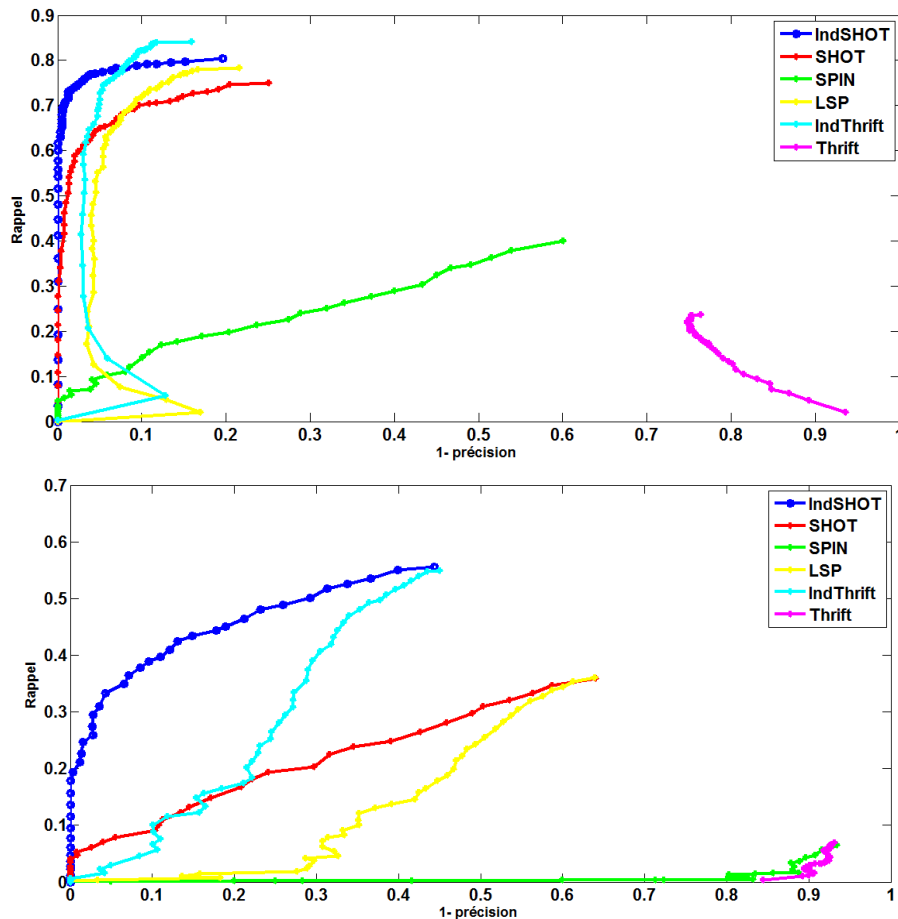


Figure 4-16 Courbe rappel-précision pour un bruit, de $\sigma = 0.1 \times MR$ -en haut, et de $\sigma = 0.3 \times MR$ -en bas

- Le Thrift, basé uniquement sur les normales avec l'absence de notion de base de référence, reste le plus vulnérable à une perturbation des données. Par ailleurs, le IndThrift se situe en deuxième position concernant la robustesse au bruit. Nous concluons que la forme 2D de ce nouveau descripteur joignant le formalisme du Thrift avec l'indice de forme a permis de créer une description plus stable et plus discriminante.

Ces résultats confirment et consolident les constatations données précédemment.

4.2.2.4 Robustesse à un lissage

❖ Description

La modification du maillage par un lissage des données est un autre type de perturbations des données initiales que nous avons étudié dans notre évaluation de la robustesse des descripteurs. Cette modification lisse les détails de la forme et rend difficile la tâche d'extraction d'informations discriminantes, comme l'illustre la Figure 4-17. Nous étudions l'impact d'une transformation de lissage Laplacien des données, avec un facteur de relaxation f de valeur proportionnelle à la résolution du maillage ($0.25 \times MR$), sur la performance des descripteurs. Similairement à ce qui a été présenté pour le bruit, nous traçons la courbe Rappel-Précision dans la Figure 4-17 pour une application de cette transformation sur 5 vues de chacun des 9 objets Minolta.

❖ Résultats

Le résultat montre clairement la forte stabilité du descripteur IndSHOT par rapport aux autres descripteurs (meilleure rappel et précision). Les courbes LSP, IndThrift et SHOT ont un comportement assez proche les uns des autres. Pareillement au cas d'ajout d'un bruit, le comportement du Spin est clairement sensible au lissage. Les mêmes raisons avancées pour la robustesse au bruit sont valables pour le lissage. La formulation donnée pour ces histogrammes paraît sensible à l'échantillonnage.

Les résultats présentés dans cette partie confirment ceux obtenus pour le bruit en invoquant la stabilité du descripteur IndSHOT grâce à son design de normalisation et son hybridité histogramme-signature pour décrire l'information locale.

- Le pouvoir descriptif et discriminant intervient pour accentuer l'invariance ou la robustesse des descripteurs. Une quantification de ce pouvoir descriptif est présentée dans la partie qui suit.

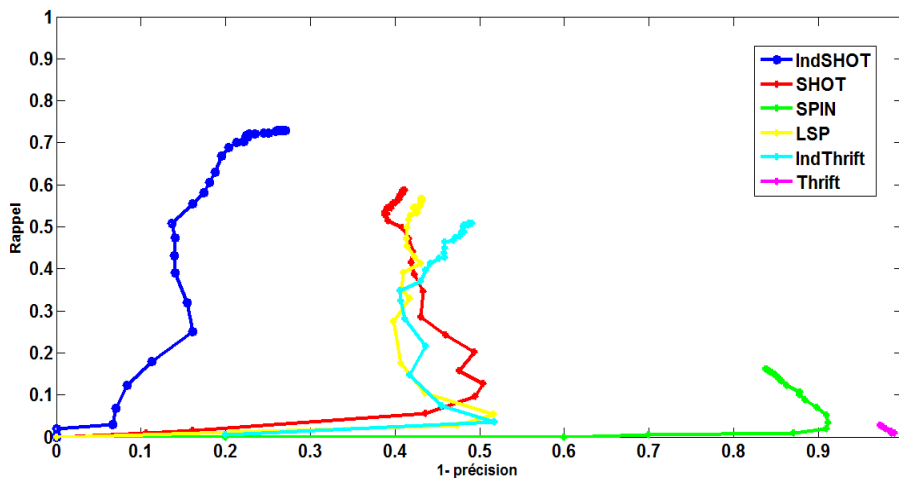
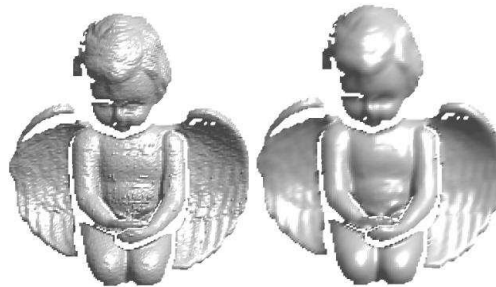


Figure 4-17 En haut-Exemple de lissage du modèle de l'ange -En bas- Courbe rappel précision pour un facteur de relaxation $f = 0.25 \times MR$

4.2.2.5 Pouvoir descriptif et stabilité pour un décalage de la position physique

❖ Pouvoir descriptif

- Quotient R_d :

Choksuriwong et al. (Choksuriwong, et al., 2005) se sont intéressés à l'évaluation de la performance de différents descripteurs invariants par transformations géométriques et ont proposé, pour quantifier le pouvoir descriptif et discriminant des descripteurs, un ratio de l'écart type des moyennes des valeurs

du descripteur sur les objets de la base sur la moyenne des écarts type du descripteur. Nous notons ce quotient R_d .

En effet, dans le but de reconnaître un objet qui apparaît sous différentes orientations ou échelles, le descripteur utilisé doit donner des valeurs similaires sous ces vues. Soit E_i le vecteur formé par les valeurs de l'écart type du descripteur j obtenu pour l'objet i . Afin d'avoir des descripteurs similaires pour le même objet, la valeur moyenne des E_i doit être minimisée.

D'un autre côté, pour pouvoir discriminer les objets entre eux, la dissimilarité entre les descripteurs doit être grande pour bien séparer ces objets. Soit M_j le vecteur formé par les valeurs de la moyenne du descripteur j obtenu pour un objet i . Pour augmenter le pouvoir discriminant du descripteur entre les objets, l'écart type σ_{M_j} des M_j , doit être maximisé. Par conséquent, le quotient $R_d = \frac{\sigma_{M_j}}{\bar{E}_j}$ présente un

critère qui caractérise la performance d'un descripteur.

Dans notre cas, les descripteurs sont sous forme d'histogrammes calculés sur des PIs. Nous proposons de calculer ce ratio de la façon suivante:

- Pour un objet i , et un point k sélectionné sur une vue x , nous cherchons les homologues de k sur les autres n vues de l'objet. Nous avons ainsi n descripteurs décrivant le même point physique k . Pour ces n histogrammes, et pour chaque bin (composante) de l'histogramme, nous calculons la moyenne M_i sur les valeurs des vues de l'objet pour obtenir un vecteur des moyennes des composantes. Puis, nous calculons, pour chaque case et pour toutes n vues, l'écart type par rapport à ce vecteur des moyennes. Nous formons l'écart type total e_i , pour l'objet i , comme la somme de ces écarts types.
- Pour chaque type de descripteur j , répéter ce traitement en prenant deux points par vue et faire la moyenne des écarts E_{ji} par objet. Ensuite, la moyenne sur tous les objets \bar{E}_j est déduite.
- Nous disposons pour chaque objet i d'un vecteur M_i des moyennes des composantes du descripteur calculé au même point physique sur ses n vues. Nous calculons, par la suite, pour chaque vecteur M_i , son écart σ_i avec l'ensemble des vecteurs M_m ($m \neq i$). Le numérateur du quotient σ_{M_j} correspond à la somme des écarts types σ_i sur tous les objets. Pour chacun des six descripteur, nous déduisons la valeur du ratio $R_d = \frac{\sigma_{M_j}}{\bar{E}_j}$. Le dispositif de ces étapes est

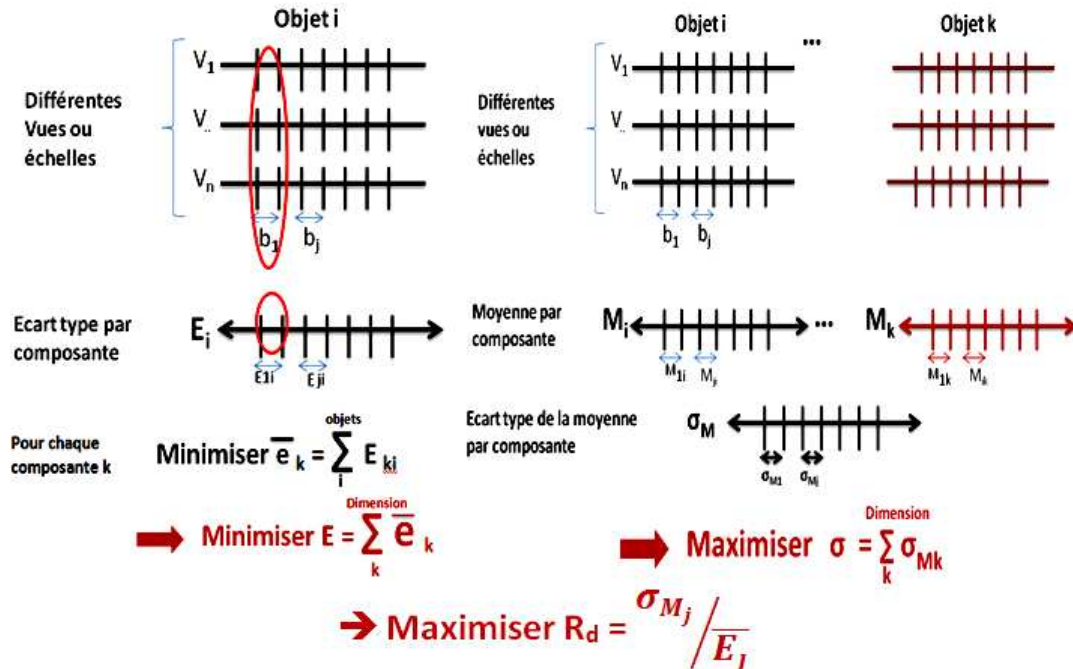
schématisé dans la Figure 4-18.

○ **Quotient C_{dis} :**

Nous introduisons un autre ratio, que nous appelons C_{dis} . Ce ratio calcule pour k points extraits, aux mêmes positions physiques sur une vue initiale et sur sa vue transformée par rotation, le rapport entre la moyenne des distances des descripteurs faussement appariés, et la moyenne des distances des descripteurs correctement appariés. Plus ce ratio est grand plus le pouvoir descriptif est grand.

Les résultats sont donnés dans le Tableau 4-2. D'après ces ratios, il est clair que le IndSHOT a le plus grand pouvoir descriptif. Le SHOT et le IndThrift se positionne à peu près au même niveau. Le ratio R_d du Thrift est élevé, cependant la faible valeur de C_{di} fait chuter son pouvoir descriptif.

Classés dans un ordre décroissant de performance, nous avons l'IndSHOT, le SHOT, le IndThrift. Ces résultats rejoignent les constatations données dans les paragraphes précédents pour conclure clairement sur la supériorité du descripteur IndSHOT.



	SHOT	IndSHOT	Spin	LSP	IndThrift	Thrift
Ratio C_{dis}	1.07	1.47	0.79	0.99	1.07	0.25
Ratio R_d	3.5	4.8	2.6	2.5	3.09	4.5

Tableau 4-2 Pouvoir descriptif des six descripteurs avec le ratio C_{dis} et le ratio R_d

❖ Stabilité pour un décalage de la position physique du point de calcul

L'étape de description des régions locale autour des PIs intervient après la phase de détection. La répétabilité de ces détecteurs est primordiale pour réussir la description et le processus de mise en correspondance. Néanmoins, cette répétabilité reste approximative puisque nous savons, qu'une erreur de décalage dans les positions spatiales des PIs extraits par les détecteurs peut subvenir d'une vue à une autre. Nous proposons dans ce paragraphe, d'étudier le comportement des différents descripteurs pour décrire la même zone décalée avec une marge ϵ croissante. Pour ce faire, nous sélectionnons un point sur une vue de l'objet d'Ange. Puis, nous cherchons les k plus proche voisins de ce point que nous ordonnons dans un ordre croissant selon leur distance du point référence. Nous calculons pour chacun de ces points les six descripteurs sur la zone qui les entoure avec le même rayon du support.

Nous illustrons dans la Figure 4-19, la distance entre le descripteur du point référence de calcul et les descripteurs de ses k (~ 200) plus proche voisins. L'axe des x représente la distance physique du point référence au $k^{\text{ème}}$ plus proche voisin et qui correspond au décalage.

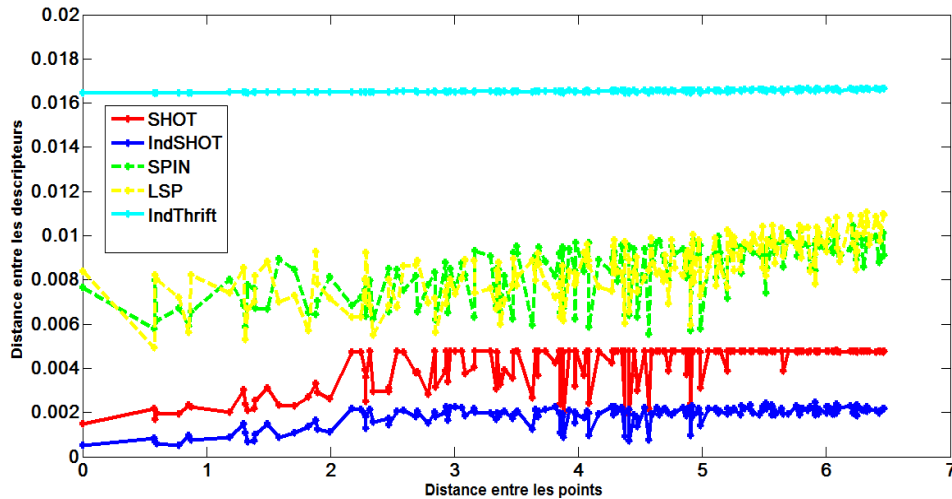


Figure 4-19 Evolution de la distance entre le descripteur d'un point référence et les descripteurs de ses k plus proches voisins (en fonction de la distance physique du $k^{\text{ème}}$ plus proche voisin)

Le IndThrift paraît le plus stable des descripteurs pour un décalage de la position du calcul. La courbe des distances pour le IndSHOT évolue avec une faible pente et reste relativement stable à partir du 20^{ème} plus proche voisin. La pente et l'amplitude de la courbe du SHOT est légèrement plus grande que celle du IndSHOT. Quant aux courbes du spin et du LSP, elles oscillent avec une amplitude plus importante exprimant une faible stabilité de ces deux derniers. D'un autre côté, la distance calculée entre les descripteurs est largement plus faible pour le IndSHOT.

- Notre conclusion sur cette partie, est que les nouveaux descripteurs proposés, IndSHOT et IndThrift, ont montré une performance qui a amélioré le pouvoir descriptif et la stabilité des descripteurs de l'état de l'art desquels nous nous sommes inspirés.

4.2.2.6 Temps de calcul et outils de développement

Dans notre travail, les algorithmes de description Spin, Thrift, IndThrift et LSP ont été implémentés. Pour le SHOT, l'implémentation originale est utilisée. L'algorithme original du CSHOT a été modifié pour l'adapter au descripteur IndSHOT. Le temps de d'exécution présenté dans cette partie n'inclue pas le temps de calcul des quantités de l'indice de forme qui est donné en sortie du détecteur.

Pour automatiser les calculs, nous avons créé une partie commune pour tous les descripteurs qui lit les données, construit le maillage, extrait les points d'intérêt (PIs) et lance la description. Le code implémenté dans le SHOT introduit un parallélisme dans le traitement des descripteurs en lançant des tâches parallèles qui traitent la description des PIs par tranche de taille (nombre des PIs/nombre de tâche). Nous utilisons ce même parallélisme pour tous les algorithmes de descriptions. Les résultats du temps de calcul sont donnés dans le Tableau 4-3. Pour avoir une idée sur la totalité du temps nécessaire pour notre système et sa comparaison avec une approche existante, nous présentons, dans le

Tableau 4-4, le temps de calcul des phases de détection et de description pour le couple (SC_HK_C, IndSHOT) et pour l'approche du SURF.

La bibliothèque graphique VTK est utilisée pour les traitements appliqués au maillage. Les filtres suivants ont été utilisés pour la partie précédente de l'évaluation des descripteurs :

vtkPolyDataNormal est un filtre qui calcule les vecteurs normaux pour un maillage polygonal. Les normales correspondent aux normales des facettes, et la normale d'un sommet est déduite de la moyenne des normales des facettes qui lui sont adjacentes.

vtkSmoothPolydataFilter est un filtre qui lisse un objet de type vtkPolyData en utilisant une méthode de type Laplacienne. Les coordonnées des sommets du maillage sont remplacées par la moyenne des coordonnées des sommets qui lui sont connexes (qui lui sont directement liés par une arête). Un facteur de relaxation contrôle le degré des remplacements des sommets. Ce processus est répété selon le nombre d'itérations donné en paramètre.

vtkDecimatePro est un filtre de décimation qui réduit le nombre de polygones d'un maillage en approximant au mieux la géométrie de la forme originale. L'algorithme procède de la manière suivante : chaque sommet du maillage est inséré dans une file de priorité. Certains points qui ne peuvent pas être supprimés ou triangulés ne sont pas traités. Après chaque suppression d'un point de la file, lui succède une triangulation des trous. Le processus se répète jusqu'à ce que la file soit vide. Le maillage avec les points qui restent est divisé en des pièces séparées selon une liaison non-manifold et les points sont réinsérés dans la file à nouveau. Puis, le même traitement est répété récursivement jusqu'à atteindre le pourcentage de réduction donné en paramètre.

❖ Résultats

Base: Nombre de Pls	IndSHOT	SHOT	IndThrift	LSP	Spin
Stuttgart : 75 points	0.56	0.57	2.9	0.54	0.52
Carotte : 72 points	0.42	0.41	2.15	0.42	0.42

Tableau 4-3 Temps de calcul (en seconde) des cinq descripteurs

Base : densité de points	Détection+description	
	Surf	SC_HK_C + IndSHOT
Stuttgart : 10948 points	15.20	18.64
Carotte : 4685 points	10.13	9.62

Tableau 4-4 Temps de calcul (en seconde) des phases de détection et de description pour le couple (SC_HK_C, IndSHOT) et pour l'approche du SURF

D'après ces tableaux :

- Le temps d'exécution des descripteurs IndSHOT, SHOT, LSP et Spin est de même ordre.
- Le temps de calcul du descripteur IndThrift est important (allant jusqu'à 7 fois le temps du IndSHOT). L'explication est que ce descripteur se base sur le calcul des normales du voisinage sur deux rayons (le premier w_{Large} est égal au rayon du support et le deuxième w_{Small} est égal à la moitié du rayon du support). Les opérations de parcours de tout les points du support du point référence et le recalcule pour chacun de deux nouvelles normales résultantes des valeurs des normales des points des deux supports (W_{Large} et W_{Small}), sont coûteuses.
- La légère supériorité dans le temps de calcul de la combinaison SC_HK_C -IndSHOT par rapport au détecteur-descripteur SURF, revient au temps mis par le détecteur pour calculer les quantités différentielles : courbure et normale. Cette étape reste à optimiser.

4.3. Conclusions

Nous avons implémenté et testé six descripteurs basés sur les histogrammes cumulant une information calculée essentiellement à base de géométrie différentielle. Le pouvoir descriptif, l'invariance aux transformations géométriques et la robustesse au bruit et à l'échantillonnage ont été évaluées et les résultats ont montré :

- Stabilité des descripteurs IndSHOT et IndThrift pour une variation d'angles de vue
- Supériorité nette du descripteur IndSHOT pour variation d'échelles
- Robustesse au bruit des deux descripteurs IndSHOT et IndThrift
- Forte stabilité du IndSHOT pour une transformation de lissage des données
- Pouvoir descriptif élevé du IndSHOT et une forte stabilité du IndThrift suite à un décalage de la position physique
- la supériorité du nouveau descripteur IndThrift par rapport à la version du Thrift. La forme 2D de ce nouveau descripteur en ajoutant l'information de l'indice de forme a haussé le pouvoir descriptif et a renforcé la stabilité et la robustesse du descripteur.
- La supériorité du nouveau descripteur proposé IndSHOT au descripteur d'origine SHOT. En effet, une nette amélioration est réalisée par ce nouveau descripteur qui combine deux informations de la forme locale (indice de forme et cosinus des normales) par rapport à une seule information (cosinus des normales) contenue dans le SHOT.
- Nous avons pu constater l'importance de l'utilisation d'une base locale de référence dans l'expression des descripteurs à travers la comparaison effectuée avec le descripteur proposé IndThrift n'incluant pas de base de référence dans sa formulation. L'étude comparative a confirmé la performance supérieure du IndSHOT utilisant une RF.
- L'évaluation quantitative a montré que le descripteur LSP est moins robuste et moins descriptif en comparaison avec les descripteurs SHOT et IndSHOT et IndThrift. Bien que ce descripteur se base sur la définition d'une RA pour l'expression des cosinus des normales, cette RA n'est pas calculée d'une façon robuste. En effet, la normale référence du PI correspond à la normale du point obtenue par une simple moyenne sur les normales des sommets voisins à la facette. Alors que la RF du SHOT et IndSHOT est obtenue par le calcul des vecteurs propres d'une matrice de covariance pondérée par les distances des points appartenant au support. Ce processus a permis d'augmenter la robustesse et la répétabilité de ces deux derniers. D'un autre côté, l'utilisation des cosinus des normales calculées sur deux

supports de taille différentes (IndThrift) paraît plus performant qu'une description par le cosinus entre la normale du point référence et les normales de ses voisins (LSP).

- Un temps de calcul quasi réel pour l'étape de description du IndSHOT contrairement à celui du IndThrift.

Dans le cadre pratique d'une application de reconnaissance d'objets, le chapitre suivant expose une comparaison de la performance des descripteurs présentés en association avec les détecteurs les plus stables et robustes du chapitre précédents. L'évaluation du taux de reconnaissance sur différentes bases de données synthétiques et réelles démontrera la supériorité de la combinaison des détecteurs et descripteurs proposés.