

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



Université Djillali Liabès de Sidi Bel Abbès
Faculté de Génie Électrique
Département d'Électronique



Thèse présentée par :
BELHALLOUCHE Lakhdar
Pour l'obtention du diplôme de :
Doctorat en Sciences
Spécialité : Electronique
Option : Signal, Image, Parole et Télécommunications

Intitulé de la thèse
Indexation des Champs de Lumière (Light Field)
Dans le Domaine
Compressé par la Transformée en Ondelettes
Adaptée à la Forme

Présentée devant le jury composé de :

Mr BOUKELIF Aoued	Pr (U.D.L. SBA) Président
Mr BELLOULATA Kamel	Pr (U.D.L. SBA) Directeur de thèse
Mr BESSAID Abdelhafid	Pr (Université de Tlemcen) Examineur
Mr BELAID Ahror	MCA (Université de Béjaia) Examineur
Mr KPALMA Kidiyo	Pr (INSA de Rennes, France) Invité

Petit oiseau, si tu n'as pas d'aile

Tu ne peux pas voler

Petit oiseau, si tu n'as pas d'aile

Tu peux marcher

*À mes parents
À mes frères
À ma sœur
À ma grand mère
À ma fille
pour leur amour
et leur soutien illimité*

Remerciements

Je remercie les membres de ma famille, pour m'avoir soutenu et encouragé, durant toutes ces années de préparation de ce travail.

Je remercie mon directeur de thèse monsieur le professeur BELLOULATA Kamel, d'avoir dirigé ce travail, de m'avoir initié dans ce domaine de recherche et d'être si patient durant toutes ces années.

Je remercie monsieur le professeur BOUKELIF Aoued, pour m'avoir fait l'honneur de présider mon jury de thèse.

Je remercie messieurs les professeurs BESSAID Abdelhafid et BELAID Ahror, pour avoir accepté de rapporter sur mon travail de thèse.

Toute ma gratitude et ma profonde reconnaissance à monsieur le professeur Kidiyo KPALMA, pour son aide et son étroite collaboration durant les stages passés à l'INSA de Rennes. Merci également, aux membres du laboratoire IETR de l'INSA de Rennes, merci au professeur Joseph ROSIN, à Meriem, à Alexandre et sa femme ainsi qu'à Pascal.

Résumé

Vu la croissance vertigineuse des bases de modèles 3D, ainsi que leurs utilisation, beaucoup de chercheurs se sont intéressé de près à la recherche dans ce genre de base. La recherche dans les bases de modèles 3D consiste à comparer les caractéristiques visuelles de l'objet requête (exemple ou mot clé), aux caractéristiques des modèles de la base. Ces caractéristiques sont souvent organisées sous formes d'histogrammes. Donc la recherche d'images 3D dans une base revient à extraire les caractéristiques visuelles des modèles 3D, de les organiser sous forme d'histogrammes et de les comparer un à un. Étant donné que la plupart des modèles 3D sont stockés actuellement dans leur formats compressés, nous nous sommes intéressé dans ce travail à la problématique de la construction d'indices des modèles 3D à partir des coefficients des transformées de l'image. Dans un contexte très particulier nous nous sommes penchés sur les méthodes d'indexation des light field.

Ce travail comprends quatre étapes : un prétraitement du modèle 3D permettant son alignement et la définition de son repère tridimensionnel, une deuxième étape concerne l'extraction de vues 2D constituant le light field, lors de la troisième étape les caractéristiques des vues 2D sont extraites et sont organisées sous formes d'histogrammes. Enfin nous utilisons ces histogrammes pour évaluer la similitude entre le modèle 3D requête et ceux dans la base de modèles.

Table des matières

Remerciements	4
1 Introduction	1
1.1 Introduction	1
1.1.1 Contexte	1
1.1.1.1 Architecture du système d'indexation et de recherche des light field	4
1.1.2 Motivations	7
1.1.3 Objectifs	7
1.2 Organisation du manuscrit	8
2 État de L'art Des méthodes d'indexation	10
2.1 Introduction	11
2.2 Etat de l'art des méthodes d'indexation des modèles 3D	12
2.2.1 Méthodes d'indexation basées sur les caractéristiques 3D . . .	13
2.2.1.1 Méthodes d'indexation basées caractéristiques globales	13
2.2.1.2 Méthodes d'indexation basées caractéristiques locales	16
2.2.1.3 Méthodes d'indexation basées sur le partitionnement de l'espace	17
2.2.2 Méthodes d'indexation 3D basées graphes	17
2.2.2.1 Méthodes d'indexation par squelettes	17
2.2.2.2 Méthodes d'indexation par les graphes de Reeb . . .	19
2.2.3 Méthodes d'indexation basées vues 2D	19
2.2.4 Méthodes d'indexation hybrides	21
2.3 Méthodologie des systèmes d'indexation et de recherche d'images ba- sés régions	22
2.3.1 Extraction des régions de l'image	23
2.3.2 Description sémantique des régions de l'image	25
2.4 Indexation des light-field	26
2.5 Principe de notre contribution dans le domaine d'indexation des vues 2D	27

2.6	Conclusion	28
3	Indexation des Vues/requêtes 2D Par la Transformée en Ondelettes Basée Région	29
3.1	Introduction	30
3.2	Schéma global	31
3.2.1	Extraction de la région d'intérêt	31
3.2.2	Transformée en ondelettes basée région	31
3.2.2.1	Étendue des bordures de segments	37
3.2.2.2	Convolution des segments avec le banc de filtres	37
3.2.2.3	Sous échantillonnage	39
3.3	Caractérisation de l'image par la transformée en ondelettes basée région	40
3.3.1	Caractérisation des attributs	41
3.3.2	Regroupement des vecteurs de caractéristiques	41
3.3.2.1	Construction des patterns de texture	41
3.3.2.2	Construction de l'histogramme de texture	42
3.3.2.3	Construction des patterns de couleur	45
3.3.2.4	Construction de l'histogramme de couleur	46
3.3.2.5	Construction des patterns de contour	47
3.3.2.6	Construction de l'histogramme de contour	48
3.4	Conclusion	51
4	Prétraitement Des Modèles Tridimensionnels	52
4.1	Introduction	53
4.2	Schéma du processus de normalisation et d'alignement des objets 3D	53
4.3	État de l'art des méthodes de normalisation	54
4.4	Normalisation des light field	58
4.4.1	Calcul du centre de l'objet	59
4.4.2	Construction de la Sphère minimale englobante	60
4.4.3	Centrage de l'objet 3D	61
4.4.4	Mise à l'échelle de l'objet 3D	62
4.5	Méthodes d'alignement	63
4.6	Alignement des light field	69
4.7	Conclusion	70
5	Indexation Des Light Field Par La Transformée en Ondelettes Basée Région	72
5.1	Introduction	73
5.2	Indexation des light-field	73
5.2.1	Light field	74

5.2.2	Extraction des vues 2D	86
5.2.3	Caractérisation des light field	88
5.2.3.1	Caractérisation de la couleur	90
5.2.3.2	Caractérisation du contour	93
5.2.3.3	Caractérisation de la texture	96
5.3	Conclusion	101
6	Recherche Des Light Field	103
6.1	Introduction	104
6.2	Stratégie de recherche des light field	105
6.3	Visualisation des résultats	109
6.4	Processus de comparaison	109
6.4.1	Conventions	110
6.4.2	Normes de Minkowski	110
6.4.3	Distance de Bhattacharyya	110
6.4.4	Distance de Earth Mover	111
6.4.5	Divergence du χ^2	112
6.4.6	L'intersection d'histogrammes	112
6.4.7	Distance de Hausdorff	113
6.5	Comparaison des light field	113
6.5.1	Comparaison des vues des light field	115
6.6	Conclusion	118
7	Simulations et résultats	119
7.1	Introduction	120
7.2	Bases de données	120
7.2.1	La Base d'images Corel	120
7.2.2	La base Princeton Shape Benchmark (PSB)	121
7.2.3	La base NTU 3D modèles benchmark	122
7.3	Paramètres d'évaluation	122
7.4	Expérimentations sur l'indexation et la recherche d'images 2D basée région	124
7.5	Protocole de recherche des vues 2D	124
7.5.1	Niveau de décomposition de la transformée SA-DWT	125
7.5.2	Taille des histogrammes	125
7.5.3	Calcul des paramètres de pondération α et β	128
7.5.4	Comparaison de la méthode proposée aux méthodes d'indexa- tion basées région	128
7.6	Expérimentations sur l'indexation et la recherche des light field	131
7.6.1	Détermination du paramètre de pondération ω	133

7.6.2	Expérimentation sur l'alignement	134
7.6.3	Expérimentation sur le nombre de vues	135
7.6.4	Expérimentations sur le processus de la recherche	136
7.6.5	Accélération et robustesse du processus de comparaison	140
7.6.5.1	Temps de comparaison des objets 3D	140
7.6.5.2	Robustesse aux déformations	141
7.6.6	Étude comparative	142
7.6.6.1	Éexpérimentations sur la base d'objets 3D DB3	143
7.6.6.2	Éexpérimentations sur la base d'objets 3D DB4	143
7.6.6.3	Comparaison de la robustesse vis-à-vis les déforma- tions des objets	146
7.6.6.4	Temps de comparaison des objets 3D	146
7.7	Conclusion	147
Annexe 1		151
Annexe 2		152
Annexe 3		153
Bibliographie		153

Table des figures

1.1	Exemple de synthèse d'une scène 3D à partir de formes géométriques	2
1.2	Exemple d'un système d'acquisition des light field.	3
1.3	<i>Exemples d'objets synthétisés à l'aide de la technique des light field.</i> . .	4
1.4	Schéma général du système d'indexation et de recherche des light field	5
2.1	Illustration du problème de l'occlusion des objets 3D.	12
2.2	<i>Classification des méthodes d'indexation de modèles 3D</i>	13
2.3	<i>Distribution de formes de cinq tanks (en gris) et de six voitures (en noir). Image extraite de [93].</i>	16
2.4	Exemples de caractéristiques des faces de l'objet 3D	17
2.5	<i>Exemple de squelette de forme 3D [133]</i>	18
2.6	<i>Exemples de graphes de Reeb de différents modèles 3D[132]</i>	19
2.7	<i>Illustration de la méthode de division de l'image en blocs et la sélection du bloc dominant.</i>	24
2.8	<i>Illustration d'une méthode de la segmentation de l'image à l'aide de l'attribut couleur.</i>	24
3.1	Principe de notre méthode d'indexation d'images basée région	32
3.2	Exemple d'extraction de la région d'intérêt	33
3.3	<i>Comparaison de différentes techniques de la transformation basée région en terme de PSNR [76].</i>	35
3.4	Comparaison entre les différentes méthodes de transformation basée région	36
3.5	Exemple d'extension d'un segment du signal x	38
3.6	Exemple d'extension d'un segment de courte longueur	38
3.7	Exemple de l'application de la transformée en ondelettes basée région	40
3.8	Exemples de la construction des patterns de texture	43
3.9	Exemples d'histogramme de texture	45
3.10	Construction des patterns de couleur	46
3.11	Exemples d'histogrammes de couleur	48
3.12	Exemple de la construction d'un pattern de contour	49

3.13	<i>Exemples d'histogrammes de contour. (a), (b) et (c) représentent les histogrammes des régions d'intérêt des objets. (d), (e) et (f) représentent les histogrammes de contour du fond de l'image.</i>	50
4.1	Schéma général du processus de normalisation et d'alignement des objets 3D	54
4.2	Exemple de l'utilisation des centres de gravité pour le centrage	55
4.3	Illustration de l'effet négatif du sous échantillonnage sur la capture du contour de l'objet	59
4.4	Illustration de la projection d'un objet 3D sur deux formes différentes	59
4.5	Exemple d'objets 3D maillés	60
4.6	Exemple de la construction de la sphère minimale	61
4.7	Influence de la normalisation sur la sémantique de l'objet	63
4.8	Représentation d'un nuage de points	65
4.9	Illustration de la construction des axes principaux	66
4.10	Résultat de l'alignement d'un ensemble d'objets à l'aide de notre algorithme	71
5.1	<i>Schéma général de l'indexation des light field par la transformée en ondelettes basée régions.</i>	75
5.2	Représentation de la fonction plénoptique	76
5.3	Le rendu des scènes 3D à partir de nouvelles points de vue	77
5.4	Système de capture basé mosaïques	78
5.5	Le rendu des scènes 3D à partir de nouvelles points de vue	79
5.6	Illustration de la capture des light field	80
5.7	<i>plan de capture (μ, ν) [116]</i>	83
5.8	<i>plan (s, t) light fields [116]</i>	83
5.9	Illustration du simulateur d'acquisition des light field	85
5.10	<i>Illustration des effets trompeurs qui peuvent être produits par les modèles 3D</i>	86
5.11	<i>Répartition des caméras sur la sphère de capture des vues 2D</i>	88
5.12	<i>Résultat de la projection 2D d'un objet 3D.</i>	89
5.13	Exemples d'histogrammes de couleur de la composante Y des vues 2D d'objets 3D	91
5.14	Illustration de l'incapacité de l'attribut couleur à caractériser des objets de la même classe et de différentes couleurs	92
5.15	Caractérisation partielle (seulement de deux projections 2D) de deux objets de classes différentes	94
5.16	Exemples d'histogrammes de contour de la composante Y des vues 2D d'objets 3D	95

5.17	Caractérisation partielle (seulement de deux projections 2D) de deux objets de la classe avion à l'aide de l'attribut contour	97
5.18	Caractérisation partielle (seulement de deux projections 2D) de deux objets de classes différentes	98
5.19	Illustration de l'effet de l'occlusion sur l'indice du contour	99
5.20	Caractérisation partielle (seulement de deux projections 2D) de deux objets de la classe avion à l'aide de l'attribut texture	100
5.21	Caractérisation de l'attribut texture	101
6.1	Exemple de l'arbre de catégorisation et du sous ensemble d'objets représentatifs	106
6.2	Schéma général de la recherche des light field	108
6.3	Illustration du système de comparaison des light field	114
6.4	Comparaison de deux objets 3D	115
7.1	<i>Echantillons d'images de la base de Corel</i>	121
7.2	Echantillons de modèles 3D de la base PSB	122
7.3	Échantillons de modèles 3D de la base NTU	123
7.4	Évaluations des performances du système d'indexation d'images pour différents nombre d'itération de la SA-DWT.	126
7.5	Évaluations des performances du système d'indexation d'images par rapport au nombre de bins de l'histogramme de texture.	126
7.6	Évaluations des performances du système d'indexation d'images par rapport au nombre de bins de l'histogramme de couleur.	127
7.7	Évaluations des performances du système d'indexation d'images par rapport au nombre de bins de l'histogramme de contour.	127
7.8	Définition des valeurs des paramètres de pondération.	128
7.9	<i>Comparaison de notre méthode d'indexation basée région à la méthode Simplicity à l'aide du paramètre Précision moyenne.</i>	129
7.10	<i>Comparaison de notre méthode d'indexation basée région à la méthode de WALRUS en terme des paramètres de précision et de rappel.</i>	130
7.11	<i>Comparaison de la méthode proposée à la méthode basée sur les curvelettes en terme de précision.</i>	131
7.12	<i>Comparaison de notre méthode d'indexation basée région aux méthodes d'indexation basées régions en terme du paramètre précision.</i> . .	132
7.13	<i>Détermination de la valeur de pondération ω</i>	134
7.14	<i>Comparaison des performances de la méthode proposée avec et sans ACP et normalisation.</i>	135
7.15	<i>Évaluation de la méthode proposée pour différentes valeurs de l'angle θ</i> 136	

7.16	<i>Comparaison des performances de notre méthode au premier et deuxième niveau de recherche, à l'aide de la courbe Rappel\Précision.</i>	137
7.17	Résultat du processus de la recherche.	138
7.18	Comparaison des performances de l'algorithme de recherche aux deux niveaux de comparaison	139
7.19	Impact de l'alignement et de la normalisation des objets sur la durée de la phase de comparaison	140
7.20	Impact temporel de la hiérarchisation sur la phase de comparaison . .	141
7.21	Comparaison des performances de notre méthode à la méthode LFD .	143
7.22	Comparaison des performances de notre méthode par rapport à la méthode HS	144
7.23	Comparaison des méthodes LFD, HS et la méthode proposée à l'aide de la courbe précision-rappel	145
7.24	Comparaison temporelle entre la méthode proposée, LFD et HS . . .	147

Liste des tableaux

7.1	Variation des distances entre les objets suite à leur déformations . . .	142
7.2	Paramètres optimaux de l'algorithme d'indexation et de recherche des light field	142
7.3	Comparaison de la robustesse de la méthode proposée aux méthodes LFD et HS	146

Chapitre 1

Introduction

1.1 Introduction

1.1.1 Contexte

Depuis plusieurs années, un intérêt particulier est porté aux objets 3D. La communauté scientifique préfère manipuler et visualiser des modèles 3D à la place des images 2D. Ceci est suite au développement de scanner et d'outils d'acquisition sophistiqués. En fait, grâce à ce type d'images il est devenu possible de plonger dans un environnement virtuel 3D proche de la réalité, on peut même synthétiser et créer des scènes 3D de caractéristiques de luminance, d'effet de l'ombre et de contraste, qui se rapprochent de la réalité.

Grâce aux nouvelles technologies de modélisation et de synthèse des scènes tridimensionnelles, on dispose actuellement d'un nombre impressionnant de modèles 3D. Trouver des solutions pour un stockage efficace et l'optimisation de l'archivage des objets 3D, constitue un challenge incontournable. Les techniques d'indexation et de recherche de l'image tentent au mieux de caractériser l'aspect visuel du contenu des modèles 3D, pour générer des descriptions plus ou moins précises. En effet, l'organisation des modèles 3D dans des bases d'images a rendu possible l'accès rapide et la récupération efficace de ces modèles par l'utilisateur.

Il est ainsi possible de faire vivre des expériences presque réalistes au simple utilisateur des nouvelles technologies. Ceci a incité les plus grandes salles de cinéma, les spécialistes de jeux, de l'industrie automobile et de l'imagerie médicale, à l'utilisation de ce type d'images, pour une meilleure visualisation de leurs produits. Grâce à l'utilisation de modèles tridimensionnels, ces acteurs socio-économiques peuvent également faire une meilleure promotion de leurs produits. En quelque sorte les objets 3D sont devenus plus attractifs et plus conviviaux comme outil de simulation.

C'est dans cet état d'esprit que sont apparues les premières méthodes d'acquisition et de rendu de modèles 3D. Ces méthodes consistent à convertir les données acquises de la scène 3D en données numériques, directement visualisables sur un dispositif de reproduction graphique tel un écran par exemple.

Le rendu des modèles 3D peut être réalisé de deux manières différentes :

- Une première famille d'approches qui sont basées sur la modélisation mathématique de la géométrie des objets 3D. Dans ce cas le rendu se réduit à la conversion des modèles mathématiques en modèles graphiques 3D[120, 88, 121]. Ces techniques sont plus adaptées au rendu des formes géométriques simples, tels qu'une sphère ou un triangle. Mais elles ne peuvent pas traiter des modèles géométriques complexes et moins réguliers. Par exemple il est très difficile de modéliser mathématiquement les feuilles d'un arbre ou les poils d'un chat. Un autre inconvénient provient de la complexité des calculs qui sont employés lors de la phase du rendu (le rendu de chaque pixel est l'intersection d'un rayon lumineux avec l'objet à visualiser). La lumière et les effets de l'ombre ne sont pas bien simulés dans les scènes synthétisées. A cet effet, certains chercheurs ont proposé de coller la texture de l'objet sur son modèle géométrique[67]. Autrement dit, la description géométrique ne suffit pas à elle seule pour modéliser tous les aspects d'un objet 3D[121]. La figure 1.1 montre une scène synthétisée à l'aide de cette technique, les formes de cette scène ne correspondent pas vraiment au monde réel : les angles sont très droits, l'échange de la lumière entre les formes est très différent de la réalité.



FIGURE 1.1 – Exemple de synthèse d'une scène 3D à partir de formes géométriques, image extraite de [45]

- La deuxième approche consiste à estimer la luminance et l'ombre des objets de la scène 3D à partir de leur vues 2D, en utilisant un système d'acquisition de plusieurs caméras. Le rendu de la scène 3D est donc calculé à partir de ces

photographies sans utiliser un modèle géométrique. Cette catégorie regroupe différentes méthodes, les premières méthodes proposées n'utilisent qu'une seule image 2D. D'autres utilisent plusieurs vues pour augmenter le degré du photoréalisme du rendu et pour mieux estimer la profondeur et les champs de la lumière à l'intérieur de la scène[120, 87, 121]. Enfin, plus récemment certains chercheurs proposent d'augmenter le nombre de caméras à une centaine de caméras, organisées toutes en un seul dispositif d'acquisition. Le dispositif le plus connu est celui de l'université de Princeton montré sur la figure 1.2. Cette dernière approche communément appelée light field permet une parfaite reconstruction des objets 3D. Etant donné qu'elle est basée sur des photos réelles et sur un emplacement bien choisi des caméras, cette technique permet une parfaite estimation et simulation des effets de la lumière. Dans notre travail, deux cas de figure se présentent. 1) dans le cas de la disponibilité des vues 2D des light field, nous nous contentons de la sélection des vues 2D nécessaires à l'opération de la caractérisation du light field. 2) Dans le cas des bases de modèles 3D, nous projetons chaque modèle dans un environnement simulateur du système d'acquisition des light field. Nous projetons ensuite le modèle sur certains angles de vues, afin d'extraire les vues 2D nécessaires à la caractérisation du light field[21, 20].



FIGURE 1.2 – Exemple d'un système d'acquisition des light field. (a) représente un mur du système d'acquisition des light field. (b) représente le système d'acquisition des light field constitué de plusieurs murs de caméras, image extraite de [45]

La figure 1.3 montre des exemples d'objets 3D de la base [1], qui ont été synthétisés à l'aide des light field. Contrairement aux méthodes multi-vues ou celles basées sur les modèles géométriques, ces objets sont très proches du monde réel. Cependant les prétraitements employés pour estimer les champs de lumière restent d'une très grande complexité de calcul, d'autre part la mémoire d'acquisition des light field est très volumineuse, elle peut facilement dépasser les 1,6 GB[89, 108, 31, 36]. On

compte ainsi des milliers de modèles 3D créés et utilisés dans le domaine de l'industrie, du loisir et du médical. Les spécialistes se retrouvent ainsi obligés de créer des bases propres à chaque domaine.



FIGURE 1.3 – *Exemples d'objets synthétisés à l'aide de la technique des light field.*

L'expansion rapide des bases de modèles 3D et leur utilisation dans différentes applications, va de pair avec le problème de leur exploitations. De plus en plus de techniques d'indexation et de recherche de modèles 3D ont été proposées, pour faciliter l'accès aux bases de données et optimiser les résultats de recherche[93]. Le schéma d'un système d'indexation comprend généralement deux phases : Une phase d'indexation des objets 3D hors ligne. Une deuxième phase de recherche d'objets similaires à l'objet requête en ligne. Le principe est de comparer l'image requête aux images de la base à l'aide d'une mesure de similarité entre les descripteurs de ces modèles et le descripteur de la requête. A cet effet, l'image requête et les images de la base doivent toutes être décrites par des descripteurs de leurs contenus.

1.1.1.1 Architecture du système d'indexation et de recherche des light field

Comme il est illustré sur la figure 1.4, la première phase de ce processus consiste à normaliser les modèles 3D. Le changement des conditions d'acquisition (La position de l'objet dans la scène, sa rotation, l'échelle de son repère et la lumière de l'environnement), peut influencer sur le résultat de la description de l'objet. L'indexation pourra ainsi devenir dépendante de ces conditions. La solution à ce problème consiste à normaliser et aligner les objets avant de les indexer[80, 40].

L'objectif de l'indexation de l'objet est de définir son espace de caractéristiques. Ces attributs doivent être discriminatifs et communs à tous les objets de la même classe. On peut distinguer par exemple la classe ballon par la forme, la classe dinosaures peut être distinguée des autres classes par la texture. Les attri-

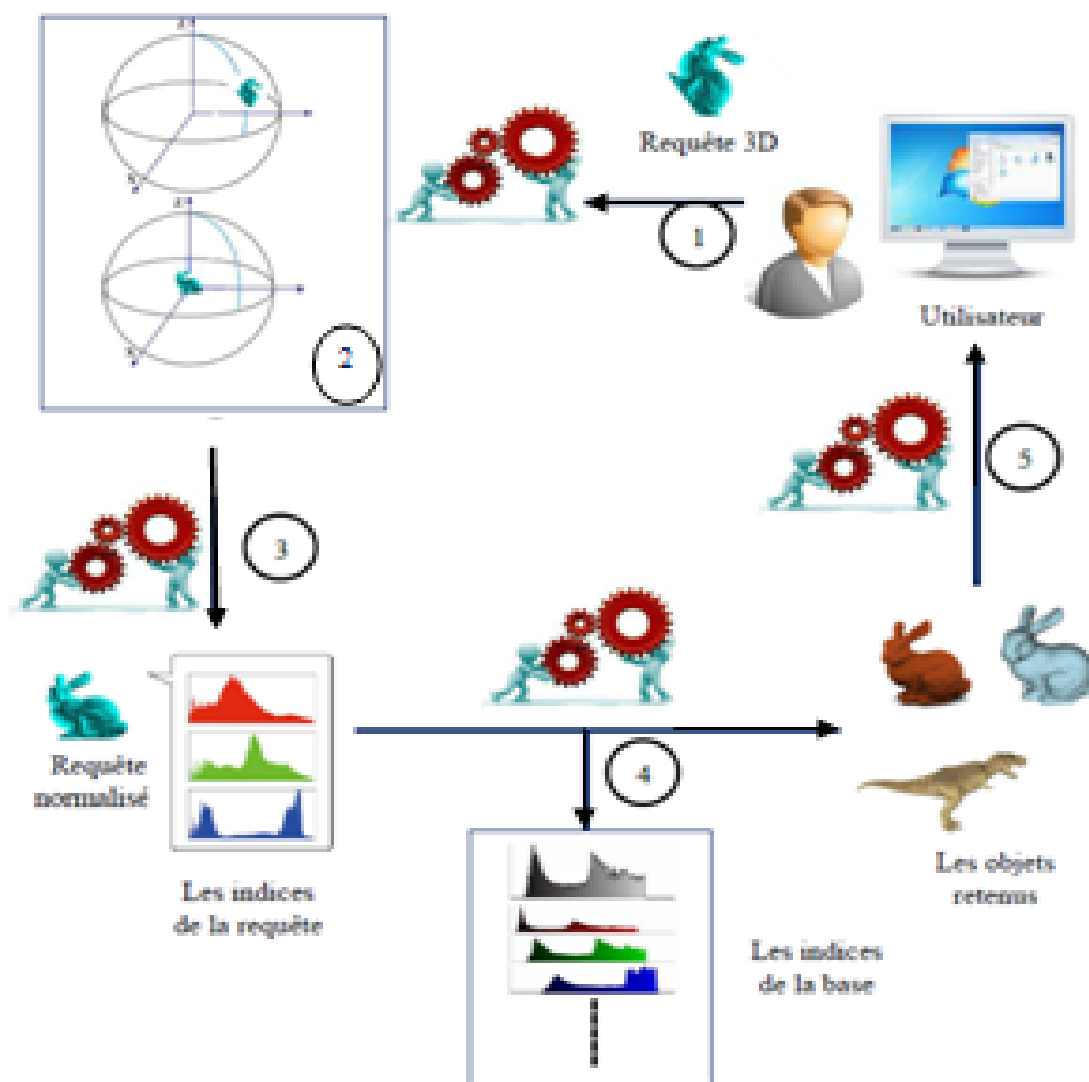


FIGURE 1.4 – Schéma général du système d'indexation et de recherche des light field. (1) représente l'initiation de la recherche par une requête. (2) représente la phase de la normalisation. (3) représente la phase de l'extraction des descripteurs de l'objet 3D. (4) représente la phase de comparaison des indices de la requête aux indices de la base. (5) représente la phase de la visualisation des objets pertinents à la requête.

but, couleur, forme et texture peuvent être extraits directement de l'image, ceci dit que l'indexation de l'objet dans ce cas est basée pixel [122]. Ce type d'indexation nécessite souvent des transformations inverses et des décodages de l'objet, qui est stocké dans son format compressé. La deuxième famille des approches de l'indexation propose la construction des indices à partir des coefficients des transformées de l'objet, sans passer par des transformations inverses ou des décodages. On peut citer par exemple l'indexation par la transformée en ondelettes, par la transformée

fractale...etc[40, 19]. L'indexation consiste à calculer l'occurrence ou la probabilité d'apparition de chaque valeur de l'attribut. Définissons par exemple l'espace de couleur C , dont les valeurs sont issues de l'ensemble $\{C_i\}_{i=1\dots N}$ (N représente le nombre de niveaux de couleur), N représente le nombre de niveaux de couleur possibles. L'indexation dans ce cas consiste à calculer l'occurrence (ou la répétition) de chaque valeur C_i de l'attribut couleur C .

L'approche usuelle pour rechercher l'image requête, consiste à comparer les descripteurs de la base d'images au descripteur de l'image requête[80]. Cette phase comprend également une phase d'indexation de l'image requête. A la fin de cette étape, seules les images dont les distances à la requête sont jugées assez faibles qui sont retenues et renvoyées à l'utilisateur comme un résultat final de la recherche. L'évaluation de n'importe quel système d'indexation et de recherche d'images est basée sur deux paramètres : 1) La pertinence des images retournées par rapport à la requête. 2) Le temps nécessaire aux comparaisons entre la requête et les images de la base. Afin d'accélérer l'opération de la recherche, nous proposons une recherche de modèles hiérarchique. Dans notre architecture, les comparaisons au premier niveau s'effectuent entre les indices de couleur, et d'autres entre les indices de contour. Ces comparaisons sont très rapides et permettront la construction d'un espace de recherche réduit R . La recherche dans le deuxième niveau s'effectue dans cet espace R , en utilisant l'indice texture.

Les systèmes de recherche de modèles 3D passent trop de temps à décompresser ces modèles et à changer leurs poses et leurs emplacements. Ceci peut être évité en précédant l'opération de l'indexation par d'autres traitements. Dans le chapitre 04, nous proposons de normaliser et aligner tous les objets de la même classe. Grâce à ces opérations, ces objets auront la même pose, le même repère ainsi que la même origine. Afin d'éviter la décompression des objets 3D, nous proposons dans le chapitre 03, la construction des indices de l'image par les coefficients de la transformée en ondelettes.

Les systèmes d'indexation multi-vues des objets 3D, consistent à caractériser les différentes vues 2D. L'indexation d'une vue 2D doit caractériser la sémantique de la région d'intérêt représentant la forme de l'objet. Afin de capturer les caractéristiques locales des régions d'intérêt, il est préférable de limiter l'opération d'indexation à ces zones. Ceci s'obtient à travers l'utilisation d'une indexation d'images basée région[139, 145]. Certaines méthodes dites basées régions dans le domaine transformée se basent sur l'extrapolation des régions d'intérêt, pour les convertir en formes régulières. La région d'intérêt dans ce cas change fortement par rapport à

son état initial. Cette extrapolation aura un effet négatif sur la caractérisation de l'objet. Cette contrainte peut être facilement traitée par l'application de transformation adaptée à la région d'intérêt. Dans le chapitre numéro trois nous proposons la construction d'indices à partir des coefficients de la transformée en ondelettes basée région.

1.1.2 Motivations

Actuellement, la majorité des techniques d'indexation de l'image, ont opté pour la description du contenu visuel de l'image. Le principe de ces techniques, est de construire des signatures à partir des caractéristiques visuelles de l'image. Ces caractéristiques peuvent être : la couleur, la texture, ou la forme de l'image, ou des régions contenues dans l'image. D'autres techniques proposent la construction de signatures de l'image dans le domaine transformé de l'image. Dans le contexte 3D, les requêtes prennent généralement trois formes : un modèle 3D, un ensemble de vues 2D, ou un ensemble de dessins ou de traits. Lors de la phase de recherche, l'algorithme effectue une comparaison de la signature requête à la signature de l'image de la base, et retient les images à faibles distances.

L'indexation des modèles 3D peut toucher différents aspects allant de l'acquisition jusqu'au choix des métriques de comparaison. L'étude bibliographique des techniques d'indexation des modèles 3D, montre que la plupart des méthodes ne traitent ce sujet que partiellement. Une partie des méthodes s'intéressent au choix des attributs, d'autres tentent d'accélérer le processus de recherche. Dans notre travail, nous essayons d'aborder chaque aspect séparément des autres avant de les traiter. Ceci nous a conduits à travailler sur plusieurs sujets :

- La normalisation et l'alignement des objets 3D.
- L'extraction des vues 2D des light field.
- La transformée en ondelettes basée région.
- La caractérisation des vues 2D par les coefficients de la transformée en ondelettes basée région.
- La hiérarchisation du processus de recherche.
- Le choix de la métrique adaptée à notre méthode de recherche.

1.1.3 Objectifs

L'objectif d'un système d'indexation de l'image, est de retrouver rapidement les images similaires à l'image requête. Plusieurs approches ont essayé de satisfaire ces exigences. Cependant, les performances n'étaient pas toujours à la hauteur. En fait, ces techniques ont toujours eu des temps de calcul très importants et les résul-

tats de la recherche des images, n'étaient pas toujours précis. Un autre inconvénient provient du fait, que l'utilisateur ne peut initier que des requêtes sous forme de modèles 3D, rendant ainsi la recherche plus difficile. Dans ce travail, nous avons pris en considération les critères de performance des systèmes d'indexation et de recherche des modèles 3D, à savoir la rapidité dans l'exécution des calculs, la robustesse des résultats aux variations de l'échelle, et de la rotation.

L'objectif majeur de notre travail est de permettre à l'utilisateur d'interroger les bases de modèles 3D, par des requêtes 2D ou 3D. Dans le contexte actuel il s'agit de : 1) trouver les primitives les plus pertinentes à notre contexte 3D. En effet, le choix de primitives avait toujours un impact crucial sur les performances des systèmes d'indexation en termes de temps de calculs et de la qualité des résultats obtenus. 2) construire des signatures ou indices robustes à la rotation, la translation et au changement d'échelle de l'objet. Ces contraintes dépendent des conditions d'acquisition et de capture du modèle 3D : le choix de l'angle de la prise de vue et le paramétrage de la lumière. Donc la prise en compte de ces contraintes lors de la phase de recherche passe forcément par une indexation des différentes vues 2D. Dans notre travail, il s'agit d'un échantillonnage régulier de l'espace d'acquisition de l'objet 3D. Ceci correspond exactement à la notion des light Field, qui offrent cette possibilité avec une prise en considération des différents paramètres. 3) Enfin, lors de la phase de recherche, nous limitons l'espace de comparaison : seules les vues les plus significatives de l'objet sont à comparer avec la requête, les modèles dont les scores sont jugés faibles, feront l'objet d'un deuxième niveau de comparaison. Cela constitue un système de comparaison hiérarchique.

1.2 Organisation du manuscrit

Dans le chapitre II, nous présentons une étude bibliographique et comparative des différentes approches d'indexation et de recherche d'images 2D et 3D. Nous décrivons ensuite notre principale contribution pour l'indexation d'images basée région. Dans le chapitre III, nous nous intéressons à l'indexation des vues 2D par la transformée en ondelettes basée région. Cette technique se propose de caractériser les régions d'intérêt correspondant aux vues 2D, par leurs coefficients résultant de la transformée en ondelettes. Le chapitre IV, décrit les étapes de normalisation et d'alignement des objets 3D. Ce chapitre aborde le problème de l'invariance aux rotations, changement de l'échelle et de la luminance de notre algorithme d'indexation.

Dans le chapitre V nous présentons une nouvelle méthode d'indexation des light field par la transformée en ondelettes basée région. Nous proposons dans ce

chapitre, l'association d'un nombre réduit de vues 2D (appelées vues clé) à chaque objet de la base. La caractérisation de ces vues est réalisée selon la méthode décrite dans le chapitre III. Ces vues clés feront l'objet d'une comparaison premier niveau, seules les modèles dont les distances sont jugés faibles, vont être comparés à la requête au deuxième niveau de notre méthode de comparaison. Finalement, dans le chapitre VII, nous validons nos choix à travers les résultats obtenus sur des bases de modèles 3D. Nous procédons à la comparaison de notre méthode aux méthodes existantes d'indexation d'objets 3D. A la fin du manuscrit, nous rappelons nos contributions, et nous présentons nos communications, et publications scientifiques.

Chapitre 2

État de L'art Des méthodes d'indexation

Résumé : Comme nous le verrons dans le chapitre V, en indexation des objets 3D, les scènes sont souvent synthétisées à partir des vues 2D, qui sont obtenus sous différents angles de vue de l'espace. A cet effet, notre schéma d'indexation des objets 3D consiste à caractériser chacune de ces vues 2D par des descripteurs du contenu visuel de la région d'intérêt correspondant à l'objet lui même. En d'autre terme le principe de notre méthode est de sélectionner des vues 2D de l'ensemble des projections de l'objet 3D. Ces vues 2D sont ensuite caractérisées à l'aide d'une technique d'indexation d'image basées régions. La région d'intérêt doit impérativement correspondre à l'objet. Dans ce chapitre, nous commençons par la présentation d'un état de l'art des différentes techniques utilisées dans le domaine de l'indexation d'objets 3D, nous présentons ensuite les différentes techniques d'indexation des vues 2D. Enfin nous expliquons le principe de notre méthode portant sur l'indexation des light field par la transformée en ondelettes basée région (Shape Adaptive Discret Wavelet Transform : SA-DWT). Nous introduisons plus précisément trois nouveaux descripteurs de couleur, texture et de contour.

2.1 Introduction

Le problème de la recherche des objets 3D dans les grandes bases de données a incité beaucoup d'intérêt de la communauté scientifique. L'objectif principal de cet intérêt est le développement des méthodes d'accès et de recherche dans les bases d'objets 3D. La recherche est généralement basée sur une comparaison implicite entre la requête introduite par l'utilisateur (mots clés ou requêtes par l'exemple) et les objets de la base. Étant donné que la comparaison pixel à pixel est souvent très compliquée et elle ne peut être accomplie qu'en temps différé ; la solution était d'associer à chaque objet une signature décrivant au mieux son contenu visuel. Cette opération est communément appelée l'indexation de l'objet. Ce type de traitement permet de reconnaître la sémantique de l'objet 3D, sans aucune interprétation haut niveau de son contenu.

Ces deux dernières décennies ont vu apparaître plusieurs techniques d'indexation de l'objet. Les chercheurs dans ce domaine, se posent tous les mêmes questions : quelle est la meilleure façon à adopter pour caractériser le contenu visuel de l'objet. Quelles caractéristiques doit on utiliser pour exprimer sa sémantique.

Dans les premiers temps, les recherches ont été focalisées sur les méthodes d'indexation géométriques et topologiques des modèles 3D. Ces méthodes sont basées sur l'exploitation des caractéristiques géométriques ou topologiques des objets 3D, pour la construction de leur indices. Ces signatures ou indices peuvent ensuite être utilisés lors de la phase de recherche. Beaucoup de standards d'indexation et de recherche des objets se sont basées sur cette approche. Surtout que les primitives utilisées caractérisent très bien les particularités morphologiques de l'objet. Il est ainsi possible de distinguer les objets aux contours réguliers de ceux aux contours irréguliers. Cependant ces méthodes sont très fastidieuses et extrêmement coûteuses en terme de temps. En outre, ces méthodes se montrent très vite inefficaces lorsqu'il est question d'indexer des objets déformés, ou dont des parties sont occultées par d'autres objets. Autrement dit, ce type de méthodes nécessite la capture intégrale de l'objet. Aucune partie de l'objet ne doit être cachée ou déformée. Sinon, on se retrouve très vite avec un indice très différent, comme c'est le cas de l'objet *cube* dans la figure 2.1, où on peut constater que l'occlusion d'une petite partie de l'objet *cube* a changé complètement sa morphologie.

De nombreuses méthodes sont référencées dans la littérature afin de contourner ce genre de contrainte. Parmi les plus utilisées, on trouve celles qui sont basées sur les attributs de la texture et de la couleur des objets 3D. L'effet négatif de l'occlusion des objets peut être éliminé par l'utilisation de la caractéristique couleur ou texture.

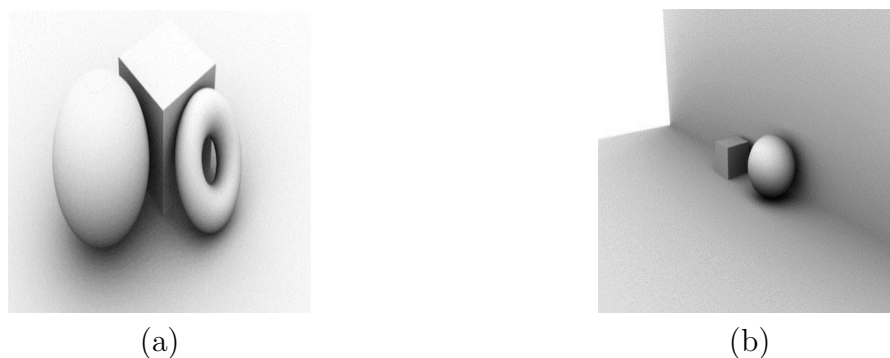


FIGURE 2.1 – *Illustration du problème de l'occlusion des objets 3D. (a) Un cube caché presque entièrement par d'autres formes 3D. (b) La partie cachée du cube est moins importante [2, 3].*

Les histogrammes construits à partir de ces caractéristiques représentent leur répartitions spatiales ou fréquentielles, aucune information topologique ou géométrique n'est prise en considération par ce genre de méthodes. Une seconde solution au problème de l'occlusion consiste à ne caractériser que les entités non occultées. Cette solution nécessite la prise de vue de différents angles et la caractérisation des vues couvrant l'intégralité de l'objet. Dans le cas du rendu 3D à l'aide des light field, il s'agit de caractériser les différentes vues 2D.

L'indexation de l'objet 3D basée vues consiste à caractériser ses projections 2D, en se basant sur son aspect visuel, chaque projection 2D est segmentée à plusieurs régions d'intérêt. Ces régions peuvent ensuite être indexées de la même façon que les images 2D, à l'aide d'une technique basée région [51, 96, 58].

Nous débutons ce chapitre avec un état de l'art des techniques d'indexation des objets 3D. Nous présentons ensuite les principales techniques d'indexation d'images basées région. Finalement, nous décrivons notre contribution dans le domaine de l'indexation et la recherche des objets 3D.

2.2 Etat de l'art des méthodes d'indexation des modèles 3D

Dans cette section, nous dressons un état de l'art des méthodes d'indexation des objets 3D. Selon le type de signature utilisée, on peut distinguer trois classes des approches de l'indexation : 1) Les méthodes basées sur les caractéristiques globales,

spatiales ou locales des modèles 3D. 2) Les méthodes basées graphes. 3) Les méthodes basées vues 2D. Les frontières dans cette classification peuvent facilement être franchies d'un type à l'autre. On peut ainsi rencontrer les méthodes hybrides, qui sont la combinaison de deux classes différentes. Par exemple les caractéristiques employées par les techniques de la classe 1 peuvent être combinées avec les caractéristiques des vues 2D, pour produire des techniques d'indexation hybrides. La figure 2.2 récapitule ces types et cite en outre les principales méthodes et techniques associées à chaque catégorie [93].

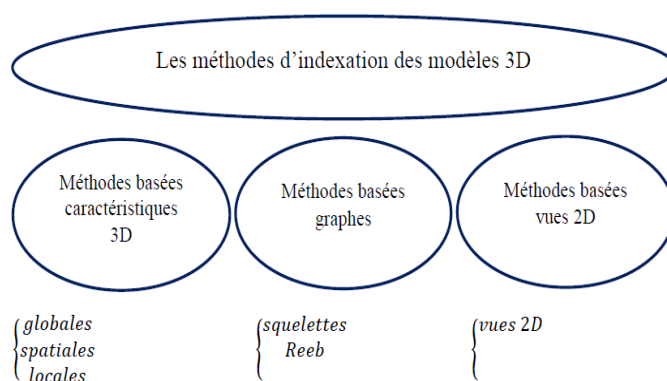


FIGURE 2.2 – Classification des méthodes d'indexation de modèles 3D

2.2.1 Méthodes d'indexation basées sur les caractéristiques 3D

Les caractéristiques 3D dans cette catégorie désignent généralement les propriétés géométriques et topologiques des formes 3D. Le choix de l'espace de description des caractéristiques a fait apparaître trois types de méthodes d'indexation : les méthodes basées caractéristiques globales, les méthodes utilisant les caractéristiques locales et les méthodes basées sur le partitionnement de l'espace.

2.2.1.1 Méthodes d'indexation basées caractéristiques globales

Le terme *globale* dans l'appellation de ce type de méthodes, vient du fait que ces méthodes caractérisent la forme 3D comme une seule entité. Autrement dit aucune caractérisation locale des régions constituant la forme 3D n'est employée. Ces caractéristiques peuvent porter sur :

- La géométrie.
- La topologie.

– La morphologie.
des objets 3D.

La caractérisation géométrique, topologique ou même morphologique des formes 3D, concerne généralement la distribution des formes. Parmi les caractéristiques employées, on peut citer par exemple :

- Les racines carrées de l'aire des triangles. Cet attribut nécessite une partition triangulaire de l'objet. L'indice est ensuite composé des racines carrées des aires de ces triangles.
- La distribution des points autour du centre de gravité de l'objet. Cette caractéristique mesure la distance entre le centre de gravité de l'objet et les sommets des triangles de son contour.
- Les degrés d'inclinaisons entre les sommets des triangles. Autrement dit, les angles entre les sommets de chaque triangle de la surface de l'objet.
- Les moments statistiques.
- Le volume et la surface.

Zhang et Chen [144, 128] se sont intéressés aux méthodes d'extraction des caractéristiques globales :

- Le volume.
- La surface.
- Les moments statistiques.
- Les coefficients de la transformée de Fourier.

Afin de s'adapter au matériel du rendu, les objets 3D sont représentés par une collection de vecteurs et de polygones 3D [144]. Cette modélisation est réalisée par le langage VRML (langage de la modélisation de la réalité virtuelle) [30]. Le volume de l'objet est égal à la somme des volumes des triangles situés à l'intérieur de l'objet. L'emplacement de ces triangles dépend du sens de leurs normales [30]. Zhang et Chen proposent également l'extraction des moments statistiques M_{pqr} , à l'aide de l'équation suivante :

$$M_{pqr} = \sum_i s_i \int \int \int x^p y^q z^r \rho_i(x, y, z) dx dy dz \quad (2.1)$$

où

i représente l'indice du polygone.

$$\rho_i(x, y, z) = \begin{cases} 1 & \text{si } (x, y, z) \in \text{objet} \\ 0 & \text{ailleurs} \end{cases} \quad (2.2)$$

s_i représente le signe du volume du $i^{\text{ème}}$ polygone.

(p, q, r) représentent l'ordre du moment statistique [4].

Enfin les auteurs proposent l'extraction des coefficients de la transformée de Fourier. Cette méthode consiste à appliquer la transformée de Fourier sur chaque triangle de la représentation VRML de l'objet.

Osada a repris la représentation VRML des objets 3D, dans leur papiers [98, 111]. La distribution aléatoire des points sur la surface de l'objet était utilisée pour la construction de quatre indices :

- L'indice des distances euclidiennes. Les auteurs mesurent les distances euclidiennes entre chaque pair de points de la surface.
- L'indice des racines aux carrés des surfaces des triangles superficielles de l'objet.
- L'indice des racines cubiques des volumes des tétraèdres de l'objet.

Les auteurs ont conçu ensuite un système de mesure de similarité entre les objets 3D. Ce système s'est basé sur ces indices pour mener l'ensemble de comparaisons entre la requête et les objets de la base. Les expériences effectuées ont montré que cette méthode est très performante. Cependant la phase de construction des indices nécessite des calculs très compliqués.

Paquet et Rioux [100] ont décrit chaque objet 3D par les orientations de ses faces. Il s'agit plus précisément de la caractérisation des objets par trois histogrammes : le premier histogramme construit à partir des angles entre la normale de la face et les deux principaux axes de l'objet. La construction du second histogramme est similaire au premier, néanmoins les deux axes sont considérés séparément ; c'est-à-dire pour chaque direction deux angles sont mesurés. Les auteurs mesurent la cooccurrence entre la direction des deux axes pour la construction du troisième histogramme. Considérons l'exemple ci-dessous :

α_1 : représente l'angle entre la normale de la face et l'axe 1.

α_2 : représente l'angle entre la normale de la face et l'axe 2.

Les auteurs proposent de construire le troisième histogramme à partir des valeurs de la co-occurrence entre les différents angles α_1 , et α_2 .

Outre leur efficacité à caractériser les objets maillés, ces méthodes se distinguent par leur robustesse aux translations et rotations. La limitation fondamentale des méthodes d'indexation basées caractéristiques 3D globales provient de l'indispensabilité de la connectivité entre les différentes faces de l'objet. En effet, ces méthodes montrent une insuffisance pour la description des objets à des faces de tailles différentes. En outre, ces méthodes ne peuvent pas capturer les petites va-

riations à l'intérieur des différentes faces de l'objet. A titre d'illustration, la figure 2.3 montre l'incapacité de ces méthodes à différencier entre les faces de deux objets distincts. En revanche, ces méthodes se montrent très rapides et efficaces dans le cas des objets à très grands degrés de connectivité entre les faces.

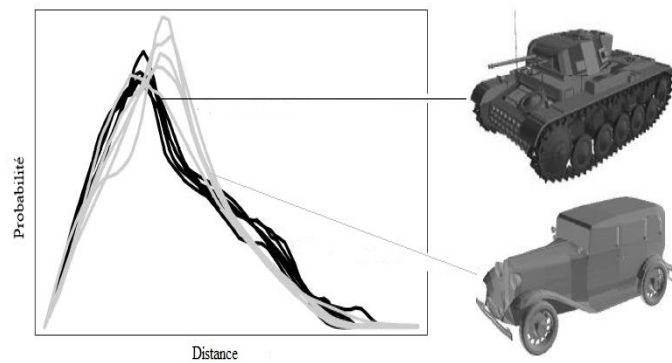


FIGURE 2.3 – *Distribution de formes de cinq tanks (en gris) et de six voitures (en noir). Image extraite de [93].*

Saupe et al ont proposé dans [115], une méthode de caractérisation des objets 3D basée sur leur harmoniques sphériques. Celles-ci résultent de l'échantillonnage de l'espace 3D de l'objet à des intervalles radiaux égaux. À l'instar de la décomposition de Fourier, les fonctions sphériques sont décomposées comme la somme des 16 premières composantes harmoniques. Les coefficients de ces harmoniques ainsi que les moments statistiques de l'objet 3D, sont ensuite utilisés pour construire les indices de l'objet. L'avantage principal de cette méthode réside dans son importante robustesse vis-à-vis la rotation de l'objet 3D.

2.2.1.2 Méthodes d'indexation basées caractéristiques locales

La notion locale est introduite dans cette catégorie grâce à l'utilisation de la sphère de Gauss. En effet, J. Little [79] a proposé le partitionnement de la sphère unité en espaces angulaires égaux. Chaque face de l'objet est ensuite projetée sur cet espace pour calculer sa normale. Afin d'améliorer les performances de ces méthodes, Wang et al ont proposé dans [138] d'ajouter une information relative à la distance entre les différents descripteurs locaux de l'objet, permettant ainsi de caractériser son aspect global.

2.2.1.3 Méthodes d'indexation basées sur le partitionnement de l'espace

L'indexation dans cette classe consiste à décrire la distribution géométrique des différentes faces du modèle 3D. Ceci en étudiant la direction des faces de l'objet 3D. Chaque face peut être caractérisée par :

- L'angle entre les différents points de la surface de l'objet.
- Le moment d'inertie autour de l'axe de l'objet.
- Les moments géométriques.
- Les moments de Zernike 3D.

La figure ci-dessous 2.4 illustre quelques propriétés statistiques de la forme, en l'occurrence les moments d'inertie, la distance moyenne à l'axe de la forme et la variance de la distance à l'axe de la forme. Ces propriétés sont calculées pour N points de la surface de l'objet [96].

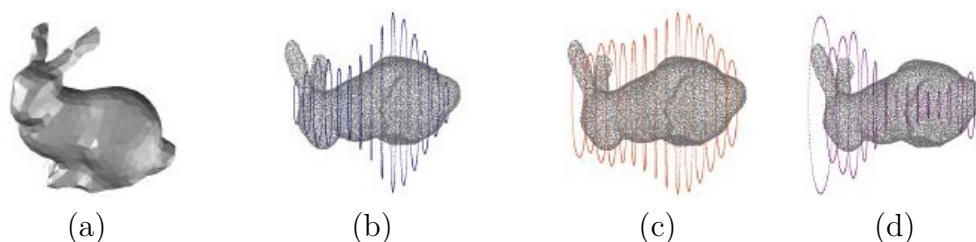


FIGURE 2.4 – Exemples de caractéristiques des faces de l'objet 3D. (a) Exemple d'un objet 3D, (b) La variation des moments d'inertie autour de l'axe de l'objet, (c) représente la variation des distances moyennes entre les points sur la surface et l'axe de l'objet. (d) Représente la variance de la distance de la surface par rapport à l'axe [96].

2.2.2 Méthodes d'indexation 3D basées graphes

Certains objets 3D peuvent être caractérisés par leurs articulations. Celles-ci sont robustes aux déformations et aux translations de l'objet, elles peuvent même donner la forme générale de l'objet à caractériser. A cet effet, plusieurs approches d'indexation se sont basées sur ces caractéristiques [78, 128].

2.2.2.1 Méthodes d'indexation par squelettes

La caractérisation des modèles 3D par les propriétés statistiques de leurs contours s'effectue généralement de différentes manières. On peut citer par exemple le cas de l'utilisation des distances entre les sommets des triangles et le centre de gravité de l'objet. En 1967 Blum a introduit un nouveau concept, basé sur un ensemble

restreint des points du contour de l'objet. S'inspirant des squelettes des animaux, Blum a proposé la caractérisation des objets 3D, par les propriétés statistiques de leur squelette. Le squelette dans ce cas est obtenu par l'intersection entre la sphère maximale à l'intérieur de l'objet et son contour. Les arrêtes correspondent aux rayons reliant le centre de gravité de la sphère aux points d'intersection. Les propriétés géométriques de ces nœuds et arrêtes sont ensuite utilisées pour construire les indices de l'objet [28].

Cette technique s'avère performante en ce sens qu'elle permet une meilleure comparaison des objets, surtout ceux qui sont articulés. De par ses attributs, elle permet aussi de faire des recherches partielles de l'objet. Dans ce même esprit, Sandar et al [125] proposent une méthode d'indexation basée sur l'utilisation du squelette. Dans ce travail la représentation de la forme par son squelette consiste à voxiliser l'objet et ensuite de réduire le nombre de voxels, grâce à l'utilisation de la transformée de distance[52]. Les voxels sont obtenus par l'application de l'algorithme de l'arbre couvrant le minimum sur l'ensemble des points. Durant ce processus, certains voxels sont supprimés, d'autres sont préservés dans la représentation finale de l'objet.

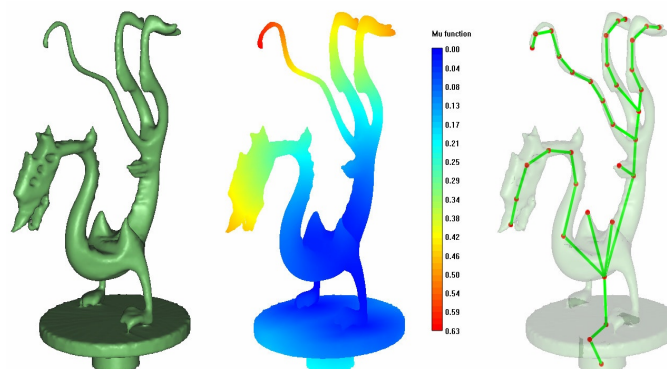


FIGURE 2.5 – Exemple de squelette de forme 3D [133]

Dans [78] S. Lin et al ont proposé une méthode d'indexation basée sur l'utilisation des squelettes. Cette méthode se déroule en deux étapes :

- Extraction du squelette de l'objet.
- Extraction des caractéristiques des squelettes.

L'extraction du squelette de l'objet consiste à simplifier l'objet 3D, grâce à l'utilisation de l'erreur quadratique [55]. Les polygones de l'objet sont ensuite réduits à l'aide de la méthode du Laplacien [15]. L'indice du squelette est enfin construit à partir des gradients orientés de ses pixels.

2.2.2.2 Méthodes d'indexation par les graphes de Reeb

Ces méthodes sont basées sur la caractérisation de la structure générale de la forme. Parmi toutes les méthodes de représentation topologique des formes, les graphes de Reeb sont sans doute les plus connus et les plus répandus, notamment pour représenter les objets à grands degrés d'articulation [128]. La caractérisation d'un objet par le graphe de Reeb, consiste à décrire la topologie de la forme en ses points critiques. Ce type de graphe est le résultat de l'application d'une fonction de segmentation μ sur la surface de l'objet. Les recherches dans ce domaine ont porté principalement sur la définition de cette fonction [110]. Dans [59] Hilaga et al proposent une fonction basée sur l'utilisation de la distance géodésique, pour la détection des nœuds et des arrêtes du graphe de Reeb. En fait, le résultat de la distance entre les sommets des différents maillages, est le critère principal pour la définition des nœuds.

C'est ce même type de graphe que l'on retrouve dans le travail proposé par Biasoti et al [26]. Etant donné que deux modèles identiques ne peuvent avoir que deux graphes similaires. La comparaison entre ces modèles consiste à évaluer le nombre de nœuds similaires entre leurs graphes. Ainsi il est possible de procéder à une comparaison partielle entre les différents modèles 3D.

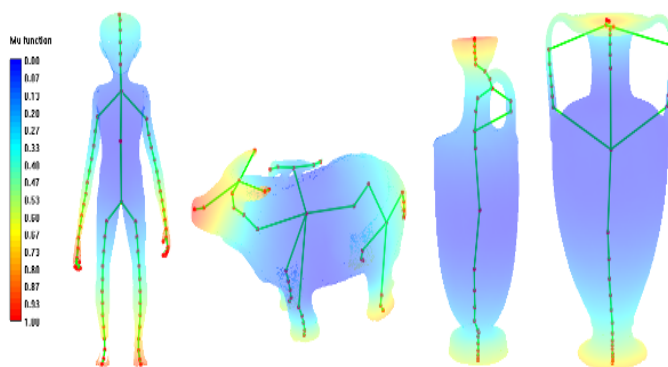


FIGURE 2.6 – Exemples de graphes de Reeb de différents modèles 3D[132]

2.2.3 Méthodes d'indexation basées vues 2D

Les méthodes d'indexation de modèles 3D basées vues 2D, portent sur la caractérisation de l'aspect visuel de ces vues. Lesquelles sont utilisées lors de la phase du rendu. Parmi les premiers travaux proposés dans cette famille d'approches, figure celui de Chen et al [32]. Les auteurs dans ce travail ont introduit pour la première

fois la notion de descripteurs de Light Field. Le principe de ce travail est de projeter le modèle 3D sur vingt sommets d'une forme dodécaèdre pour extraire dix prises de vues, et éliminer les dix autres par opposition orthogonale. Chen et al caractérisent ensuite chacune de ces silhouettes par un descripteur de 36 moments de Zernike [143] et un descripteur de dix coefficients de Fourier. L'avantage de cette méthode réside dans l'échantillonnage de l'espace de prises de vue, grâce à l'utilisation de la fonction pléoptique des light field. Il est ainsi possible de réduire le nombre de vues, sans l'utilisation des méthodes de calculs compliquées.

Un autre avantage viendra cette fois-ci de la prise en compte de toutes les rotations possibles lors de la phase de la caractérisation, garantissant ainsi une invariance aux rotations de ce système d'indexation. En outre cette méthode supporte les requêtes à deux et trois dimensions, on parle donc de système d'indexation à requêtes multidimensionnelles. Cependant, cette méthode emploie des calculs très compliqués lors de la phase de recherche.

L'approche usuelle de l'indexation des objets 3D basée vues consiste à :

– *Capter les vues 2D*

Cette étape est basée sur un mur de caméras pour la capture simultanée des vues 2D.

– *Sélectionner les vues clés*

Il est communément admis que l'augmentation du nombre de vues, améliore les performances du processus de l'indexation. Or, ceci peut donner lieu à des calculs très compliqués et à l'augmentation de l'espace de stockage. À cet effet, un compromis *Nombre de vues/Qualité d'indexation* doit être assuré. Ceci est possible, si on ne retient que les vues clés lors de la phase de l'indexation [54].

– *Caractériser les vues 2D*

Cette étape consiste à construire les indices des vues 2D, à partir de leur attributs.

– *Mesurer la similarité*

Cette opération est basée sur la comparaison entre les indices des vues 2D.

Plusieurs méthodes d'indexation des objets 3D par les vues 2D ont été proposées. D'une façon générale tous les descripteurs de forme peuvent être utilisés pour caractériser les vues 2D. Ce qui change plus particulièrement d'une méthode à une autre c'est le nombre de vues 2D. Ces vues doivent être descriptives du contenu de l'objet, leur choix doit obéir à plusieurs critères :

- La forme de l'objet dans ces vues doit être intégrale. Aucune occultation ni déformation n'est tolérée.
- L'intervalle angulaire entre ces vues doit être régulier.

Parmi les méthodes proposées, on peut en particulier citer celle de T.F. Ansary et al, qui ont utilisé un regroupement adaptatif des vues 2D. Le résultat de celui-ci est un ensemble de vingt à quarante vues 2D [14]. S. Mahmoudi et M. Daoudi ont conçu une méthode qui permet de sélectionner sept vues à partir de trois directions principales et quatre directions secondaires [85]. L'indexation dans cette méthode est basée sur l'attribut contour. Dans son travail de thèse Thibault Napoléon a proposé un système d'indexation multi-vues [93]. Ce système est basé sur trois indices des vues 2D :

- Un indice de la forme de la région d'intérêt.
- Un indice de contour basé sur les coefficients de la transformée en cosinus discret (DCT).
- Un indice construit à partir des coefficients de la transformée de Fourier.

2.2.4 Méthodes d'indexation hybrides

Comme leur nom l'indique, les méthodes hybrides sont basées sur la combinaison de plusieurs caractéristiques de la forme 3D. La forme peut alors être caractérisée par une combinaison géométrique, topologique de descripteurs, comme il est possible de construire des descripteurs combinant les caractéristiques géométriques et les caractéristiques de vues 2D. Plusieurs configurations de descripteurs sont alors envisageables, relativement au type de l'application. Li et al dans [141] proposent par exemple de combiner les caractéristiques géométriques et les caractéristiques des vues 2D de l'objet. La forme est alors caractérisée par un descripteur à quatre dimensions : une dimension caractérisant les moments de Zernike, une deuxième caractérisant les coefficients de Fourier. Les deux derniers caractérisent l'aspect géométrique de la forme : l'un est basé sur les paramètres de profondeur, et l'autre est construit à partir des rayons géométriques de la forme.

De nombreux auteurs ont ensuite proposé d'autres méthodes hybrides d'indexation, en particulier des méthodes caractérisant les aspects géométriques et topologiques de la forme. La méthode d'indexation proposée dans [107] représente la géométrie de la forme par la signature de kernel, la caractéristique topologique est représentée par la fonction de diamètre de la forme, et de la distance géodésique moyenne entre points de la surface de l'objet.

Syntèse

L'état de l'art que nous venons de présenter donne un aperçu des méthodes d'indexation des objets 3D. Quoiqu'elle soit non exhaustive, cet état de l'art montre clairement que les chercheurs dans ce domaine ont abordé le sujet de l'indexation de

différentes manières. D'autres parts, cette étude bibliographique révèle que l'indexation des objets 3D basée vues 2D est plus appropriée à ce type d'images, car elle est robuste aux changements des conditions d'acquisition des objets 3D. En effet, cette méthode offre la meilleure solution aux problèmes d'occultation et de déformation des objets 3D. Ceci en capturant l'objet de différents angles de vues. A cet effet, il nous a semblé intéressant d'utiliser cette méthode pour l'indexation des objets 3D. A l'instar des méthodes basées vues, dans ce travail, nous traitons les points suivants :

- La normalisation et l'alignement des objets 3D.
- L'extraction des vues 2D.
- La caractérisation basée région des vues 2D.
- La mesure de similarité.

Dans la suite de ce chapitre, nous étendons notre étude des méthodes d'indexation basées vues 2D. Nous nous intéressons plus particulièrement à la phase de caractérisation des vues 2D, par les méthodes d'indexation basées région.

2.3 Méthodologie des systèmes d'indexation et de recherche d'images basés régions

Depuis le début des années 90, les scientifiques dans le domaine de l'indexation de l'image se posent tous le problème de l'indexation sémantique de l'image. La RBIR (Région Based Image Retrieval : l'indexation de l'image basée région) se présente comme une alternative aux techniques d'indexation CBIR (Content Based Image Retrieval : l'indexation de l'image basée contenu). Contrairement aux techniques CBIR, les techniques RBIR se basent essentiellement sur la mesure de similitude entre les différentes parties de l'image.

L'état de l'art des méthodes d'indexation d'images basée régions, met en évidence deux catégories d'approches [66, 41] :

- Indexation d'images basée région dans le domaine pixel.
- Indexation d'images basée région dans le domaine transformée.

A l'instar des méthodes d'indexation basées contenu, l'indexation de l'image basée région dans le domaine pixel, consiste à extraire les attributs (couleur, texture ou forme) de la région d'intérêt, de mesurer les statistiques de ces attributs des différentes régions de l'image et enfin d'utiliser les statistiques obtenues pour la construction d'un indice de l'image.

En indexation de l'image à l'aide des attributs du pixel, les axes de recherche portent particulièrement sur le choix de l'attribut primaire et la méthode de segmentation de l'image en régions d'intérêt. La grande difficulté rencontrée dans ce type de travaux vient de la nécessité de segmenter automatiquement et objectivement toute l'intégralité de la base d'images, sans aucune intervention humaine [142, 113]. L'objectif de la segmentation dans les systèmes de l'indexation d'image est de partitionner l'image en un ensemble de régions.

2.3.1 Extraction des régions de l'image

La première contrainte imposée par n'importe quelle technique d'indexation de l'image basée région concerne le partitionnement de l'image en un ensemble de régions d'intérêt. La littérature dense de ce thème montre qu'il n'y a pas un algorithme de segmentation optimal pour toutes les bases d'images. Certains chercheurs se contentent de partitionner l'image en un ensemble de blocs de tailles fixes. Zhang par exemple, a proposé de segmenter l'image en 32 segments. Pour chaque bloc une moyenne de niveau de gris est ensuite calculée, les 32 valeurs des niveaux de gris sont enfin regroupées à l'aide de l'algorithme *K-means* (en utilisant autant de groupes que de segments). Cette méthode permet une bonne limitation des régions d'intérêt, cependant elle est trop dépendante de leur positions, de l'échelle et des rotations des régions d'intérêts dans l'image [60].

Dans [106], les auteurs se sont basés sur la couleur dominante de l'image, comme une caractéristique discriminante pour la segmentation. Dans l'exemple de la figure 2.7 l'image est divisée en neuf blocs de tailles fixes, la couleur dominante est celle du bloc 5. À cet effet, ce bloc est choisi pour qu'il soit indexé et exprimer la sémantique de l'image.

La robustesse de ces approches aux conditions de translation, rotation et du changement de l'échelle, ont encouragé d'autres chercheurs à s'en inspirer. À l'instar de Tsai, qui a proposé de segmenter l'image en 5 régions (le coin haut de gauche, le coin haut de droite, le coin en bas à gauche, le coin en bas à droite et le coin du centre). D'autres ont proposé de diviser l'image en 9 sous régions, en se basant toujours sur la primitive couleur et sur des algorithmes de regroupement des données [60].

Le travail proposé dans [44] se distingue par la modélisation des sémantiques dans les images naturelles par des fonctions mathématiques. L'application de ces fonctions repose sur une segmentation initiale à l'aide de l'algorithme de décalage

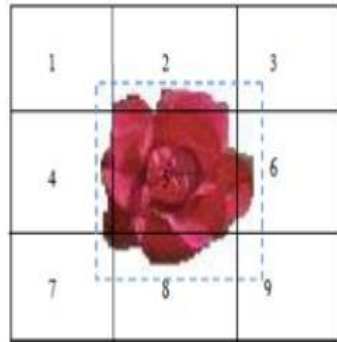


FIGURE 2.7 – Illustration de la méthode de division de l'image en blocs et la sélection du bloc dominant.

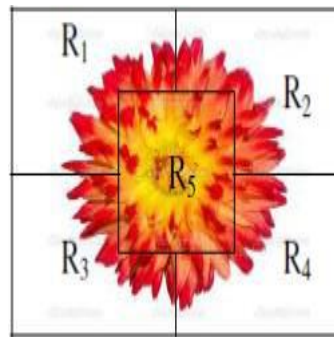


FIGURE 2.8 – Illustration d'une méthode de la segmentation de l'image à l'aide de l'attribut couleur.

moyen (mean shift algorithm) [37].

L'une des approches de la segmentation des images naturelles consiste à simuler le système de perception visuelle humaine. Vu son efficacité et l'objectivité de ses résultats, cette approche est largement utilisée. Marques avait proposé dans [86] l'utilisation d'un modèle de simulation de la perception visuelle humaine, ce simulateur est appliqué sur une grille de régions issues d'une segmentation initiale de l'image, où chaque région est extraite à l'aide de ses caractéristiques couleur et texture. La région d'intérêt correspond à la région dominante. Si l'image contient plusieurs régions d'intérêt, un algorithme de regroupement est appliqué sur l'ensemble de régions obtenues [86].

2.3.2 Description sémantique des régions de l'image

Concernant maintenant la deuxième contrainte, à savoir la description du contenu sémantique de l'image, celle-ci a été traitée de plusieurs façons dans la littérature. Par exemple on peut citer le travail de Tsai, où chaque image est segmentée en 5 régions. La description dans cette méthode est basée essentiellement sur la concaténation des histogrammes de la couleur et de la texture. De façon presque similaire, Vimini a proposé de segmenter l'image en 09 régions. Dans ce travail, chaque région est décrite par une concaténation des histogrammes de couleur et de texture. L'histogramme de la couleur est basé sur la quantification des niveaux de couleur de l'espace HSV. Tandis que la caractéristique texture est calculée à partir de la matrice de co-occurrence. Cette description est complétée par un histogramme de toute l'image. La limitation principale de ces approches réside dans leur dépendance des conditions d'éclairage et de luminance. Ces approches contraignent également l'utilisateur à un temps d'attente très intéressant, à cause des complexités de calculs et des opérations de transformation inverses (surtout que la plupart des images se trouvent sous formats compressés dans les bases d'images). Plus récemment les chercheurs ont proposé la construction des indices directement dans le domaine compressé. Ce type d'approches fournit plus d'information concernant la variation de la texture et de la couleur dans les domaines spatial et spectral [39].

La méthode proposée dans [61] est basée sur une description globale et locale de l'image, par les attributs couleur et texture. Ces descripteurs sont issus de l'application des filtres de Gabor sur l'image et ses régions. Lesquelles sont regroupées au préalable, en se basant sur la similarité entre leurs descripteurs. Contrairement à cette méthode, la méthode IRM [75] est basée sur la comparaison de chaque région de l'image requête aux différentes régions de l'image candidate. La méthode Fuzzy-club extrait les régions de chaque image, à l'aide d'un algorithme de segmentation non supervisée [146]. Ces régions sont ensuite caractérisées à l'aide de trois indices (indice de couleur, de texture et de forme). Rao et al ont proposé dans [109], une nouvelle présentation de l'histogramme de couleur, qui est basée sur la distribution spatiale des sous-ensembles de l'attribut couleur. Lakshmi et al se sont intéressés dans [72] davantage à l'aspect caractérisation que l'aspect segmentation. Cette méthode caractérise chaque région à l'aide de trois descripteurs (de l'attribut couleur et texture). Le descripteur de couleur est la moyenne des occurrences de l'histogramme de couleur. La texture est caractérisée à l'aide de l'homogénéité et du contraste.

La transformée en ondelettes par exemple peut être employée pour capturer les caractéristiques spatiale et spectrale à plusieurs niveaux de résolution. Surtout qu'elle se dispose d'importantes propriétés de concentration de l'énergie et de cor-

relation des coefficients. Vu ces propriétés, beaucoup d'approches l'ont utilisé pour capturer les informations spatiale et spectrale des régions d'images. Cependant, cette technique se montre inefficace pour la caractérisation des singularités des contours dans l'image [22]. Pour cette raison, certains chercheurs ont proposé l'utilisation des filtres de Gabor [81], et d'autres utilisent la transformée en curvelet [145].

Contrairement à la transformée en ondelettes, les filtres de Gabor peuvent capturer facilement les caractéristiques de texture à plusieurs orientations et échelles [145]; ceci permet une significative amélioration dans les performances de la caractérisation des régions de l'image. D'autre part, la transformée par curvelet permet une adaptation de la transformée aux inclinaisons et courbures des formes dans l'image [145]. Par conséquent les curvelet capturent mieux les contours des formes. La méthode de [94] a proposé de caractériser les différentes régions de l'image à l'aide de descripteurs de texture et de couleur, ces descripteurs sont construits à partir des coefficients de la transformée en ondelettes des régions de l'image. La particularité de cette méthode est l'utilisation de fenêtres de tailles variables pour l'extraction des régions de l'image. Ces régions sont ensuite regroupées, en mesurant la similarité entre leur descripteurs [94]. Contrairement à cette méthode, Wang et al ont proposé dans [139], la segmentation de l'image à l'aide des fenêtres de tailles fixes 4*4 pixels. Les descripteurs dans cette méthode sont extraits de la transformée en ondelettes des régions de l'image.

2.4 Indexation des light-field

Maintenant que nous avons passé en revue les principaux travaux de l'indexation 3D, penchons-nous sur le processus d'indexation des light field proprement dit. Celle-ci consiste à caractériser le light field (ensemble de vues 2D de l'objet 3D), ou une projection 2D de l'objet 3D. Autrement dit, l'objet 3D est ici considéré non pas en tant que tel, mais plutôt comme un ensemble de projections 2D. La caractérisation d'un objet 3D est une opération complexe. En effet, cette opération est subjective du dispositif d'acquisition utilisé, et des conditions dans lesquelles s'est déroulée cette acquisition du modèle 3D : la forme de ce modèle dépend principalement de l'orientation de la caméra, et de l'illumination du milieu dans lequel il se trouve durant l'opération de l'acquisition. Autrement dit chaque objet 3D est fournis avec une orientation et une échelle quelconque [53].

Un système d'indexation et de recherche de modèles 3D est toujours confronté à plusieurs contraintes, l'une d'entre elles est la robustesse aux changements de

l'échelle, la translation et la rotation dans l'espace du modèle 3D. En fait, pour rendre un indice invariant à : la translation, le changement de l'échelle, et à la rotation du modèle, de nombreux prétraitements doivent être effectués avant de procéder à la caractérisation du modèle lui-même. Ces prétraitements consistent à :

- Définir le centre de l'échelle des coordonnées.
- Mettre à l'échelle le modèle.
- Définir une représentation canonique pour chaque classe.

Dans ce travail de thèse les opérations de prétraitement sont abordées et expliquées dans le chapitre 04.

Dans notre cas, il est également nécessaire de passer d'un modèle en trois dimensions à une représentation en deux dimensions. Afin d'accomplir cette tâche, nous avons pris en compte deux contraintes. La première concerne le nombre de projections issues de ce passage $3D \rightarrow 2D$. Quoiqu'il ait un impact positif sur les performances de l'indexation, au-delà d'un certain seuil, le nombre de projections pourrait être un facteur négatif. Surtout que l'augmentation du nombre de projections rend les calculs plus compliqués, et nécessite un temps de calcul très important. La deuxième contrainte concerne l'échantillonnage de l'espace d'acquisition du modèle 3D. Grâce à l'utilisation d'une dense fonction d'échantillonnage les light field, permettent de capturer des modèles 3D à très haute résolution. Ceci nous a encouragés à opter pour cette technique d'acquisition et de rendu des modèles 3D. Dans notre cas, les modèles 3D sont téléchargés, aucune capture ou acquisition d'objets 3D n'est effectuée. Nous nous contentons de simuler le système d'acquisition des light field pour réaliser la projection du modèle 3D sur une sphère.

2.5 Principe de notre contribution dans le domaine d'indexation des vues 2D

Dans notre travail d'indexation des vues 2D basée région, on a essayé de palier aux problèmes posés par l'utilisation de la transformée en ondelettes pour la caractérisation des vues 2D. On a ainsi proposé d'inclure la caractérisation du contour des régions d'intérêt dans les descripteurs de l'image, en se basant sur l'algorithme DLEP. D'autres part, on a développé une transformée en ondelettes qui peut s'adapter aux formes à indexer, et qui est capable de limiter la caractérisation aux régions d'intérêt. Nos descripteurs, qui se présentent sous formes d'histogrammes sont construits à partir de patterns de couleur, texture et de contour. Nos descripteurs sont les résultats de combinaisons pondérées de descripteurs associés à chaque région de l'image. Cet algorithme est détaillé dans le chapitre suivant.

2.6 Conclusion

Dans ce chapitre, nous avons dressé un état de l'art de l'indexation des objets 3D. Cette étude bibliographique a mis le doigt sur les avantages et les inconvénients de chaque méthode. L'indexation multi-vues semble être la plus robuste vis-à-vis la déformation des objets 3D. Cet avantage nous a motivés pour utiliser cette approche dans notre système d'indexation des objets 3D. Cependant celle-ci nécessite des prétraitements permettant de convertir l'objet 3D en un ensemble de vues 2D.

Étant donné que les objets 3D dans notre cas sont constitués d'un ensemble de vues 2D, nous pensons qu'il est important d'aborder le sujet de l'indexation d'images 2D. Nous avons même présenté un état de l'art des techniques d'indexation d'images 2D, pour comprendre mieux ce sujet et repérer les approches les plus performantes dans ce domaine.

En outre, nous avons soulevé dans ce chapitre plusieurs aspects relatifs à la complexité de calcul du processus d'indexation et à l'adaptation de la caractérisation de la région d'intérêt. Ces contraintes feront l'objet d'une étude plus approfondie dans les chapitres 3, 4, 5 et 6.

Chapitre 3

Indexation des Vues/requêtes 2D Par la Transformée en Ondelettes Basée Région

Résumé - Nous avons vu dans la partie précédente que la caractérisation des régions d'intérêt 2D est au centre des méthodes d'indexation de modèles 3D. La tâche d'indexation d'un objet 3D peut être traduite par une indexation de ses projections 2D. On a même justifié notre choix d'utiliser la transformée en ondelettes basée région pour caractériser ces vues 2D. Ce chapitre sera consacré à la présentation de notre schéma de caractérisation basée région des images 2D. On commence par l'explication de la transformée en ondelettes basée région, avant de passer à la présentation de notre contribution dans le domaine de l'indexation basée région sans extrapolation. Dans ce chapitre nous introduisons les descripteurs de couleur, de texture et du contour et la notion de la concaténation des histogrammes des différentes régions de l'image.

3.1 Introduction

Généralement le rendu des scènes tridimensionnelles est basé sur la modélisation géométrique ou topologique [13, 99]. Ceci fût parmi les premières approches dans le domaine du rendu 3D. C'est d'ailleurs ce type de technique qui est le plus utilisé dans les applications 3D du net. Cependant le rendu basé géométrique n'est pas sans inconvénient. En fait, il est souvent difficile de modéliser mathématiquement des formes arbitraires et de distributions irrégulières de détails, comme c'est le cas des feuilles des arbres par exemple. De ce fait, beaucoup sont les chercheurs qui ont proposé d'augmenter le degré du photoréalisme des scènes 3D, en plaquant des textures des images réelles sur les modèles 3D [57, 119]. Un peu plus tard, d'autres ont proposé le rendu basé images, en synthétisant les modèles 3D à partir d'un ensemble d'images 2D. Celles-ci représentent les vues 2D prises à partir de différents angles de vue. Du fait, que le processus du rendu de modèles 3D doit être en mesure de visualiser tous les angles de vue de l'espace, il était nécessaire de concevoir une technique d'échantillonnage dense de l'espace des angles de vue[119, 92].

Ce contexte a vu naître en 1946 la technique des light field, qui est basée sur une représentation compacte de la fonction plénoptique. Cette approche permet en fait de se passer de la modélisation géométrique, grâce à l'estimation des champs de lumière aux alentours de l'objet 3D [74]. L'indexation des modèles 3D basée vues exploite les informations et attributs, issus des projections 2D du modèle lui-même. Contrairement aux méthodes d'indexation classiques qui sont basées sur les informations géométriques et topologiques de la modélisation mathématique de l'objet, ces méthodes n'utilisent aucune information relative au modèle mathématique[21].

L'indexation 3D basée vues 2D fonctionne en deux étapes. La première étape consiste à générer les vues 2D et à l'extraction de la région d'intérêt. Dans la deuxième étape, le nombre de projections 2D se voit réduit grâce à une opération d'échantillonnage des angles de vue. Les projections qui résultent de cette opération contribuent à la construction du descripteur final.

Les objets que nous caractériseront dans le chapitre V ne représentent que des régions plus ou moins intéressantes du modèle 3D (la région du fond et la région d'intérêt de l'image de l'objet), nous nous sommes tournés dans cette thèse, vers l'indexation basée région. Au début de ce chapitre, nous présentons le schéma de principe de notre méthode d'indexation d'images basée région. Cette section sera suivie par un aperçu comparatif des différentes méthodes de transformation de l'image basée région. La seconde partie de ce chapitre est consacrée à l'explication de notre principale contribution, qui a fait l'objet d'une publication dans *International*

Journal Image, Graphiques and Signal Processing [22]. Cette partie concerne plus particulièrement la construction des indices de l'image à partir des coefficients en ondelettes.

3.2 Schéma global

Comme nous venons de le présenter, notre système d'indexation d'images basé région est un processus complet de traitement de la requête et des images de la base. Sa conception doit prendre en compte des contraintes diverses et variées. Celles-ci concernent l'ensemble des étapes, qui vont de l'extraction des régions d'intérêt, jusqu'à leur indexations. La figure 3.1 illustre les étapes de notre schéma d'indexation :

1. Extraction des régions d'intérêt
2. Extraction des composantes Y , C_b , C_r .
3. Transformation des trois composantes par la transformée en ondelettes basée région.
4. Regroupement des caractéristiques.
5. Construction des histogrammes des patterns de : Couleur, Texture et de forme.

3.2.1 Extraction de la région d'intérêt

Avant d'appliquer la transformée en ondelettes basée région, l'image doit être segmentée en régions d'intérêt. Dans ce travail, nous avons utilisé une méthode de segmentation des images naturelles, dont le nombre de régions d'intérêt est défini par l'utilisateur ou par l'homogénéité des régions. Pour faire face à ces contraintes, nous utilisons la méthode décrite dans le papier [90]. Celle-ci est basée sur la modélisation des régions de texture homogènes par une distribution gaussienne et la modélisation des contours par un code adaptatif. Un résultat de l'application de cet algorithme de segmentation est illustré sur la figure 3.2

3.2.2 Transformée en ondelettes basée région

Les standards de codage et de compression basés objets tel que le MPEG 4, offrent une meilleure manipulation et une très bonne qualité visuelle des objets par rapport aux standards basées pixel[70]. Dans les premières méthodes de codage basé objet, les formes irrégulières dans l'image sont converties à des formes rectangulaires, avant d'être transformées. Cette conversion nécessite l'ajout de pixels supplémentaires

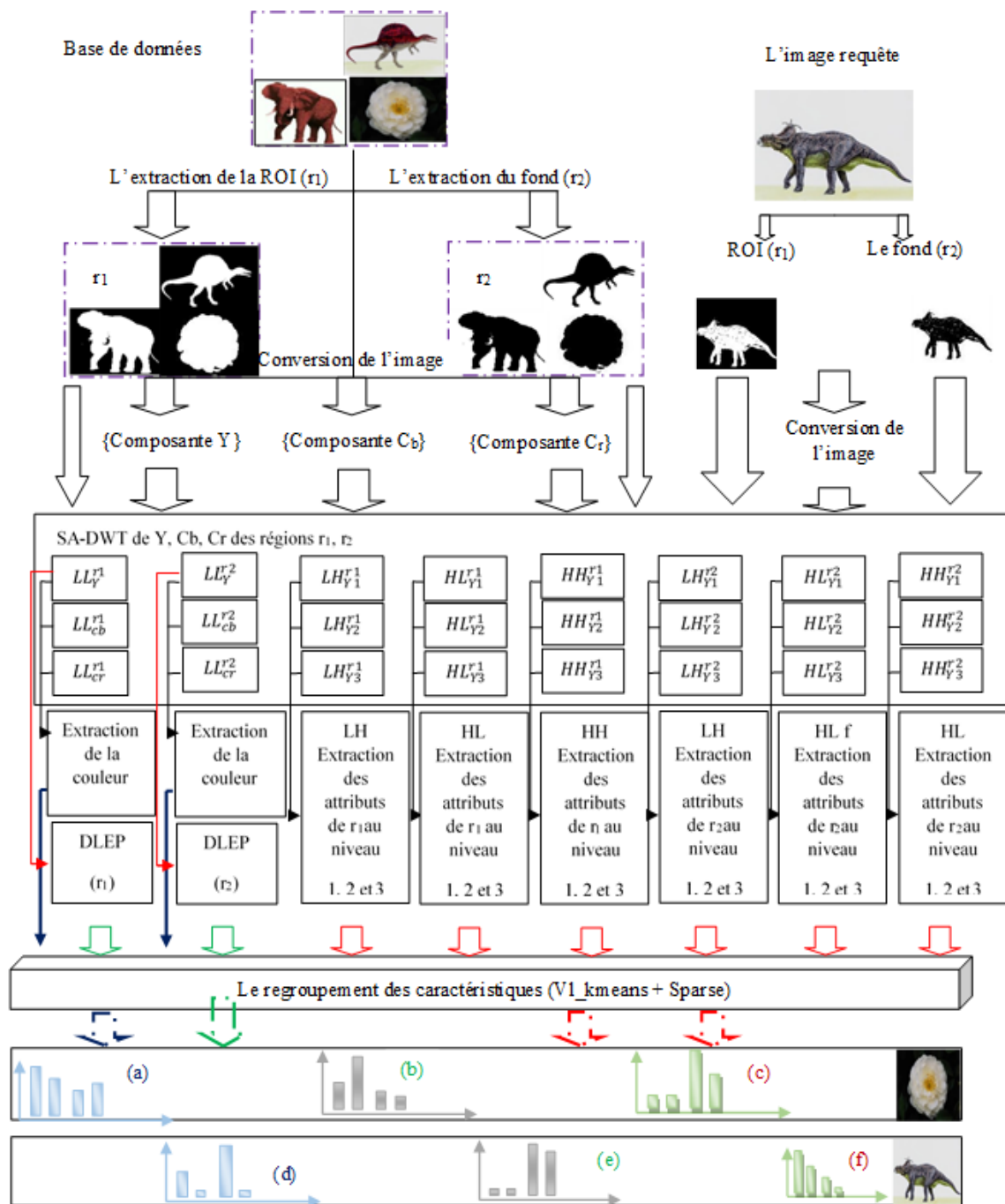


FIGURE 3.1 – Principe de notre méthode d'indexation d'images à l'aide de la transformée en ondelettes basée région. L'image candidate et l'image requête sont segmentées en deux régions chacune. (a), (b), (c) représentent les histogrammes de couleur, de contour et de texture respectivement de l'image candidate. (d), (e), (f) représentent les histogrammes de couleur, de contour et de texture respectivement de l'image requête.

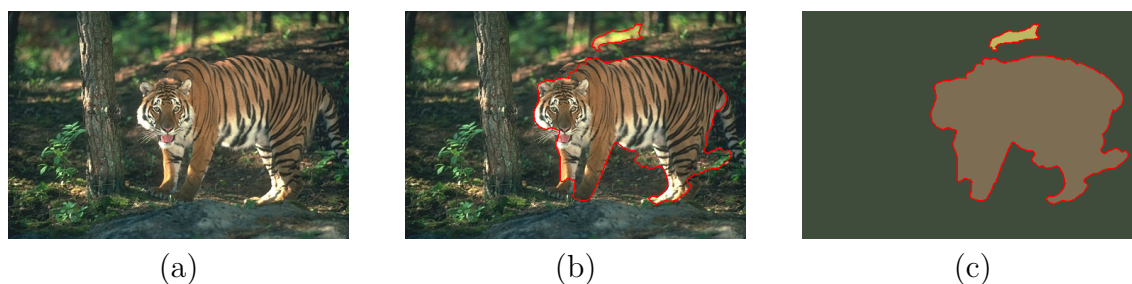


FIGURE 3.2 – Exemple d'extraction de la région d'intérêt, (a) représente l'image initiale à segmenter, (b) représente l'extraction du contour de la région d'intérêt, (d) représente la région d'intérêt.[90]

aux formes initiales. Ce remplissage (*le padding*) des régions d'intérêt peut se réaliser de différentes manières, des valeurs nulles peuvent être ajoutées. D'autres travaux sont basés sur l'extrapolation entre les régions intérieures et extérieures de l'objet, pour remplir ces régions. Quoiqu'elles ont de faibles taux d'erreur binaire et une complexité très faibles, les opérations d'extrapolation et du padding citées ci-dessus influencent négativement sur la qualité du codage, en modifiant la corrélation entre les différentes lignes et colonnes de la région d'intérêt. En d'autres termes, la région transformée ne correspond pas exactement à la région initiale. C'est pourquoi une alternative permettant la préservation des caractéristiques de la région d'intérêt a été proposée.

Dans cette optique, la transformée en cosinus discrète basée région (Shape Adaptive Discret Cosinus Transform : SA-DCT) fût la première transformée basée région, quia été proposée. Le principe de cette technique consiste à ramener toutes les valeurs à l'intérieur des blocs 8×8 pixels au coin haut à gauche, avant d'appliquer la transformée en cosinus discrète (DCT) classique. En fait, les pixels de chaque ligne sont tout d'abord décalés à gauche, le vide crée par ce décalage est remplacé par des éléments nuls. Les lignes qui en résultent sont ensuite transformées à l'aide de la DCT unidimensionnelle. Les colonnes résultantes sont à leurs tours déplacés en haut, leurs emplacements initiaux sont remplacés par des éléments nuls, une transformation en cosinus discrète verticale est ensuite appliquée[68, 24].

Cette technique se caractérise par une grande faculté de concentration de l'énergie des blocs de l'image en un seul coefficient appelé le coefficient DC (Direct Coefficient). Elle est également très adaptée aux applications d'indexation et de recherche de l'image, du fait qu'elle permet de garder le contenu de l'image sans modification. Cependant, la principale limitation provient de la transformation ba-

sée bloc, qui engendre une caractérisation très localisée du contenu de l'image. En outre, cette transformation cause une significative perte des corrélations entre les pixels des différents blocs.

Dans la même optique, J. Muller et al utilisent dans [91] le même principe, en remplaçant la transformée en cosinus discrète par la transformée en ondelettes (DWT). Les pixels qui n'appartiennent pas à la région d'intérêt sont tout d'abord remplacés par des éléments nuls. Les lignes sont ensuite transformées à l'aide de la transformée en ondelettes classique, les colonnes ainsi obtenues sont à leur tour transformées à l'aide de la DWT. Cette méthode ne permet pas de garder les propriétés statistiques des blocs de l'image sans modification. Elle provoque même la création d'un effet de flou dans le voisinage des contours des régions de l'image, augmentant ainsi le nombre de pixels à transformer.

Une alternative est de se passer de la décomposition de l'image en blocs, et de transformer les régions directement sans décomposition. Plus récemment, une famille spécifique d'ondelettes appelée les Curvelettes a été proposée, afin de combler les insuffisances de la transformée en ondelettes. Les curvelettes sont des fonctions qui s'adaptent bien aux contours des formes, et spécialement les courbes. Elles offrent même la possibilité de capturer l'information de contour à différentes échelles et orientations[145]. Le seul inconvénient provient toujours du fait que cette transformée ne peut être appliquée qu'aux régions rectangulaires. Il est ainsi indispensable de passer par des opérations d'extrapolation et d'extension des régions d'intérêt.

Réalisant que la transformée en ondelettes préserve les relations spatiales entre ses coefficients, Egger avait proposé dans [43, 42], des algorithmes de transformée en ondelettes qui s'adaptent facilement aux formes arbitraires des régions d'intérêt. Le fonctionnement de cette technique est basé sur son adaptation à la longueur des segments. Quand il est question de transformer un segment de longueur impaire, la transformée en ondelettes s'applique en premier lieu sur la longueur paire du segment, le pixel restant est transformé indépendamment et il est placé dans la sous bande basse fréquence des coefficients en ondelettes. Ceci provoque automatiquement une certaine discontinuité dans la décomposition en ondelettes au deuxième niveau.

Une nouvelle approche qui permet de préserver le nombre de pixels a été proposée par Li dans[76]. Cette approche offre également l'avantage de maintenir les caractéristiques de la corrélation spatiale entre les différentes sous bandes de la transformation en ondelettes. Il est clair sur la figure3.3, que la transformée en on-

delettes basée région donne de très bons résultats par rapport aux autres techniques de transformation basées région. La courbe correspondant au SA-DCT est obtenue à l'aide du standard MPEG 4 version 7, tandis que les résultats obtenus à l'aide des schémas de codage basés ondelettes, sont obtenus à l'aide du standard MPEG 4 utilisant la technique du codage ZTE.

En terme de PSNR, nous pouvons voir clairement sur la figure 3.3, que la transformée en ondelettes basée région enregistre une amélioration de 1,5 jusqu'à 2 dB par rapport aux résultats enregistrés par le SA-DCT, et une amélioration de 2 jusqu'à 2,5 dB par rapport aux résultats enregistrés par la méthode de Egger et al.

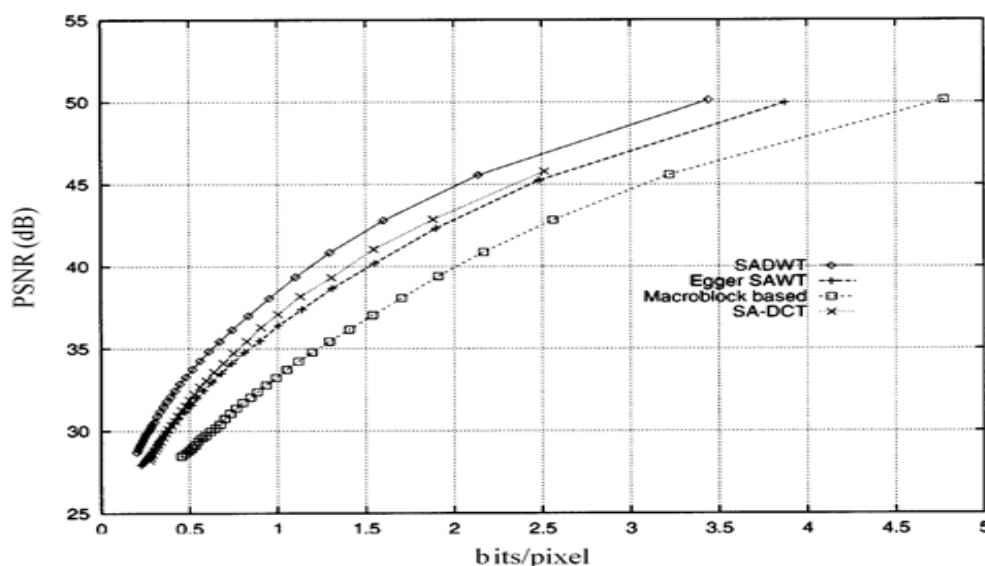


FIGURE 3.3 – Comparaison de différentes techniques de la transformation basée région en terme de PSNR [76].

Une deuxième comparaison entre les différentes approches de transformation basée région, est présentée sur la figure 3.4. En effet, grâce à cette comparaison nous nous rendons bien compte, que la transformée en ondelette basée région offre la meilleure qualité visuelle de l'objet décodé. En fait, nous pouvons facilement observer l'effet de décomposition en blocs sur la figure (c), qui a été transformée à l'aide du SA-DCT. L'utilisation de la méthode de Egger produit un bruit significatif sur la figure (e). Nous entendons par là, que l'utilisation de la transformée en ondelettes basée région produit un taux d'erreur faible tout en gardant l'aspect visuel de la région transformée presque intact.

La transformée en ondelettes basée région consiste à appliquer la DWT unidi-

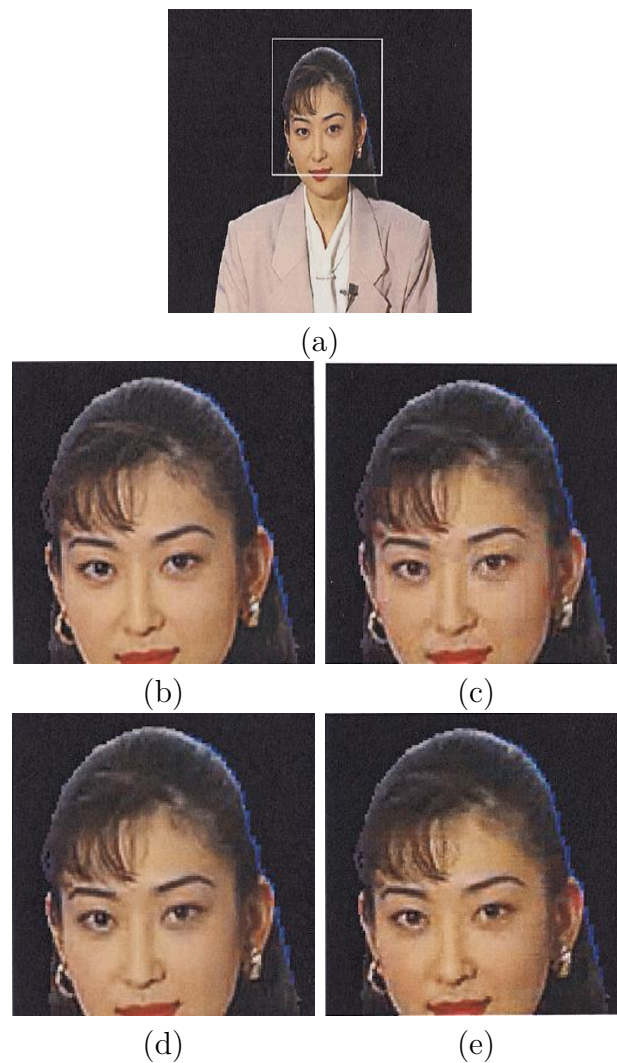


FIGURE 3.4 – *Comparaison entre les différentes méthodes de la transformation basée région. (a) représente l'image originale, le cadre en blanc définit la région d'intérêt à transformer. (b) représente la région d'intérêt avant transformation. L'image (c) représente la région d'intérêt après décodage, dans le (c) la transformée utilisée est le SA-DCT, (d) l'image est reconstruite à l'aide du SA-DWT, (e) le résultat visuel du codage à l'aide de la méthode de Egger [76]*

mensionnelle sur les lignes de la région d'intérêt segment par segment, et de répéter cette opération sur les colonnes de la région d'intérêt. Aucune opération de pré-traitement n'est effectuée sur les lignes et les colonnes de la région d'intérêt, ce qui permet de préserver leur caractéristiques. La transformée en ondelettes consiste à :

- Étendre les segments à transformer.
- Convoluer les segments avec les filtres passe bas et passe haut.
- Sous échantillonner les segments.

Ces étapes feront l'objet d'une étude détaillée, dans les paragraphes suivants. L'approche utilisée dans notre thèse est celle du travail[76].

3.2.2.1 *Étension des bordures de segments*

Étant donné que les signaux sont des grandeurs finies, un effet de bord fait son apparition lorsqu'on convolue les filtres avec les bordures de la région d'intérêt. Plusieurs solutions ont été présentées, certains ont proposé de prolonger le signal, en ajoutant des zéros sur les bordures du signal figure 3.5. Cela engendre généralement des discontinuités du signal. Une autre solution consiste à étendre le signal en ajoutant des valeurs dérivées des pixels internes au segment. Dans ce contexte deux types d'extension ont été proposées : L'extension symétrique figure 3.5 et l'extension périodique.

Les extensions aux bordures du signal doivent satisfaire certaines conditions. Afin d'éviter l'effet d'aliasing sur les bordures, les valeurs du signal aux extrémités doivent être symétriques. Autrement dit les pixels sur les extrémités du segment doivent être symétriques, par rapport à son contour.

La figure 3.5 illustre les différents types de l'extension d'un segment x , l'extension périodique consiste à ajouter $L/2$ valeurs de l'extrémité droite du segment aux bordures gauche. Et d'ajouter le même nombre de valeurs de l'extrémité gauche aux bordures droites. Le type 1 de l'extension symétrique applique une symétrie du segment x par rapport à son extrémité gauche ou droite. Le point de symétrie dans le deuxième type correspond à la dernière valeur de chaque côté. Le principe du troisième type est le même, tout en inversant les signes des valeurs à ajouter. Si le segment est plus court que la longueur des filtres, l'extension consiste à répéter les valeurs du segment $L/2$ fois aux extrémités gauche et droite 3.6.

3.2.2.2 *Convolution des segments avec le banc de filtres*

La difficulté rencontrée par la transformée en ondelettes basée région, vient du fait qu'elle traite souvent des segments de longueurs et de positions impaires. En fait, le sous échantillonnage de ce type de segment ne produit pas une répartition équitable des coefficients entre les différentes sous bandes. Dans la section ci-dessous, nous expliquons la solution proposée par Li et al dans[76]. Celle-ci permet la meilleure synthèse de l'image en utilisant une extension symétrique de type 2. Ce type d'extension permet la préservation de la corrélation à l'intérieur et sur les

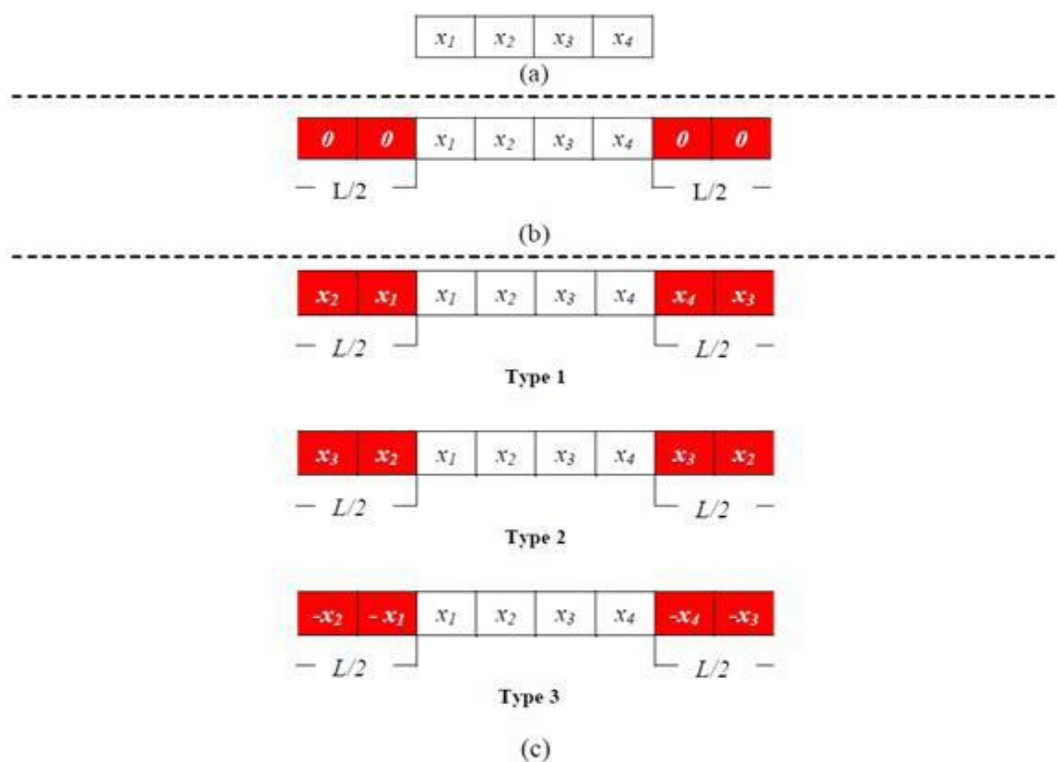


FIGURE 3.5 – Exemple d’extension d’un segment du signal x , (a) représente le segment à transformer. (b) représente l’utilisation de l’extension par l’ajout de zéros aux bordures du segment x , (c) illustre les trois types de l’extension symétrique. Les filtres de la transformée en ondelettes, qui sont utilisés dans cet exemple sont de longueurs L .

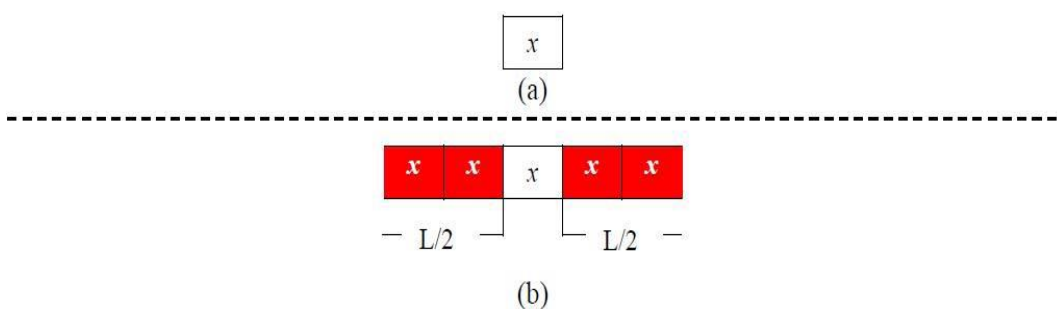


FIGURE 3.6 – Exemple d’extension d’un segment de courte longueur, (a) représente le segment à transformer. (b) représente le segment après extension.

bordures du segment sans modification. Nous pouvons ainsi éviter l’effet de transition brusque dans les contours des formes. Dans ce travail, nous utilisons les filtres de la transformée en ondelettes de Debauchies (voir annexe 7.7).

Prenons l’exemple d’un signal x , tel que : $x = x_0, x_1, x_2, x_3 \dots x_{N-1}$.

N représente la longueur du signal x . L'analyse de la transformée en ondelettes est supposée être réalisée par les deux filtres : G, Q , tel que : $G = G_0, G_1, G_2, G_3 \dots G_{L_G-1}$. Filtre d'analyse passe bas. $Q = Q_0, Q_1, Q_2, Q_3 \dots Q_{L_Q-1}$. Filtre d'analyse passe haut. L_G : représente la longueur du filtre passe bas. L_Q : représente la longueur du filtre passe haut. Les convolutions entre le segment et les filtres passe bas et passe haut, sont données par les formules suivantes :

$$x_G^*(u) = \sum_{v=1}^{L_G/2} G(u + v - L/2 + 1)x(L_G - 1 - v) \quad (3.1)$$

$$x_Q^*(u) = \sum_{v=1}^{L_Q/2} Q(u + v - L/2 + 1)x(L_Q - 1 - v) \quad (3.2)$$

R_G^*, R_Q^* représentent les réponses impulsionnelles passe bas et passe haut des filtres.

3.2.2.3 Sous échantillonnage

Le sous échantillonnage et la répartition des coefficients entre les différentes sous bandes, doivent respecter les règles suivantes :

- 1) Si la longueur du segment N est paire.
 - si la position du segment p est impaire : $N/2$ coefficients passe bas sont placés dans la position $p/2+1$.
 $N/2$ coefficients passe haut sont placés dans la position $p/2$.
 - si p est paire : $N/2$ coefficients passe bas sont placés dans la position $p/2$.
 $N/2$ coefficients passe haut sont placés dans la position $p/2$.
- 2) Si la longueur du segment N est impaire.
 - si p est impaire : $N/2$ coefficients passe bas sont placés dans la position $p/2+1$.
 $N/2+1$ coefficients passe haut sont placés dans la position $p/2$.
 - si p est paire : $N/2+1$ coefficients passe bas sont placés dans la position $p/2$.
 $N/2$ coefficients passe haut sont placés dans la position $p/2$.

Afin d'illustrer le principe de la transformée en ondelettes basée région, nous représentons sur la figure 3.7, l'application de cette transformée sur deux images. Seules les régions d'intérêt encadrées par des rectangles rouge, qui ont été transformées. Les résultats de leur transformations sont colorées en blanc. La zone noire représente la partie non transformée, qui correspond au fond de l'image.

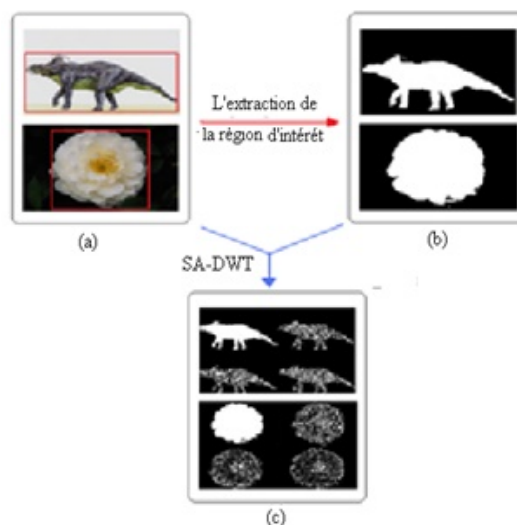


FIGURE 3.7 – Exemple de l'application de la transformée en ondelettes basée région (SA-DWT) sur deux régions d'intérêt de deux images différentes. Les régions d'intérêt obtenu grâce à un processus de segmentation sont encadrées par des rectangles rouge.

3.3 Caractérisation de l'image par la transformée en ondelettes basée région

Les vues 2D d'un modèle tridimensionnelle sont généralement déformées, suite à l'occlusion de l'objet par d'autres objets de la scène, ou à cause de la variation des conditions de capture. Celui-ci rend la caractérisation des formes par leurs attributs de contour insuffisante. L'ajout de la caractérisation par la primitive couleur et texture, qui sont robustes aux occlusions et aux déformations des régions d'intérêt ; permet d'améliorer les performances des systèmes d'indexation de modèles 3D.

La transformation en ondelettes n'étant pas capable de capturer l'information relative aux contours, il est indispensable de compléter la description des attributs de texture et de couleur par une description des orientations du contour. Dans notre thèse, nous proposons de caractériser les contours des régions d'intérêt par leur patterns des extrêmes locales (DLEP). nous introduisons également une caractérisation de la texture par les patterns issus des sous bandes de détails de la transformée en ondelettes. La caractéristique la couleur est issue du coefficient d'approximation de la transformée en ondelettes.

3.3.1 Caractérisation des attributs

La discrimination entre les caractéristiques de différentes images, est un critère primordial de n'importe quel système d'indexation. Étant donné que la transformée en ondelettes est basée sur la décomposition spatiale et spectrale de l'image à plusieurs niveaux ; l'utilisation des patterns (associations) de coefficients issus de différents niveaux, permet de garder la corrélation entre les sous bandes de la transformée en ondelettes. Chaque pattern est constitué des valeurs d'une caractéristique donnée appartenant à différentes sous bandes. Les patterns de la texture sont constitués par les coefficients de détail, issues des trois sous bandes LH , HL et HH . Les patterns de la caractéristique couleur sont composés des coefficients d'approximation des trois composantes Y , C_b et C_r . Les patterns de la caractéristique contour caractérisent les directions de la variation des coefficients dans la sous bande d'approximation.

3.3.2 Regroupement des vecteurs de caractéristiques

Contrairement à la représentation classique des histogrammes de patterns, le principe de la représentation parcimonieuse est de projeter l'histogramme sur un dictionnaire creux (le nombre d'éléments nuls est supérieur à la dimension de l'histogramme lui même), d'où l'appellation représentation parcimonieuse.

La représentation parcimonieuse sous forme d'histogramme d'une matrice de données X , est définie par la formule suivante :

$$h_i = \sum_{i=1} NC_{ij} \quad (3.3)$$

C_i représente une ligne de la matrice de données.

h_j représente le $j^{\text{ème}}$ bin de l'histogramme, où $j = 1, 2, 3, \dots, k$.

Autrement dit les valeurs des bins représentent le poids total des vecteurs de base, qui correspondent à ces bins dans la représentation parcimonieuse de la matrice X .

3.3.2.1 Construction des patterns de texture

L'application de la transformée en ondelettes basée région sur un segment donné, permet de répartir le détail de chaque pixel suivant les trois directions : horizontale, verticale et diagonale. Autrement dit, le détail de chaque pixel est représenté par les coefficients des trois sous bandes LH , HL et HH . Ceci dit, la caractérisation de la texture de chaque segment nécessite la combinaison des coefficients de ces trois sous bandes. Comme il est illustré sur la figure 3.8 le pattern de texture est constitué des coefficients hautes fréquences, situés tous au même emplacement dans les trois sous bandes hautes fréquence : LL , LH , HL . Cependant, la répartition des coeffi-

cients n'est pas toujours équitable entre les différentes sous bandes de la transformée en ondelettes. La décomposition d'un segment de longueur impaire produit plus de coefficients hautes fréquences, que de coefficients basses fréquences, ou le contraire. Dans ce contexte, nous cherchons à construire des patterns de texture de même longueur, égale à trois. Ces contraintes, nous incitent à compléter les emplacements vides des coefficients hauts fréquence par des éléments nuls.

La figure 3.8 montre la construction de quelque patterns de texture d'une région d'intérêt représentant un dinosaure. Dans cet exemple, nous avons pris quelques échantillons de pixels composant la région d'intérêt, et nous avons illustré l'influence de leur transformation sur le processus de la caractérisation de la texture. Afin de bien illustrer ces échantillons, on a simulé la région d'intérêt par un rectangle de carreaux bleus, la composante LH par un rectangle de carreaux rouges, la composante HL par un rectangle de carreaux verts, et la composante HH par un rectangle de carreaux jaunes.

La première ligne de la région d'intérêt par exemple contient 4 pixels (colorés bleu), sa transformation par la transformée en ondelettes basée région produit 2 coefficients (en couleur rouge) dans la sous bande LH , 0 coefficients (verts) dans la sous bande HL et 0 coefficients (jaunes) dans la sous bande HH . Donc le premier pattern de texture qui doit contenir les valeurs du premier coefficient de la transformée en ondelettes, ne contient qu'un seul coefficient de la sous bande LH . Les deux autres coefficients sont remplacés par des zéros. La même méthode est utilisée pour la construction des patterns de texture qui en reste.

3.3.2.2 Construction de l'histogramme de texture

L'histogramme de texture est la combinaison des histogrammes de texture de la région 1 (histogramme de la texture du fond), et celui de la région 2 (histogramme de la texture de l'objet) :

L'histogramme de la texture noté H_T de l'image est donné par :

$$H_T = [\omega_{pT} * H_{T_{r1}} \quad (1 - \omega_{pT}) * H_{T_{r2}}] \quad (3.4)$$

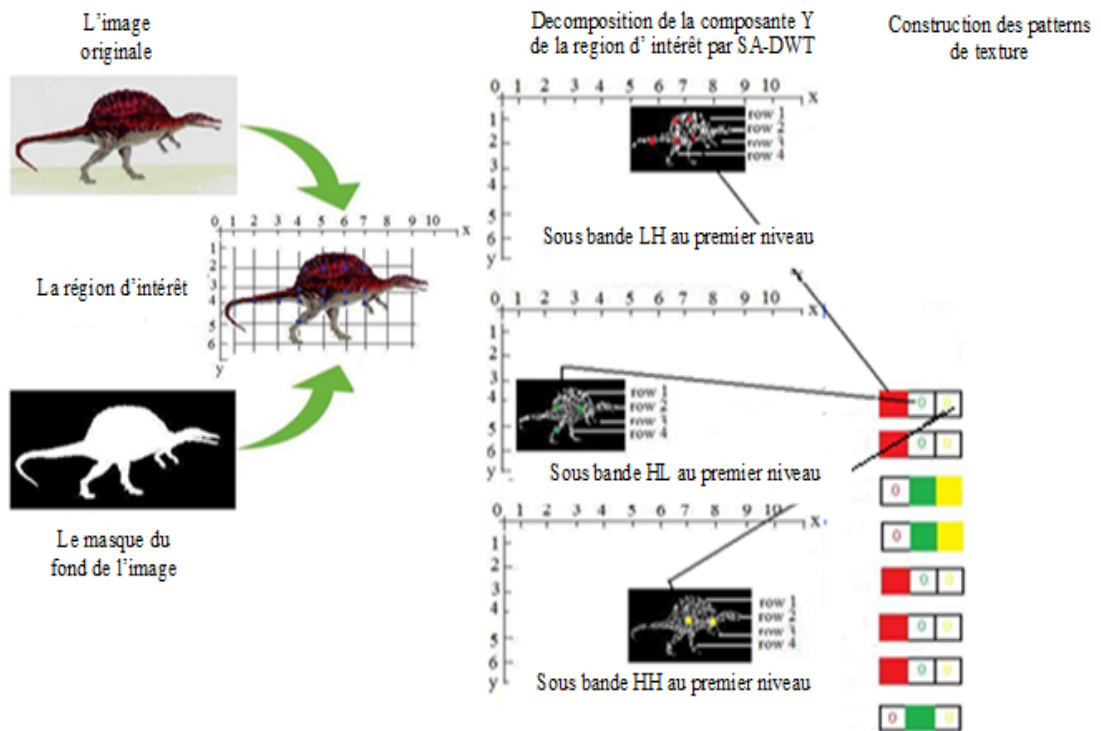
Où

L'histogramme du fond de l'objet $H_{T_{r1}}$ est donné par :

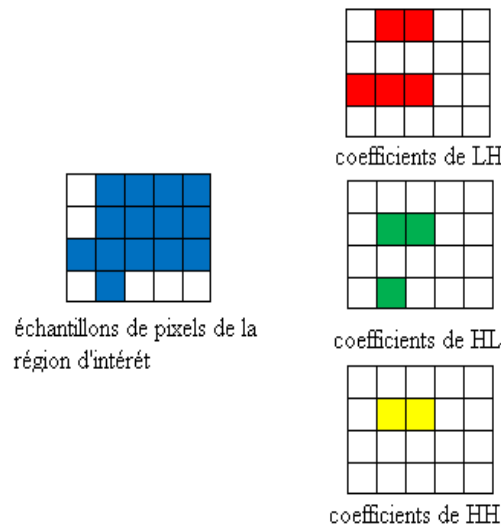
$$H_{T_{r1}} = [H_{T_{r1}^1} \ H_{T_{r1}^2} \ H_{T_{r1}^3}] \quad (3.5)$$

$H_{T_{r1}^i}$ représente l'histogramme de texture du fond, au niveau de décomposition i de la transformée en ondelettes.

3.3. CARACTÉRISATION DE L'IMAGE PAR LA TRANSFORMÉE EN ONDELETTES BASÉE RÉGION



(a)



(b)

FIGURE 3.8 – Exemples de la construction des patterns de texture d'une région d'intérêt. (a) représente les étapes de la construction des patterns de texture. (b) simule la région d'intérêt par un rectangle de carreaux bleus et les composantes des sous bandes par des rectangles avec d'autres couleurs.

L'histogramme du fond $H_{T_{r_1}^2}$ est donné par :

$$H_{T_{r_2}} = [H_{T_{r_2}^1} H_{T_{r_2}^2} H_{T_{r_2}^3}] \quad (3.6)$$

$H_{T_{r_2}^i}$ représente l'histogramme de texture de l'objet, au niveau de décomposition i de la transformée en ondelettes.

ω_{pT} représente le coefficient de pondération des régions de l'image.

L'algorithme ci-dessous introduit la méthode utilisée pour la construction de l'histogramme de la texture d'une région quelconque de l'image. Pour des raisons de simplification, cet algorithme décrit la construction de l'histogramme de la texture au premier niveau de la décomposition. La construction des histogrammes au deuxième et troisième niveau se fait de la même façon.

Algorithme de la construction de l'histogramme de texture

Algorithme 3.1

Arguments d'entrée

les sous bandes hautes fréquence de la transformée en ondelettes (le résultat de la décomposition en ondelettes au niveau 1 : LH , HL , HH).

Étape 1 : Construction des patterns de texture.

Chaque pattern est construit à partir de trois coefficients de détail, qui sont issus de la même position dans les trois sous bandes.

Étape 2 : Construction des histogrammes de référence

Pour chaque classe d'images, tous les patterns doivent être représentés dans un seul histogramme, appelé histogramme de référence (His_{ref}). Ces patterns sont ensuite regroupés à l'aide de la fonction `vlkmeans` [5].

Étape 3 : Construction de l'histogramme de texture

La représentation sous forme parcimonieuse de l'histogrammes de la texture. Cet histogramme (indice) représente l'évolution de l'occurrence des patterns de la texture. Cette étape est un processus itératif de recherche des patterns obtenus dans l'histogramme de référence. À chaque itération, l'histogramme de référence définit l'emplacement du pattern de la texture et l'occurrence (le nombre de bins) de ce pattern est incrémentée.

Sorties : l'histogramme de la texture de la région $H_{T_{r_i}}$ au premier niveau de décomposition

Les histogrammes de texture montrés sur la figure 3.9 sont le résultat de la concaténation des histogrammes de texture dans les directions horizontal, vertical et diagonal. Nous pouvons remarquer que les histogrammes des régions d'intérêt de la

classe (éléphant) des images similaires (a, b) sont différents de l'histogramme de la région d'intérêt (montagne). Cependant, nous pouvons également remarquer que les régions de fond partagent tous le même histogramme de texture. L'attribut qui permet de partager entre ces régions de fond, est la distribution des couleurs de la région ciel entre les trois images. C'est pour cette raison qu'on a introduit l'utilisation de l'histogramme de couleur.

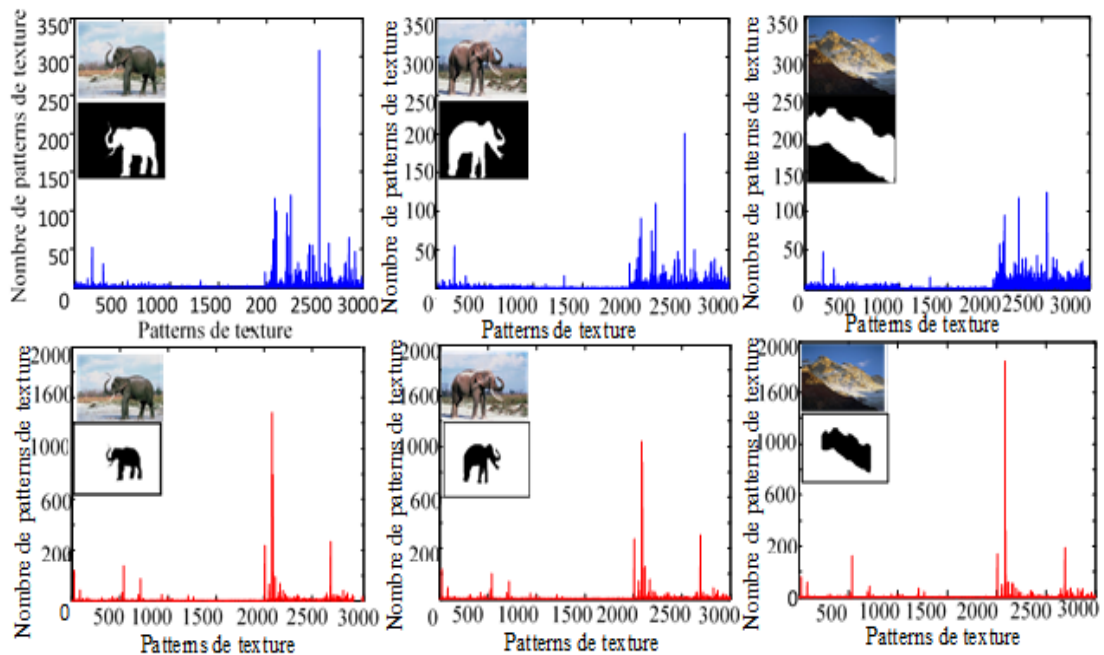


FIGURE 3.9 – Exemples d'histogrammes de texture. (a), (b) et (c) représentent les histogrammes des régions d'intérêt en blanc (éléphant) et (montagne). (d), (e) et (f) représentent les histogrammes des fonds des images.

3.3.2.3 Construction des patterns de couleur

La couleur dans les images RGB peut être caractérisée par les patterns des coefficients DWT des composantes Y , C_b , C_r . Étant donné que la sous bande d'approximation LL de la transformée en ondelettes représente significativement l'image en niveaux de gris, nous proposons de construire les patterns de la couleur à partir des coefficients LL des trois composantes. La figure 3.10 montre un exemple de la construction des patterns de la couleur de la région d'intérêt (Éléphant).

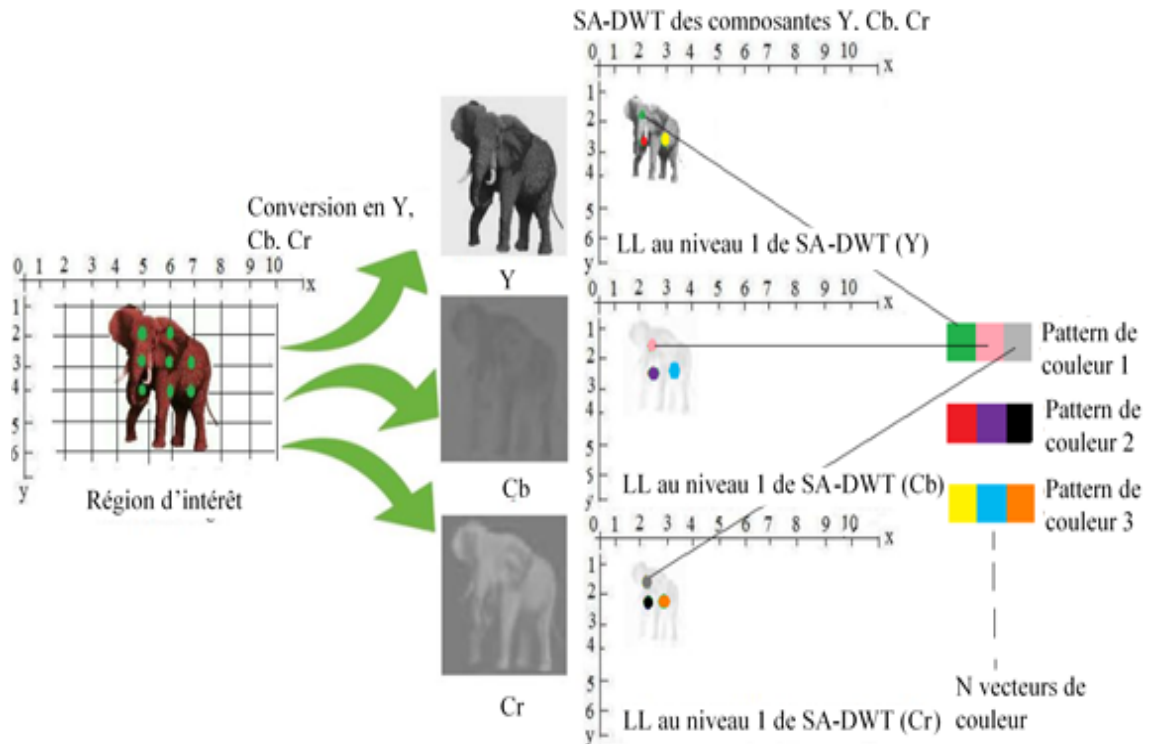


FIGURE 3.10 – Construction des patterns de couleur en utilisant la transformée en ondelettes basée région. Les points verts représentent des pixels pris arbitrairement de la région d'intérêt. N représente le nombre de coefficients des ondelettes dans la sous bande LL.

3.3.2.4 Construction de l'histogramme de couleur

L'histogramme de la couleur est la combinaison des histogrammes de la couleur des différentes régions de l'image :

$$H_c = [\omega_{pc} * H_{LLr1} \quad (1 - \omega_{pc}) * H_{LLr2}] \quad (3.7)$$

ω_{pc} représente le coefficient de pondération de l'attribut couleur.

H_{LLr1} représente l'histogramme de couleur de la région 1 (l'objet).

H_{LLr2} représente l'histogramme de couleur de la région 2 (le fond de l'image).

L'algorithme ci-dessous introduit la méthode utilisée pour la construction de l'histogramme de couleur.

Algorithme de la construction de l'histogramme de couleur

Algorithme 3.2

Arguments d'entrée

les sous bandes basses fréquence de la transformée en ondelettes (le résultat de la

décomposition en ondelettes au niveau 3 : LL_Y, LL_{C_b}, LL_{C_r}).

Étape 1 : Construction des patterns de couleur.

Chaque pattern est construit des trois coefficients d'approximation issus de la même position dans les trois composantes Y, C_b, C_r .

Étape 2 : Construction des histogrammes de référence

Les patterns doivent tout d'abord être classés dans un seul histogramme, appelé l'histogramme de référence (His_{ref}). Ces patterns sont regroupés à l'aide de la fonction `vlkmeans` [5].

Étape 3 : Construction de l'histogramme de couleur

Les patterns de couleur constituent un ensemble à redondance très élevée. La construction de l'histogramme de couleur consiste à calculer l'occurrence de chaque pattern de l'histogramme de référence dans cet ensemble.

Sorties : l'histogramme de couleur $H_{C_{ri}}$

La figure 3.11 montre quelques histogrammes de couleur des régions d'intérêt et de fond des images exemples. Nous pouvons remarquer que l'histogramme de couleur permet de différencier entre les régions du fond similaires (ciel), Ceci dit l'utilisation de l'histogramme de couleur permet d'améliorer les performances de l'algorithme de caractérisation de l'image. Cette caractérisation peut être améliorée davantage en intégrant une description du contour des régions.

3.3.2.5 Construction des patterns de contour

Les caractéristiques couleur et texture ne permettent pas de capturer l'information de la frontière contour des régions d'intérêt, surtout dans le cas des régions similaires. À cet effet, nous proposons d'introduire des descripteurs de contours. Ce processus est basé sur l'algorithme DLEP décrit dans le papier [123]. Les patterns du contour sont extraits dans la direction $0^\circ, 45^\circ, 90^\circ$ et 135° . La construction des patterns du contour consiste à décomposer la sous bande LL de la composante Y en blocs de 3×3 coefficients. Chaque coefficient du contour est caractérisé par un pattern de 4 valeurs. ces valeurs correspondent aux directions $0^\circ, 45^\circ, 90^\circ$ et 135° . Chaque coefficient du contour de la forme est ainsi caractérisé par un pattern de 4 coefficients.

La figure 3.12 illustre la construction du premier coefficient du pattern DLEP dans la direction 0° (noté $DLEP^0$). Dans ce cas, la composante Y de l'image est tout d'abord transformée à l'aide de la SADWT. La composante LL au troisième

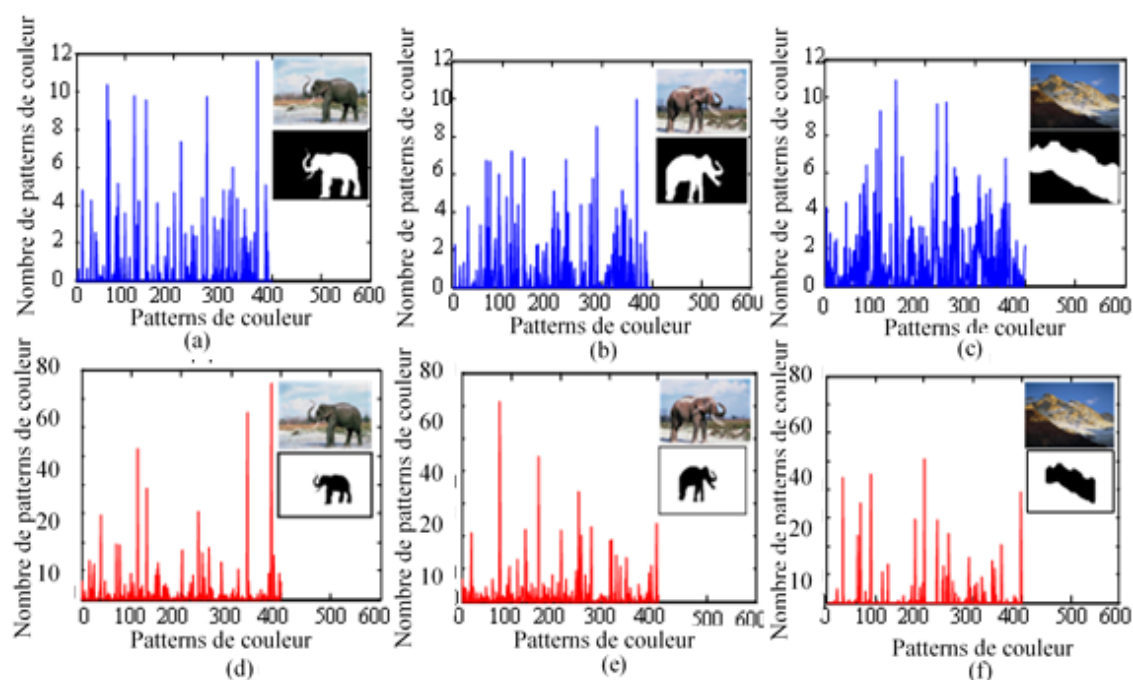


FIGURE 3.11 – Exemples d’histogrammes de couleur. (a), (b) et (c) représentent les histogrammes des régions d’intérêt des objets. (d), (e) et (f) représentent les histogrammes de couleur du fond de l’image.

niveau de la transformation est décomposée en blocs de 3×3 coefficients. À titre d’exemple, nous nous intéressons au codage du coefficient au centre du bloc encadré en rouge. Cet exemple montre que les coefficients en dehors de la région d’intérêt sont remplacés par des zéros. Le calcul du $DLEP^0$ consiste à :

1. coder les coefficients du bloc. Chaque coefficient est codé par un 0 si les variations à ses extrémités sont opposées. Sinon il est codé par un 1. À cet effet, on aura une séquence de neuf bits. Dans cet exemple cette séquence est égale à 000100110.
2. Convertir la représentation binaire de la séquence obtenue au format décimale. Dans cet exemple, le résultat est égal à 38. Donc $DLEP^0$ de ce premier coefficient est égal à 38.

3.3.2.6 Construction de l’histogramme de contour

L’histogramme de contour est la combinaison pondérée des histogrammes de contour des différentes régions de l’image :

$$H_E = [\omega_{pE} * H_{E_{r1}} + (1 - \omega_{pE}) * H_{E_{r2}}] \quad (3.8)$$

3.3. CARACTÉRISATION DE L'IMAGE PAR LA TRANSFORMÉE EN ONDELETTES BASÉE RÉGION

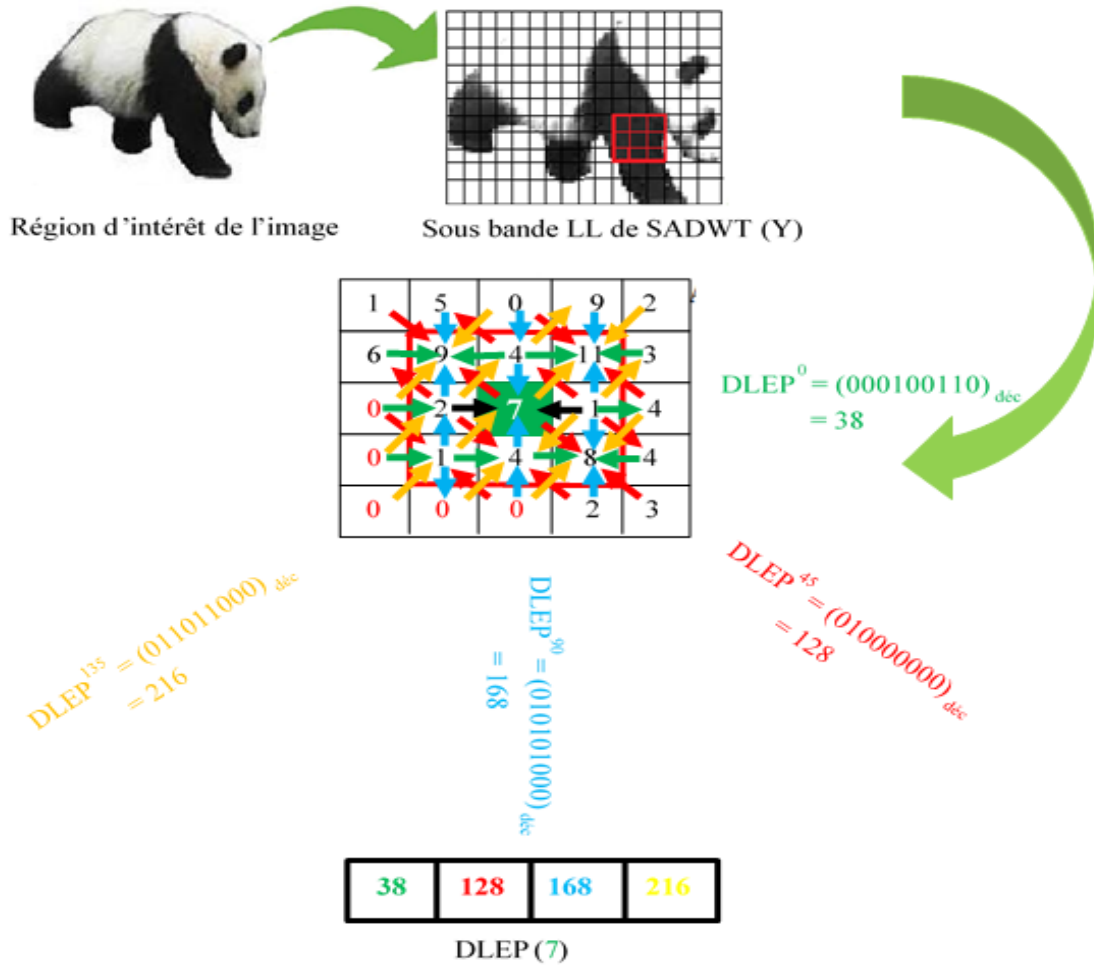


FIGURE 3.12 – Exemple de la construction d'un pattern de contour, le bloc à caractériser en rouge est choisi arbitrairement.

Ou

$$H_{E_{r_1}} = [H_{0_{r_1}} H_{45_{r_1}} H_{90_{r_1}} H_{135_{r_1}}] \quad (3.9)$$

$$H_{E_{r_2}} = [H_{0_{r_2}} H_{45_{r_2}} H_{90_{r_2}} H_{135_{r_2}}] \quad (3.10)$$

ω_{pE} Représente le poids du paramètre contour..

$H_{0_{r_1}}$ Représente l'histogramme de contour dans la direction 0° de la région r_i

$H_{45_{r_1}}$ Représente l'histogramme de contour dans la direction 45° de la région r_i

$H_{90_{r_1}}$ Représente l'histogramme de contour dans la direction 90° de la région r_i

$H_{135_{r_1}}$ Représente l'histogramme de contour dans la direction 135° de la région r_i

Algorithme de la construction de l'histogramme de contour

Algorithme 3.3

Arguments d'entrée

les sous bandes basses fréquence de la transformée en ondelettes (le résultat de la décomposition en ondelettes au niveau 3 : LL).

Étape 1 : la construction des patterns de contour.

Chaque pattern est construit des trois coefficients d'approximation issus de la même position dans les trois composantes Y, C_b, C_r .

Étape 2 : la construction des histogrammes de référence

Pour chaque classe d'image, tout les patterns doivent être représentés dans un seul histogramme, appelé histogramme de référence (Hisref). Ces patterns sont ensuite regroupés à l'aide de la fonction vl_kmeans .

Étape 3 : la présentation sous forme parcimonieuse de l'historgramme du contour.

Sorties : l'historgramme de contour de la région $iH_{C_{ei}}$

Cet algorithme est ensuite appliqué dans les autres directions $45^\circ, 90^\circ$ et 135° . Afin d'analyser l'impact de l'utilisation de l'historgramme de contour sur la discrimination entre les régions différentes, nous avons construit les histogrammes de contour pour chaque région de deux images d'éléphant, et une image de montagne. Dans la

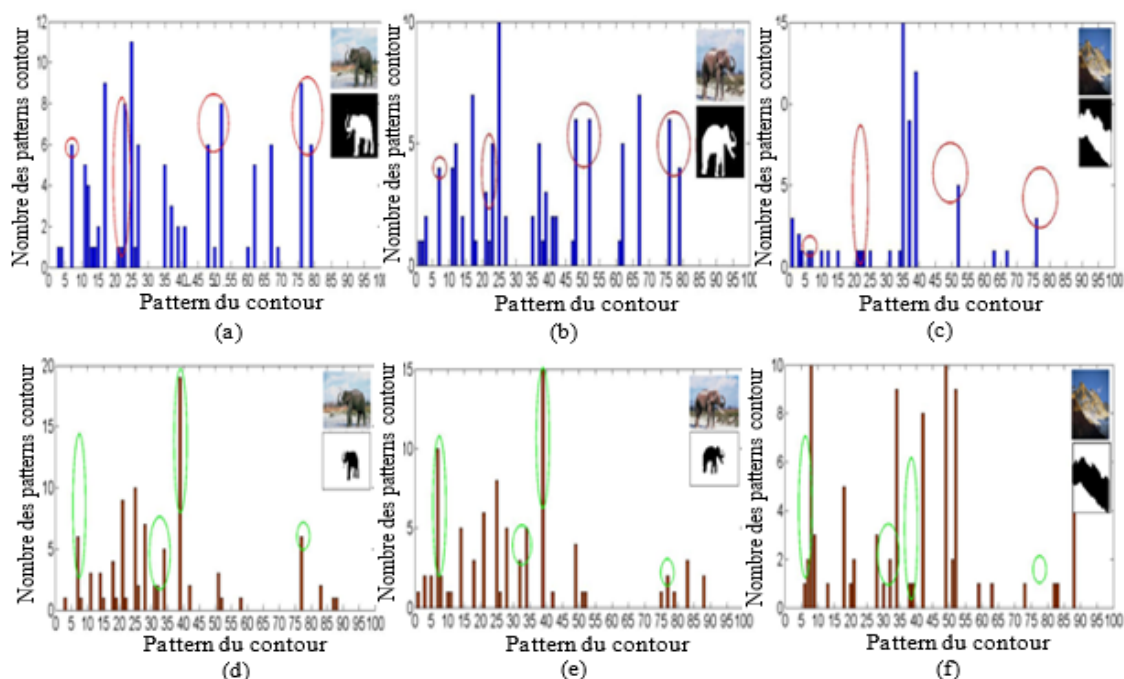


FIGURE 3.13 – Exemples d'historgrammes de contour. (a), (b) et (c) représentent les histogrammes des régions d'intérêt des objets. (d), (e) et (f) représentent les histogrammes de contour du fond de l'image.

figure 3.13, il est clair que les objets similaires (éléphant) partagent tous les mêmes

patterns de contour, qui sont entourés d'un cercle rouge. Ces patterns n'apparaissent pas dans l'histogramme de l'objet différent montagne.

D'autre part, les précédents attributs (la couleur, et la texture), ont prouvé de grandes difficultés pour différencier entre la région du fond, qui est identique dans les trois images. Cette région qui est le ciel possède les mêmes caractéristiques : couleur et texture, mais comme on peut le remarquer elles n'ont pas la même forme et les mêmes orientations aux bordures de la région. À cet effet, nous ajoutons l'histogramme de contour pour bien comparer ces différentes régions. Dans cet exemple, nous pouvons observer que les patterns entourés de cercles verts n'apparaissent que dans les histogrammes des fonds des images similaires, dans ce cas il s'agit du fond des images d'éléphant.

3.4 Conclusion

Dans ce chapitre nous avons présenté notre principale contribution, qui porte sur l'indexation des vues 2D. la caractérisation des projections 2D d'un modèle tridimensionnel, est effectué après des opérations de prétraitement, permettant la projection du modèle 3D, son alignement et sa mise à l'échelle, en vue d'une caractérisation robuste aux variations subies par le modèle 3D. Ces opérations ainsi que le choix des projections représentatives de l'objet seront traités dans le prochain chapitre.

Chapitre 4

Prétraitement Des Modèles Tridimensionnels

Résumé : dans ce chapitre nous nous intéressons au problème du changement de l'emplacement et de la pose des objets 3D. Les conditions d'acquisition des objets 3D n'étant pas identiques, leur prétraitements s'avère très vite nécessaire, afin de normaliser leur centres, échelles et orientations. Nous commençons ce chapitre par un état de l'art des techniques de normalisation et d'estimation de pose des objets 3D. Nous concluons ce chapitre par la présentation de notre approche de normalisation et d'alignement, ainsi que les résultats obtenus.

4.1 Introduction

L'utilisation de différents outils d'acquisition et de numérisation, produit des modèles 3D de différents centres, échelles et orientations. Étant donné que la caractérisation d'un modèle 3D dépend de sa pose et de son échelle ; on doit précéder cette opération par une étape de normalisation [93]. Celle-ci consiste à mettre à l'échelle et à réorienter tous les modèles de la base. L'objectif de la normalisation est d'arriver à des conditions de pose et d'alignement propres à chaque classe d'objets. Cela permet de se passer des changements de l'orientation lors de la phase de la recherche [93]. Deux modèles similaires peuvent ainsi être comparés sans se soucier de leurs conditions de rotation et de pose [93].

Afin de traiter correctement les problèmes posés par la normalisation et l'estimation de la pose, la première partie de ce chapitre est consacrée à la présentation d'un état de l'art non exhaustif des techniques de la normalisation des modèles 3D. Cette étude permet de souligner les avantages et les limites de chaque méthode. La deuxième partie de ce chapitre porte sur les méthodes d'estimation de la pose et de la réorientation des formes tridimensionnelles. À la fin de ce chapitre nous présentons notre méthode de normalisation et quelques exemples.

La première étape de notre méthode porte sur le centrage et la mise à l'échelle des objets 3D. Nous introduisons la notion de la forme sphérique minimale pour la normalisation des objets 3D. Ce choix est imposé par notre système de projection des light field, qui est basé sur cette forme. La deuxième partie est consacrée à l'alignement des modèles 3D à l'aide de l'analyse en composante principale. Cette dernière partie a fait l'objet d'une communication à la 5^{ème} conférence *ICEE International Conference On Electronics Engineering 2013*.

4.2 Schéma du processus de normalisation et d'alignement des objets 3D

Deux opérations coexistent dans le schéma présenté à la figure 4.1 , la normalisation et l'alignement des objets 3D. La première opération est basée sur :

- Le calcul du centre de gravité.
- La construction de la sphère minimale englobante.

Ces deux opérations sont ensuite utilisées pour centrer et mettre à l'échelle l'objet 3D. Lors de la phase de l'alignement, on s'intéresse au calcul des vecteurs unitaires du repère 3D (xyz) , ceci grâce à l'application de l'analyse en composantes principales sur l'ensemble des objets appartenant à la même classe. L'alignement consiste ensuite

à réorienter l'objet 3D, en se basant sur les signes de ses parties dans le nouveau repère.

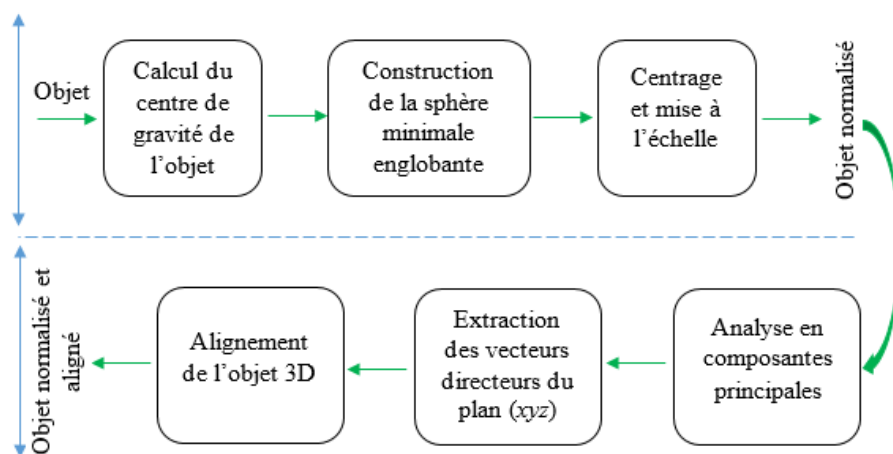


FIGURE 4.1 – Schéma général du processus de normalisation et d'alignement des objets 3D.

4.3 État de l'art des méthodes de normalisation

La normalisation est une étape indispensable dans le processus d'indexation et de recherche des objets 3D. Elle est également connue au milieu des spécialistes de l'indexation et de la recherche d'images, comme étant une opération extrêmement complexe [50, 103]. La difficulté de cette opération, réside dans le calcul des centres et des échelles des milliers d'objets. Ceux-ci étant capturés de différentes manières, leur positions et orientations changent considérablement d'un modèle à l'autre. Les recherches dans ce domaine se sont donc focalisées sur la définition du centre et de l'échelle de chaque objet, pour pouvoir ensuite le déplacer au centre d'un repère commun à tous les modèles de la base [47][141].

L'analyse des différentes techniques de la normalisation, montre qu'il existe de nombreuses méthodes de calcul des centres et des échelles des formes. Dans ce cadre nous avons choisi de séparer les méthodes de centrage des modèles 3D, des méthodes de la mise à l'échelle. Concernant la première question, une première solution consiste à considérer le centre de gravité de l'objet comme étant l'origine de son repère [93] :

$$C_g = \frac{1}{n} \sum_{i=1}^n S_i \quad (4.1)$$

S_i représente le $i^{\text{ème}}$ sommet de l'objet 3D.

C_g représente le centre de gravité de l'objet 3D.
 n représente le nombre de sommets de l'objet.

Dans [23], le centre de l'objet est le centre de gravité des sommets de son maillage. La figure 4.2 illustre l'utilisation des centres de gravité pour le centrage et la mise à l'échelle des formes. Considérons par exemple les visages de la figure 4.2 (a) ; leur centrage et leur mise à l'échelle nécessite la définition de leur centres respectifs. Dans ce travail, on s'est contenté d'utiliser le centre de gravité comme un centre de l'objet. La figure 4.2 (c) montre clairement le décalage et la désorientation qui existent entre les différents visages. Le centrage de ces visages dans la figure 4.2 (d) permet d'aligner et de mettre à l'échelle les différents visages.

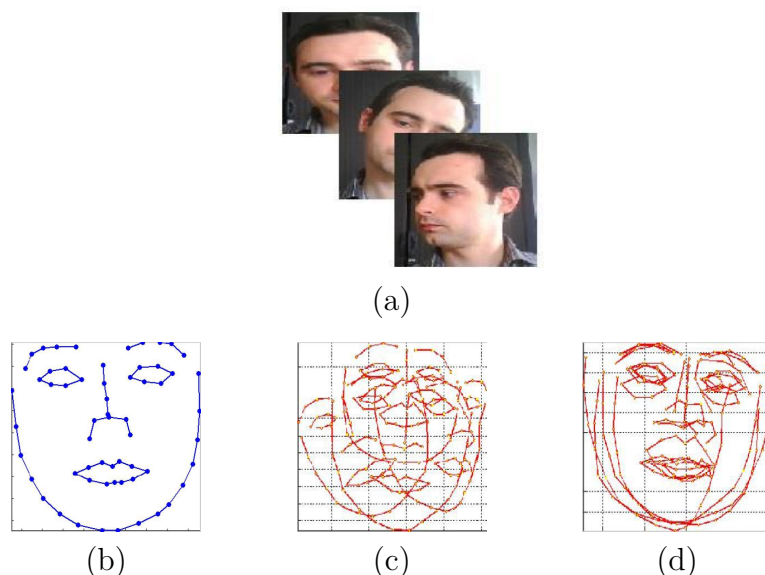


FIGURE 4.2 – Exemple de l'utilisation des centres de gravité pour le centrage et la mise à l'échelle d'un ensemble de visages. (a) représente un ensemble de visages de tailles et de positions différentes. (b) représente le modèle d'annotation utilisé pour définir les coordonnées des points caractéristiques sur le visage. (c) représente le résultat de l'annotation des visages. (d) représente les visages après les opérations de centrage et de mise à l'échelle.

Dans [29], le centre de l'objet correspond à l'intersection entre ses axes majeurs et mineurs. Les auteurs représentent les objets 3D sous la forme d'un ensemble de voxels. Le déplacement des objets au centre commun (la moyenne des centres de tous les objets) est une opération rigide de l'espace R^3 . Autrement dit, cette opération ne doit en aucun cas changer l'apparence de l'objet 3D. Le déplacement de l'objet 3D se traduit par le déplacement des différents voxels de la même distance et dans la même direction. Prenons par exemple ν^1 et ν^2 deux voxels d'un objet donné, $\nu^{1'}$ et $\nu^{2'}$

représentent les deux voxels après le déplacement. Afin que l'objet déplacé garde la même forme, ces deux voxels doivent garder la même distance après le déplacement [29]. Cette méthode du centrage est rapide, et simple à mettre en oeuvre. Cependant elle reste peu robuste aux déformations des facettes des objets 3D [29]. La solution à ce problème est l'utilisation du centre de gravité surfacique. Cette méthode est née du besoin de normaliser les objets maillés, dont les sommets sont fortement corrélés. Son principe est de pondérer les sommets des facettes par leur surfaces respectives lors du calcul du centre de l'objet [53]. Le centre de gravité surfacique est défini par la formule suivante :

$$C_g = \frac{1}{S} \sum_{i=1}^n S_i C_g^i \quad (4.2)$$

C_g représente le centre de gravité de l'objet 3D.

S_i représente la surface de la $i^{\text{ème}}$ facette de l'objet 3D.

n représente le nombre de facettes de l'objet.

C_g^i représente le centre de gravité de la $i^{\text{ème}}$ facette. Ce centre est calculé par la formule 4.1.

Dans [53] le déplacement du centre de l'objet est résumé par l'équation suivante :

$$T = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \longrightarrow \begin{bmatrix} x + T_x \\ y + T_y \\ z + T_z \end{bmatrix} \quad (4.3)$$

Les paramètres de décalage du centre (T_x, T_y, T_z) , sont obtenus en respectant la condition suivante :

$$m_{100} = m_{010} = m_{001} \quad (4.4)$$

m_{ijk} représentent les moments d'ordre i, j, k de la surface S_{ijk} [53]. La normalisation est faite selon les critères suivants :

- La version normalisée d'un objet 3D doit impérativement garder l'intégralité du contenu initial de l'objet.
- Les critères du calcul des moments m_{ijk} doivent être respectés, ceux-ci sont détaillés dans le papier[53].

Vranic dans [135] propose la méthode EBB *Extended Bounding Box*, qui est une version particulière de la méthode appelée la boîte englobante. Le centre de l'objet dans cette est calculé par la formule suivante :

$$C_b = \frac{x_{min} + x_{max}}{2}, \frac{y_{min} + y_{max}}{2}, \frac{z_{min} + z_{max}}{2} \quad (4.5)$$

$x_{min}, y_{min}, z_{min}$ représentent les coordonnées carstisiennes minimales de l'objet dans le repère de la boîte englobante.

$x_{max}, y_{max}, z_{max}$ représentent les coordonnées carstisiennes maximales de l'objet dans le repère de la boîte englobante.

La particularité de cette méthode réside dans l'utilisation d'une forme englobant l'objet dans son intégralité. Cette qualité facilite la projection de l'objet sur les facettes de cette forme, ceci correspond parfaitement au cas de l'indexation et de la recherche d'objet 3D multi-vues.

L'objectif de Vranic dans [135] a été de définir l'échelle de la sphère englobante l'objet. Étant donné que le centre de l'objet dans ce travail correspond au centre de la forme englobante ; l'auteur utilise le rayon maximum de l'objet comme un rayon à tous les objets :

$$R = \max ||C_b - S_i||^2 \quad (4.6)$$

C_b représente le centre de l'objet à normaliser. S_i représente l'ensemble des sommets de l'objet.

Tous les objets sont ensuite ramenés dans ce repère commun. Pour pallier aux problèmes causés par les bruits sur le contour des objets lors de l'acquisition. Vranic modifie l'algorithme de calcul de l'échelle présenté précédemment, en optant pour le rayon moyen comme un rayon du repère de l'objet. Le nouveau rayon est calculé par la formule suivante :

$$R = \frac{1}{n} \sum_{i=1}^n ||C_b - S_i||^2 \quad (4.7)$$

n représente le nombre de sommets sur le contour de l'objet.

Dans le cas des approches employant la boîte englobante. Une autre méthode de calcul de l'échelle est basée sur le calcul du rayon de la boîte englobante par la formule suivante :

$$R = ||C - S_{max}||^2 \quad (4.8)$$

S_{max} représente le point le plus éloigné de la surface de la boîte, par rapport à son

centre [135].

Dans le cas des objets présentant une parfaite symétrie entre leur différentes parties, Podolak et al [105] considèrent le point d'intersection entre les trois plans orthogonaux de symétrie comme le centre de l'objet. Cette méthode est testée sur différentes bases d'objets 3D, les résultats montrent qu'elle est robuste aux déformations et aux translations. Cependant cette méthode est jugée inappropriée aux asymétries [105].

4.4 Normalisation des light field

Notre étude étant principalement portée sur la caractérisation des vues 2D des light field, la projection de ceux-ci doit être prise en considération dès la phase de la normalisation. Sachant que notre système de projection est basé sur une forme sphérique, cette même forme doit être utilisée lors de la phase de la normalisation des objets 3D. À cet effet, nous proposons un système de normalisation des light field basée sur une sphère englobante.

La normalisation que nous présentons ici, est orientée dans deux directions principales. En premier lieu, les approches de l'indexation basées vues 2D, nécessitent la projection intégrale des objets. Aucune occlusion des formes n'est tolérée. Le deuxième problème concerne l'angle de la prise de vue, ceci doit être aussi large qu'il puisse capturer la plus grande partie du contour de l'objet, d'autant plus que notre méthode de l'indexation est basée sur l'utilisation de l'attribut contour.

Dans la section précédente, nous avons souligné que les méthodes de la normalisation basées sur les formes englobantes, ne se soucient pas des contraintes citées ci-dessus. La mise d'un objet 3D dans une forme englobante peut dégrader sa qualité, en fait cette opération nécessite le sous échantillonnage de l'objet, ce qui peut provoquer une perte de certains détails de son contour (voir figure 4.3), dans d'autres cas, la forme ne peut même pas envelopper l'intégralité de l'objet (voir figure 4.4). Or, une bonne caractérisation des objets 3D nécessite une préservation parfaite de la forme à indexer. Ces considérations étaient derrière la conception d'une nouvelle méthode de la normalisation basée une forme englobante sphérique. Contrairement à la boîte englobante, la forme utilisée dans notre méthode garantit une bonne projection de l'objet 3D. Autrement dit, la boîte englobante ne garantit pas une optimisation des surfaces des projections. La forme qui est utilisée dans notre méthode, est constituée de formes régulières qui sont bien réparties sur son contour. Cette organisation des points de vue permet de maximiser la surface de l'objet dans les différentes projections obtenues.



FIGURE 4.3 – *Illustration de l'effet négatif du sous échantillonnage sur la capture du contour de l'objet. (a) représente un contour bien échantillonné. (b) représente un contour sous échantillonné.*

4.4.1 Calcul du centre de l'objet

Dans cette étape, nous pouvons calculer le centre de l'objet par la formule 4.5. Néanmoins, l'utilisation des sommets dans cette formule n'est pas appropriée à notre cas, où les objets sont fortement maillés. À cet effet, nous nous sommes tournés vers une nouvelle formule, où les sommets sont remplacés par ceux des triangles de l'objet. Ceci dit, le calcul du centre de l'objet dans notre travail est réalisé par la formule [?]



FIGURE 4.4 – *Illustration de la projection d'un objet 3D sur deux formes différentes. (a) représente un objet bien projeté. (b) représente un objet rogné.*

La représentation géométrique des objets 3D est très avantageuse, dans la mesure où elle applique une triangulation surfacique à l'intégralité de l'objet. Étant donné que notre algorithme de la normalisation des objets 3D nécessite une telle triangulation, cela nous facilitera la tâche. La structure employée décrit chaque triangle par un ensemble de sommets et de faces. Un exemple de cette structure est présentée dans l'annexe 7.7.

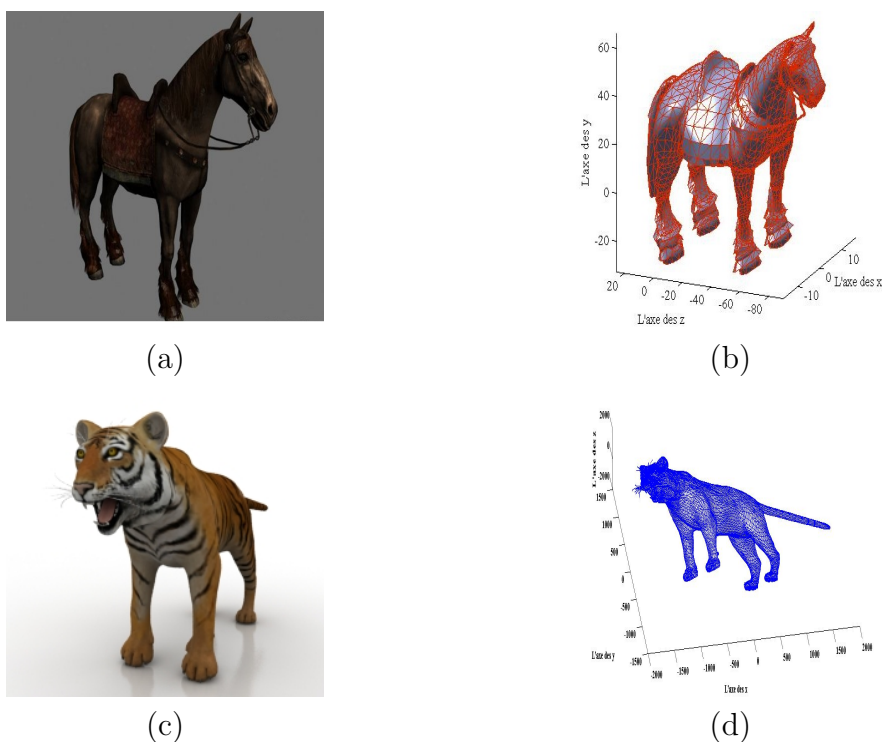


FIGURE 4.5 – Exemples d’objets 3D maillés. (a, c) exemples d’objets 3D. (b, d) représente l’objet triangulé dans un repère 3D.

4.4.2 Construction de la Sphère minimale englobante

La normalisation des objets 3D, consiste à : 1) Construire la sphère minimale englobante les objets 3D de la base. 2) Centrer tous les objets de la base. 3) Mettre à l’échelle les objets. Dans cette section, nous nous intéressons à la construction de la sphère minimale. L’objectif de cette étape est de construire une forme, qui enveloppe tous les objets de la base. La construction de la sphère minimale est un problème très compliqué, l’intérêt à ce sujet remonte au début du dix-neuvième siècle. Les études faites dans ce domaine ont permis de mieux cerner ce sujet et de proposer des algorithmes rapides et efficaces, à l’instar du travail proposé par Fisher et al dans [129]. Dans notre travail, la construction de la sphère minimale est basée sur l’algorithme de Gartner [25]. Cet algorithme est détaillé dans l’annexe 7.7.

L’objectif de cet algorithme est de construire une sphère minimale enveloppant l’objet à normaliser. Cet algorithme est un processus itératif de minimisation d’une forme sphérique initiale S_0 , dont le rayon r est égale à la distance d entre le sommet p_1 le plus éloigné et l’origine θ de la sphère S_0 . Lors de la première itération, cette origine se déplace en direction de p_1 permettant ainsi la minimisation de la sphère

S_0 . Cette itération ne s'arrête que si et seulement si d'autres sommets de la forme viennent se positionner sur le contour de la sphère. L'arrêt de cet algorithme est sujet à l'emplacement de l'origine θ , si celui-ci se trouve à l'intérieur d'une forme convexe constituée par les sommets de la forme, la sphère obtenue est considérée comme une sphère minimale (voir figure 4.6).

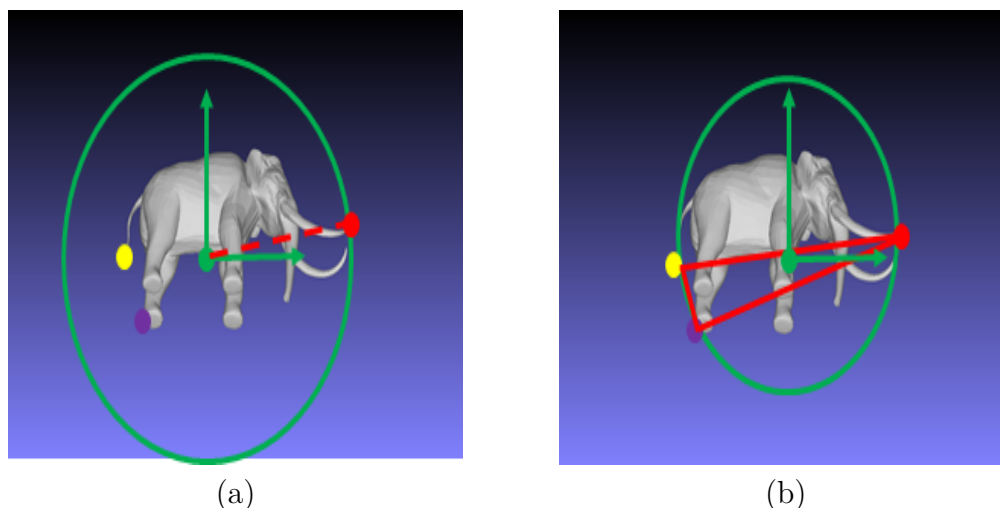


FIGURE 4.6 – Exemple de la construction de la sphère minimale. (a) représente la première itération de l'algorithme de la construction de la sphère minimale, le cercle vert représente la sphère initiale avec le repère en vert, le rayon représenté par un trait discontinu rouge, le point rouge représente le point le plus distant par rapport à l'origine du repère de la sphère initiale. les points jaune et violet représentent deux points du contour de la forme. (b) représente la deuxième itération de l'algorithme de Gartner

4.4.3 Centrage de l'objet 3D

Le centrage des objets 3D, désigne la technique qui permet de déplacer tous les objets à la même origine du repère 3D. D'une manière plus générale le problème appelé *centrage des objets 3D* est le suivant : on dispose d'un ensemble d'objets dispersés dans le repère 3D et on souhaite déterminer les paramètres T_x , T_y et T_z de la fonction de translation de ces objets à une origine commune.

Une fois que la sphère minimale est calculée, tous les objets de la base sont placés au centre de cette sphère. Les coordonnées cartésiennes de l'origine θ de cette sphère (θ_x , θ_y et θ_z) représentent les paramètres de la translation énoncée précédemment. À cet effet, les nouvelles coordonnées (p'_x , p'_y et p'_z) des points p de l'objet,

sont calculées par les formules suivantes :

$$p'_x = p_x - \theta_x \quad (4.9)$$

$$p'_y = p_y - \theta_y \quad (4.10)$$

$$p'_z = p_z - \theta_z \quad (4.11)$$

4.4.4 Mise à l'échelle de l'objet 3D

La mise à l'échelle est un point clé de la normalisation des objets 3D. Dans [23] par exemple, on a précisé que le changement de l'échelle peut influencer considérablement sur la capacité discriminante de l'algorithme de la reconnaissance de visages. À titre d'illustration, considérons deux objets similaires O_1 et O_2 , dont les tailles S_1 et S_2 sont très différentes ($S_1 \gg S_2$). Les indices de l'attribut couleur de l'objet O_1 , révèle une très grande occurrence des patterns de cet attribut par rapport à celle de l'objet O_2 . À cet effet, la mesure de la similarité dans cet exemple ne sera pas objective et ne reflètera pas la vraie sémantique des deux objets.

Afin que notre système d'indexation soit robuste aux variations de l'échelle, nous proposons d'introduire la notion de la mise à l'échelle. Celle-ci est basée sur l'utilisation du rayon r de la sphère minimale englobante. La nouvelle échelle est calculée par la formule suivante :

$$r' = 1/r. \quad (4.12)$$

Une mise à jour des coordonnées des points de l'objet est réalisée à l'aide des formules suivantes :

$$p''_x = p'_x/r. \quad (4.13)$$

$$p''_y = p'_y/r. \quad (4.14)$$

$$p''_z = p'_z/r. \quad (4.15)$$

La normalisation s'effectuant sur des objets de différentes sémantiques, elle doit prendre en considération les caractéristiques intrinsèques de chaque classe d'objets. Dans [93], l'auteur a préféré séparer la normalisation de l'objet de son indexation sémantique. En employant la même sphère minimale pour tous les objets de la base, la normalisation des objets ne dépend pas de leur taille. Les objets chien et girafe dans la figure 4.10 sont normalisés à l'aide de la même sphère englobante. Or ceci peut causer un sous échantillonnage très important de l'objet girafe, qui se traduit par une mauvaise caractérisation de son contenu sémantique. Dans notre approche, nous proposons d'adapter la taille de la sphère minimale à la sémantique de l'objet,

on construit autant de sphères minimales que de classes sémantiques.

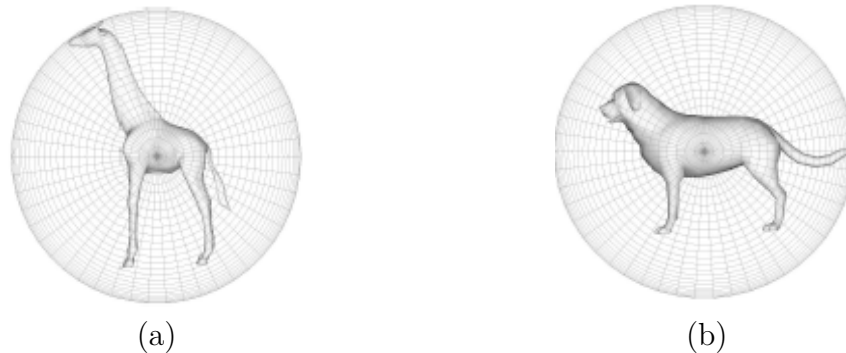


FIGURE 4.7 – Influence de la normalisation sur la sémantique de l'objet. (a) représente l'objet girafe à l'intérieur de la sphère minimale. (b) représente l'utilisation de la même sphère pour envelopper l'objet chien (figure extraite de [93]).

4.5 Méthodes d'alignement

Une étude approfondie des approches de l'alignement des objets 3D, montre le grand intérêt porté à l'utilisation de l'analyse en composante principale (ACP). Cette technique est connue pour être la plus simple et la plus utilisée. Son principe est d'utiliser les moments principaux des matrices de covariance comme des axes du repère de l'objet 3D [135]. Initialement utilisée pour définir les axes principaux des nuages des points, l'utilisation de l'ACP a ensuite été étendue à d'autres domaines, tel que le domaine de l'alignement des modèles 3D. L'objectif principal de l'analyse en composante principale, est de construire une base de projection de l'ensemble des facettes de l'objet. Pour cela, l'ACP est appliquée sur un nuage de points représentant les sommets du modèle 3D.

L'analyse en composante principale ou la transformée de Karhunen-Loève a été utilisée pour la première fois en 1901 par le mathématicien britannique Karl Pearson [11]. Cette transformée permet de représenter la géométrie du nuage des valeurs quantitatives. Soit par exemple un ensemble de données de n variables quantitatives $\{\mu\}$, qui sont les mesures de m unités (ou individus) ν . L'association des deux ensembles est représentée par une matrice de variables x .

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ x_{31} & x_{32} & x_{33} & \dots & x_{3n} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & \dots & x_{mn} \end{bmatrix} \quad (4.16)$$

La figure 4.8 illustre la représentation des ensembles $\{\mu\}$ et $\{\nu\}$ dans un repère spatial, dont l'origine est un vecteur particulier de l'espace R . Dans cet exemple l'origine est le vecteur dont toutes les coordonnées sont nulles. Ces nuages de points peuvent avoir plus de deux repères, rendant ainsi leur visualisation quasi impossible. La solution vient cette fois-ci des méthodes factorielles, dont l'objectif est de trouver un système d'axes et de plans pour ce genre de nuage de points. La projection de ce nuage de points sur cet ensemble d'axes, ne doit pas causer d'importantes déformations. En fait, la reconstitution des positions des points dans ce système de plans enregistre généralement une certaine déformation. Celle-ci ne doit pas être importante, afin que le nuage de points garde son apparence originale et les mêmes distances entre ses composantes [134].

Afin de simplifier les illustrations graphiques des nuages des points. Nous représentons les deux ensembles $\{\mu\}$ et $\{\nu\}$ dans le même espace. Les figures 4.8 (a) et (b) montrent que les nuages des points peuvent être regroupés sous la forme d'ellipsoïdes. Les deux ellipsoïdes qui sont employées dans ces deux exemples ne contiennent pas le même nombre de points. En fait, à chaque valeur de l'ensemble des données $\{\nu\}$ (boules bleues) un vecteur de variables quantitatives (boules rouges) est associé. Ce vecteur étant de longueur supérieure à trois, sa représentation graphique s'avère vite impossible. L'analyse en composante principale cherche une représentation des m données, dans un sous-espace de dimension R_n , qui est inférieure ou égale à trois. Cela ne sera possible qu'en combinant linéairement les variables quantitatives, tout en préservant leur apparence initiale. Ces variables seront appelées *composantes principales*. Les axes qu'elles déterminent sont les *axes principaux* [134].

– Choix de la métrique de comparaison

La recherche d'une représentation géométrique sans perte de l'information, à ramener à la définition d'une métrique de calcul de la distance entre les différents éléments du nuage des points. Ceci dit cette métrique est d'autant importante que son choix représente un aspect primordial dans la construction de la représentation

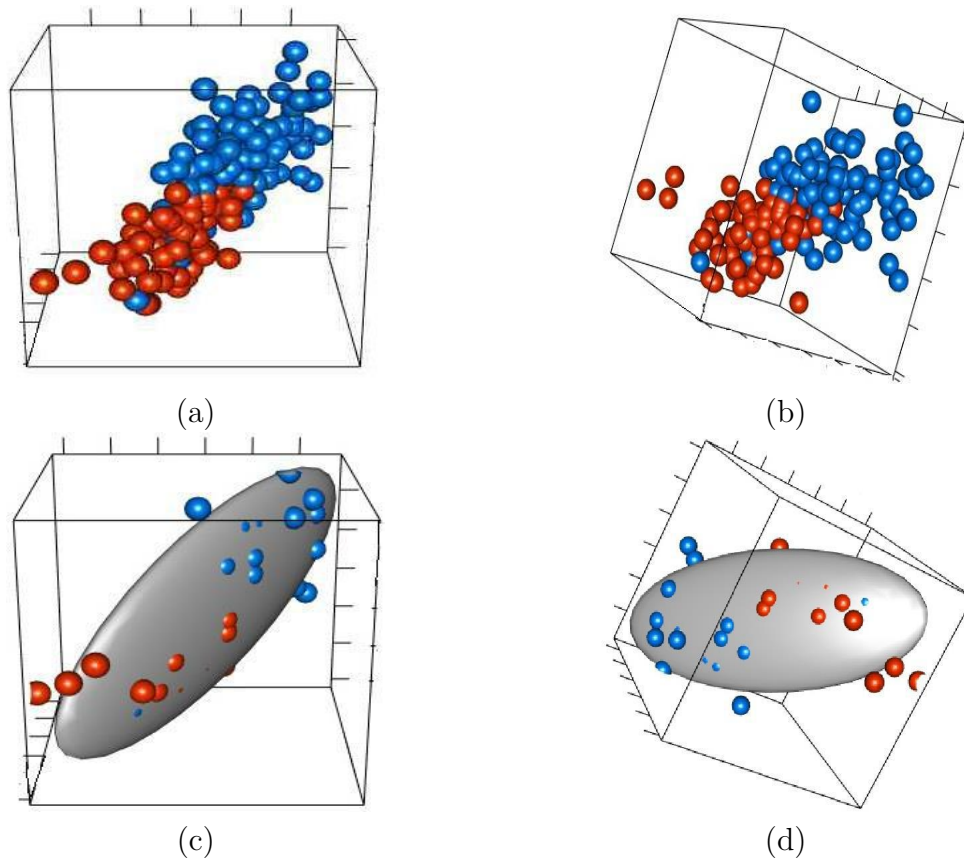


FIGURE 4.8 – *Représentation d'un ensemble $\{\nu\}$ de données et de leur variables quantitatives $\{\mu\}$. (a),(b) Les boules rouges représentent les variables quantitatives, et les boules bleues représentent l'ensemble des données, les deux ensembles sont représentés dans deux espaces, celui de la figure (a) et celui de la figure (b). (c), (d) représentent les formes générales des ensembles des données des deux figures (a) et (b) [11].*

géométrique [11]. Dans le cas de l'analyse en composante principale, on utilise la distance euclidienne, qui est définie par la formule suivante :

$$d^2(\nu_i, \nu_{i'}) = \sum_{j=1}^n (x_{ij} - x_{ij'})^2 \quad (4.17)$$

– Choix de l'origine du repère

Le point aux coordonnées nulles est souvent choisi comme une origine du repères des axes. Ce choix n'est pas approprié au nuage de points dispersés (où les cordonnées des points du nuage sont grandes). Afin de palier à ce problème, il est plus judicieux

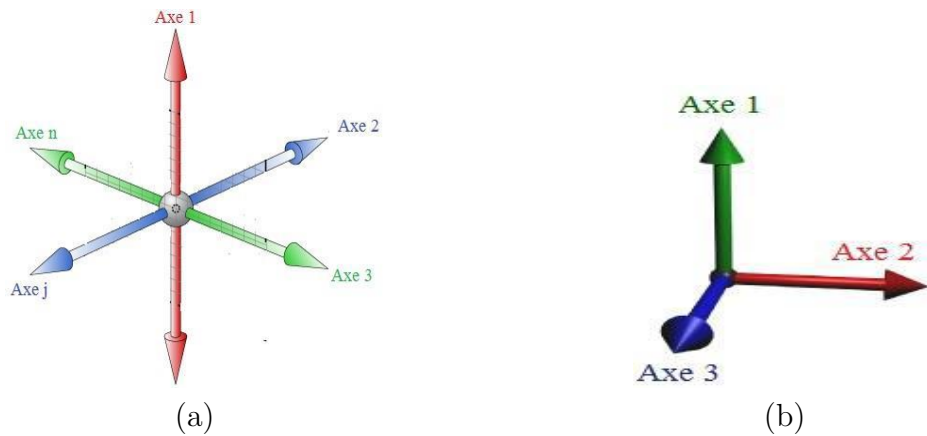


FIGURE 4.9 – Illustration de la construction des axes principaux. (a) Représente le système d'axes de l'ensemble des variables quantitatives $\{\mu\}$. (b) représente les axes principaux du système des axes de la figure (a).

d'opter à une origine liée au nuage des points lui même. C'est le cas du centre de gravité du nuage des points, dont les coordonnées sont données sous la forme suivante :

$$G = \begin{bmatrix} \frac{1}{m} \sum_{i=1}^m x_{i1} \\ \frac{1}{m} \sum_{i=1}^m x_{i2} \\ \vdots \\ \frac{1}{m} \sum_{i=1}^m x_{in} \end{bmatrix} \quad (4.18)$$

– Définition des axes principaux et des vecteurs directeurs

Outre la définition de la distance et de l'origine, la représentation géométrique d'un nuage de points concerne également la définition des différents axes et de leurs vecteurs directeurs. Cette phase est responsable de l'estimation de la pose du nuage des points, qui pourrait être une forme bi- ou tri-dimensionnelle. La principale contrainte qui guide la construction du nouveau repère des axes porte sur la déformation du nuage des points. La projection de celui-ci sur le nouveau repère ne doit produire qu'une faible déformation. Cette contrainte peut se traduire par le choix des axes les plus proches des différents points de notre nuage. Autrement dit, la dispersion des points autour de ces axes doit être faible, les points doivent être regroupés autour de ces axes. La dispersion d'un nuage de points par rapport à un axe V est mesurée par l'inertie de ce nuage par rapport à cet axe. L'inertie est donnée par la formule suivante :

$$I_V = \frac{1}{n} \sum_{i=1}^n d^2(P_{V_i}, \nu_i) \quad (4.19)$$

P_{V_i} représente la projection du point ν_i sur l'axe V_i .

d^2 mesure la distance euclidienne entre le point ν_i et sa projection P_{V_i} .

La distance euclidienne peut être exprimée à l'aide du produit scalaire par la formule suivante :

$$d^2(P_{V_i}, \nu_i) = \langle \overrightarrow{P_{V_i}}, \overrightarrow{\nu_i} \rangle^2 \quad (4.20)$$

$$\langle \overrightarrow{P_{V_i}}, \overrightarrow{\nu_i} \rangle^2 = P_{V_i}^* \cdot \nu_i \cdot \nu_i^* \cdot P_{V_i} \quad (4.21)$$

$P_{V_i}^*$ représente la transposé du vecteur P_{V_i} .

ν_i^* représente la transposé du vecteur ν_i .

En utilisant la symétrie du produit scalaire, l'inertie du nuage des points autour du transposé de l'axe V s'écrira sous la forme suivante :

$$I_{V^*} = \frac{1}{n} \sum_{i=1}^n P_{V_i}^* \cdot \nu_i \cdot \nu_i^* \cdot P_{V_i} = P_{V_i}^* \left[\frac{1}{n} \sum_{i=1}^n \nu_i \cdot \nu_i^* \right] P_{V_i} \quad (4.22)$$

Sachant que la matrice de covariance des données ν_i est égale à $[\frac{1}{n} \sum_{i=1}^n \nu_i \cdot \nu_i^*]$, l'équation 4.22 s'écrira sous la forme suivante :

$$I_{V^*} = P_{V_i}^* \text{cov}(\nu_i) P_{V_i} \quad (4.23)$$

Le premier axe se présente comme un vecteur V_1 passant par le centre de gravité et dont l'inertie est minimale. Son vecteur directeur $\overrightarrow{G_{a_1}}$ est défini sous la contrainte suivante :

$$\|\overrightarrow{G_{a_1}}\|^2 = 1. \quad (4.24)$$

La définition de l'axe V_1 nécessite le calcul de son inertie, qui est égale à la plus grande valeur propre de la matrice de covariance des variables quantitatives. Le vecteur directeur unitaire de cet axe est le premier vecteur propre, qui est associé à la plus grande valeur propre de la matrice de covariance des variables quantitatives. Le second axe V_2 se présente comme un axe perpendiculaire au premier et dont l'inertie est minimale. Le vecteur directeur unitaire de ce deuxième axe est égale à la deuxième plus grande valeur propre de la matrice de covariance des variables quantitatives. Ces deux axes peuvent ainsi définir l'espace 2D du nuage des points. Les n vecteurs propres de la matrice de covariance des variables quantitatives peuvent ainsi définir n axes de cet espace, dont les vecteurs directeurs unitaires correspondent aux valeurs propres de cette matrice.

– *Calcul des valeurs quantitatives dans le nouveau plan*

Chaque point peut être projeté orthogonalement dans ce nouveau plan. L'analyse en composante principale permet en outre de définir l'orientation des parties de la forme les unes par rapport aux autres. Autrement dit, Si deux individus possèdent des coordonnées de signes opposés, cela signifie que les deux points s'opposent sur cet axe.

Considérons par exemple un objet tridimensionnel O de n points e_i et d'une orientation quelconque. L'application de l'analyse en composante principale sur cet ensemble de points produit un repère d'orientation appelé R , qui doit vérifier les conditions suivantes :

$$\min \sum_{i=1}^n \|e_i - e_i^p\|^2 \quad (4.25)$$

$$\|R\| = 1 \quad (4.26)$$

e_i^p représente la projection du point e_i dans le nouveau repère R .

Selon ces conditions, la somme des distances au carré entre les points e_i et leurs projections respectives e_i^p , doit être faible[135]. Ceci passe par la recherche des axes principaux de la matrice de covariance. Les vecteurs propres de cette matrice représentent les axes d'orientation de la forme 3D.

Au sens de la notion du moindre carré, l'ACP se présente comme une régression linéaire des distances entre les différents points de l'objet et leurs projections. Cependant l'ACP se révèle très fragile devant les déformations des facettes de l'objet. Autrement dit, tout changement de la silhouette de l'objet peut produire un repère d'orientation très différent, c'est le cas des objets à articulations où les sommets sont souvent exposés aux déformations et aux changements d'apparence [49].

Cette façon d'appliquer l'analyse en composante principale s'avère très vite inappropriée aux objets 3D, lesquels sont souvent représentés par un ensemble inéquitable de points. Le nuage de points doit donc être considéré comme un ensemble de variables, dont chacune d'elle contribue d'autant plus à la confection des axes du repère, que sa variance est forte. Une solution classique consiste à normaliser les variables en divisant leurs valeurs par leurs écart-types. Ce traitement va être à l'origine d'un ensemble d'opérations, qui aboutiront à la création d'un espace de variables centrées et réduites. Il convient de constater que cette normalisation entraîne une reformulation du calcul de la distance euclidienne, qui est utilisée dans

l'ACP. La nouvelle formule est définie par :

$$d^2(\nu_i, \nu_{i'}) = \sum_{j=1}^n \frac{1}{\sigma_j^2} (x_{ij} - x_{ij'})^2 \quad (4.27)$$

σ_j^2 représente l'écart-type du nuage des points.

Dans la même optique, Paquet et al ont proposé la pondération des centres de gravité des facettes par leur surfaces respectives. Vranic et Saupe ont également proposé une nouvelle méthode appelée l'analyse en composante principale continue (ACPC), qui peut être appliquée sur n'importe quelle distribution de points à condition qu'elle soit triangulée au préalable. L'analyse en composante principale continue est formalisée ainsi : Soit un objet O constitué d'un ensemble de surfaces $T_{i=1}^m$. m représente le nombre de surfaces T_i engendrées par la triangulation de cet objet. Soient α_i , β_i et λ_i les sommets de la surface T_i , G_i représente le centre de gravité de cette surface. L'application de l'ACP sur cet ensemble de surfaces consiste à calculer la matrice de covariance selon la formule suivante :

Pour parvenir à une estimation optimale de la pose des objets 3D, Vranic et Saupe [137, 136, 135] ont enrichi l'ACP en rajoutant de nouvelles méthodes de calcul de la matrice de covariance.

4.6 Alignement des light field

En parallèle à la normalisation des objets 3D, une étape d'alignement doit être réalisée. L'intérêt de cette étape, à terme, est de pouvoir orienter les objets d'une même classe de la même façon. La tête des objets de la classe homme par exemple, doivent être orienter selon l'axe des y , les pieds doivent être orientés selon l'axe des y , et les mains selon l'axe des x . Prenons par exemple le cas des objets de la classe voiture, il est plus intéressant d'orienter le toit selon l'axe des y , les feus de l'avant et de l'arrière selon l'axe des x ou celui des z

Dans notre travail [23], l'application de l'analyse en composantes principales a permis d'aligner tous les visages de la base. Ces visages sont tout d'abord projetés sur le même plan 2D, cette projection produit un nuage de points correspondants aux points caractéristiques des visages. Cet ensemble de points se distingue par une concentration de la majorité des points et une forte variation du reste. L'application de l'ACP élimine les fortes variations et maintient le reste, elle permet également d'extraire les vecteurs propres de cet espace, lesquels vont être utilisés comme vecteurs unitaires du nouveau repère.

Contrairement à cette méthode, les méthodes d'alignement citées ci-dessus proposent la construction du nuage de points, à partir de tous les objets de la base. Les objets alignés par ces méthodes sont des entités 3D et non pas des images 2D.

A l'instar de la méthode de [23], les sommets des objets de chaque classe constitue un nuage de points. L'application de l'ACP réduit l'espace de variation et produit trois vecteurs propres. Le repère 3D dans notre travail, est constitué de ces vecteurs propres. Chaque objet est ensuite projeté dans ce nouveau repère, les signes des différentes parties sont ensuite utilisés pour les orienter les unes par rapport aux autres. La figure [?] montre le résultat de l'alignement d'un ensemble d'objets 3D de la base PSB.

4.7 Conclusion

Les deux principaux volets traités dans ce chapitre sont : la normalisation et l'alignement des objets 3D. En premier lieu, nous avons proposé de centrer les objets de chaque classe à l'aide de la sphère minimale englobante. La particularité de cette méthode est l'adaptation de cette sphère à chaque classe d'objets. Cela permet la prise en compte de la sémantique de l'objet dès la phase de normalisation. Nous avons ensuite utilisé l'analyse en composante principale pour réduire la variation du nuage des points des objets de chaque classe, les vecteurs propres calculés sont ensuite utilisés pour construire le nouveau repère 3D. Les résultats obtenus satisfaites le critère de robustesse vis-à-vis les rotations des objets et les changements de leurs échelles.

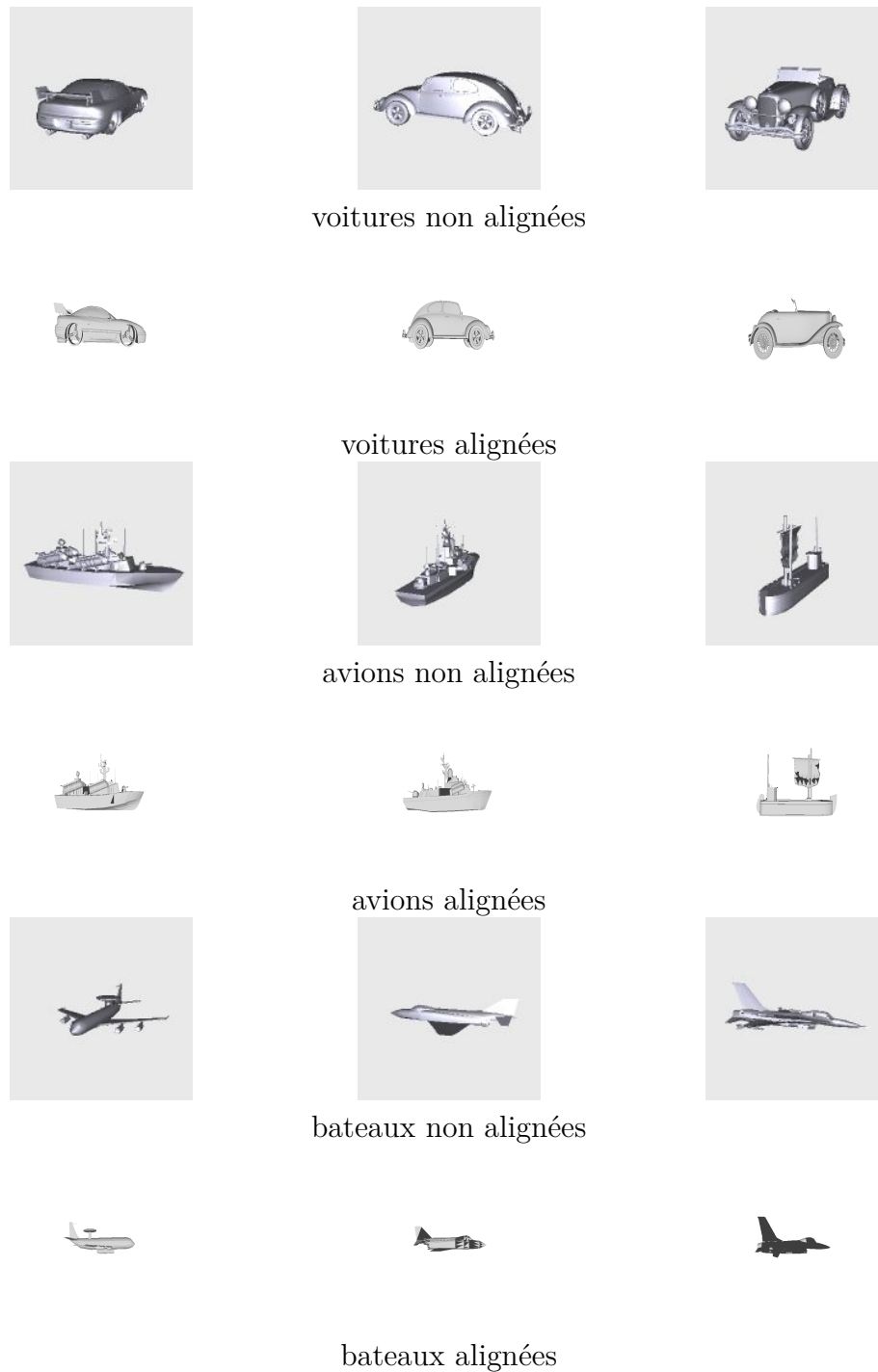


FIGURE 4.10 – *Résultat de l'alignement d'un ensemble d'objets à l'aide de notre algorithme.*

Chapitre 5

Indexation Des Light Field Par La Transformée en Ondelettes Basée Région

Résumé : Dans ce chapitre nous proposons notre principale contribution, portant sur l'indexation des modèles 3D. L'originalité de notre approche réside dans l'utilisation de la transformée en ondelettes basée région dans un contexte d'indexation multi-vues des objets tridimensionnels. L'idée principale est d'exploiter les avantages offerts par l'indexation basée régions des projections 2D pour construire les indices de la forme, de la couleur et de la texture. Dans ce chapitre nous présentons le schéma général de l'approche d'indexation des light field par la transformée en ondelettes basée régions. Nous expliquons ensuite notre méthode d'extraction des vues 2D. À la fin de ce chapitre, nous analysons les propriétés de caractérisation et de discrimination de quelques indices d'objets 3D, qui ont été construits à l'aide de notre technique.

5.1 Introduction

Les progrès des systèmes d’acquisition et de rendu de modèles 3D durant les dernières décennies ont encouragé les chercheurs à s’intéresser à leur stockage et à leur compression. Cet intérêt qui a été dédié principalement à des applications spécifiques, a connu une évolution importante avec l’expansion massive de l’utilisation du 3D dans tous les domaines. Dans les premiers temps, les recherches ont été focalisées sur les méthodes d’indexation géométrique et topologique des modèles 3D. Ces méthodes sont basées sur l’exploitation des caractéristiques géométriques ou topologiques des objets 3D, pour la construction des indices qui sont ensuite utilisés lors de la phase de la recherche.

L’application proposée dans ce travail, permet la recherche de deux types de requêtes : les requêtes sous forme d’objets 3D, ainsi que les requêtes 2D (correspondant à une seule vue de l’objet 3D). Dans le premier cas, nous projetons l’objet 3D, afin d’extraire les vues 2D selon le modèle employé dans les systèmes d’acquisition des light field. En fait, nous proposons de projeter chaque requête 3D sur une forme sphérique, les caméras d’acquisition des light field sont placées dans les coins de cette forme. Contrairement aux requêtes 3D, l’utilisation des requêtes 2D représentant la forme de l’objet recherché, ne nécessite aucune opération de prétraitement.

Ce chapitre ayant pour but la présentation de notre méthode d’indexation des light field par la transformée en ondelettes basée région, nous commençons ce chapitre par la présentation d’un schéma global de cette approche, nous proposons ensuite quelques notions de base concernant le rendu des objets 3D à l’aide des light field. Dans la dernière section de ce chapitre, nous nous focaliserons sur la caractérisation des modèles 3D par la transformée en ondelettes basée région, c’est-à-dire sur l’application de la technique d’indexation 2D proposée dans le chapitre 3 dans le cas de l’indexation des modèle 3D. Néanmoins, nous gardons sans cesse à l’esprit que les vues 2D du modèle 3D représentent des échantillons bien localisés de l’espace 3D. Autrement dit chaque vue à caractériser doit capturer l’intégralité de la forme de l’objet à caractériser.

5.2 Indexation des light-field

L’indexation des light field consiste à caractériser l’ensemble des vues 2D par leur attributs. Autrement dit, l’objet 3D est ici considéré non pas en tant que tel, mais plutôt comme un ensemble de projections 2D. Sa caractérisation est une opération complexe, parce qu’elle est subjective du dispositif d’acquisition utilisé, et des

conditions dans lesquelles elle se déroule : la forme de ce modèle dépend principalement de l'orientation de la caméra et de l'illumination du milieu dans lequel il se trouve. Autrement dit chaque objet 3D est fourni avec une orientation et une échelle quelconque [53].

Les systèmes d'indexation et de recherche des modèles 3D sont toujours confrontés à plusieurs contraintes, l'une d'entre elles concerne la robustesse aux changements de l'échelle, de la translation et de la rotation du modèle 3D. En fait, pour rendre un indice invariant aux variations citées ci-dessous, de nombreux prétraitements doivent être effectués avant de passer à la phase de la caractérisation. Ces prétraitements consistent à :

- Centrer le modèle.
- Mettre à l'échelle le modèle.
- Aligner les objets de chaque classe.

Dans ce travail ces prétraitements sont traités dans le chapitre 04.

Dans notre cas, il est également nécessaire de passer d'un modèle en trois dimensions à une représentation en deux dimensions (voir figure 5.1). Afin d'accomplir cette tâche, deux contraintes doivent être prises en compte. La première concerne le nombre de projections. Quoiqu'il ait un impact positif sur les performances de l'indexation, au-delà d'un certain seuil, ce facteur pourrait avoir un effet négatif, à cause de l'augmentation de la complexité de calculs et du temps du traitement. La deuxième contrainte concerne le choix des angles de vue. Comme nous l'avons précisé dans le chapitre précédent, les angles de vue doivent capturer l'objet dans son intégralité et doivent optimiser la forme de la vue 2D. À cet effet, nous proposons d'utiliser une sphère régulière pour la capture de l'objet.

La dernière étape de l'indexation des light field concerne la construction de leur descripteurs appelés les descripteurs des light field. Dans notre cas, ces descripteurs sont constitués des indices de texture, de couleur et du contour de l'objet (voir figure 5.1). La construction de ces indices est présentée dans le chapitre 03 et elle est basée sur les coefficients de la transformée en ondelettes basée région de l'objet.

5.2.1 Light field

Dans les chapitres précédents, on a souvent parlé de la problématique du rendu des modèles 3D et comment ce sujet est devenu d'actualité ces dernières années au milieu des infographistes. Les méthodes du rendu dites traditionnelles (basées sur la modélisation géométrique), ont déjà prouvé leurs limites, quant à la création de nouvelles scènes à très haute définition. Le rendu de ces méthodes n'a jamais pu

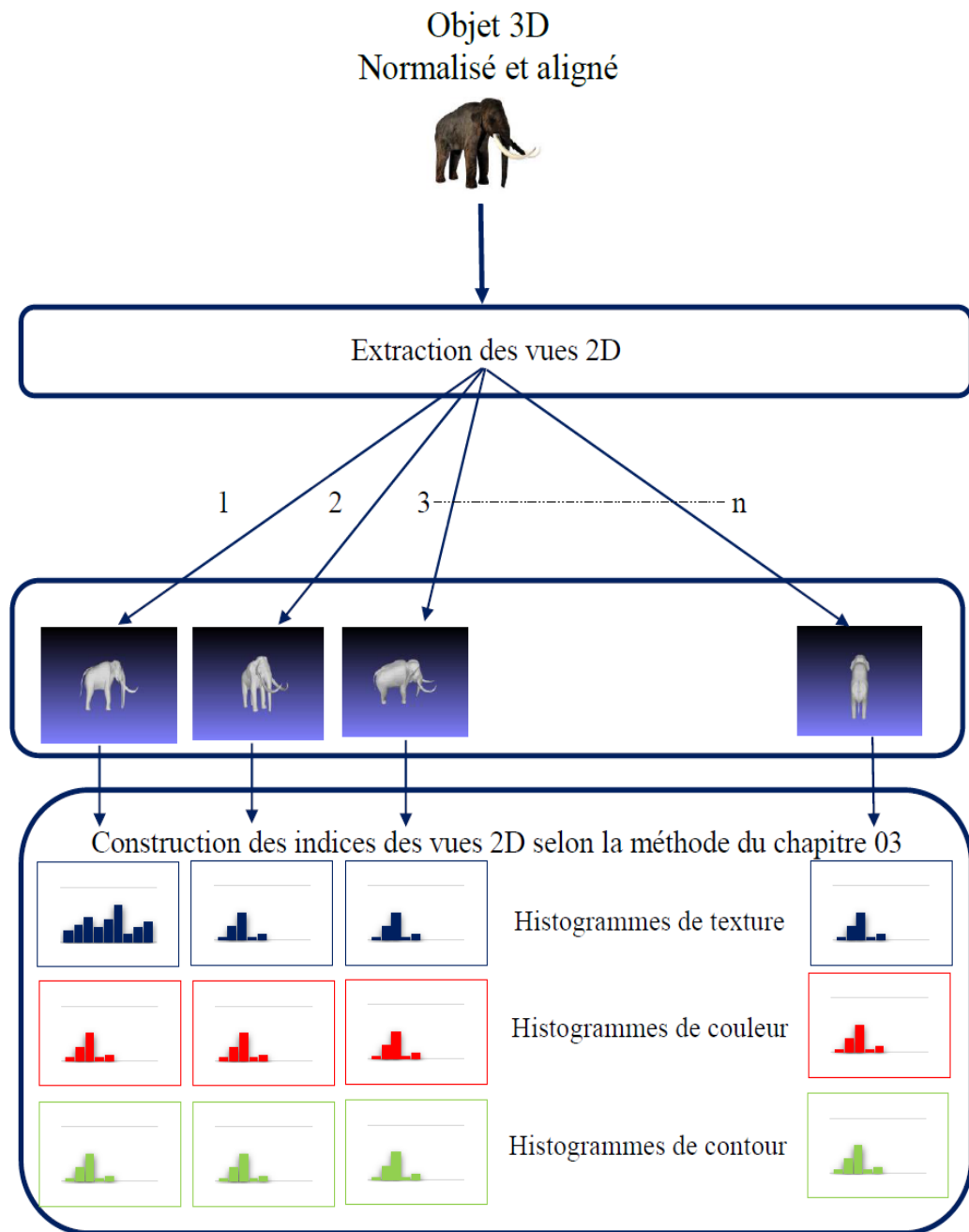


FIGURE 5.1 – Schéma général de l'indexation des light field par la transformée en ondelettes basée régions.

concurrer la qualité des scènes réelles. L'alternative était la proposition des méthodes de rendu basé images (RBI), dont le principal avantage est leur simulation des effets de la lumière des scènes 3D. La représentation des scènes 3D dans ces méthodes, est basée sur une collection d'images 2D. Ceci a permis aux scientifiques

de s'affranchir de la modélisation géométrique des scènes 3D et de réduire la complexité géométrique de ces scènes. Un autre avantage offert par ces méthodes, vient cette fois-ci de la modélisation parfaite des effets de la lumière de la scène 3D.

En 1996, E. Adelson et J. Bergen ont proposé la modélisation des effets de la lumière par une fonction P à 7 paramètres variables, dite la fonction plénoptique [10]. Celle-ci permet la mesure du flux lumineux émis par n'importe quel point de l'espace (V_x, V_y, V_z) , selon les deux directions θ et φ et selon la variation des paramètres temps (t) et la longueur d'onde (λ) (voir figure 5.2) :

$$P(V_x, V_y, V_z, \theta, \varphi, \lambda, t) \quad (5.1)$$

Dans [73] McMilan et Bishop ont réduit le nombre de paramètres de la fonction

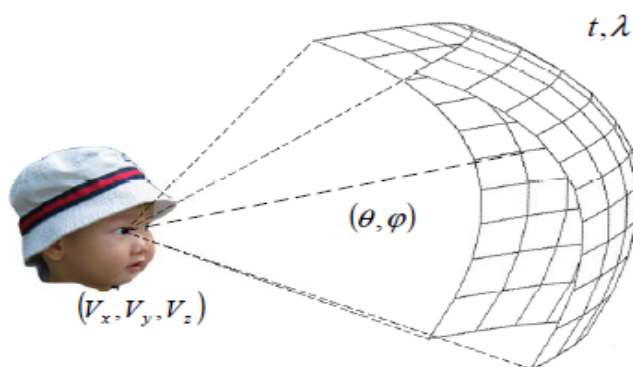


FIGURE 5.2 – Représentation de la fonction plénoptique en un point (V_x, V_y, V_z) et dans une direction (θ, φ) , (image extraite de [27])

plénoptique à cinq, grâce à l'élimination des paramètres θ et φ . Ceci en considérant que les images sont fixes et les conditions d'éclairage de la scène sont statiques :

$$P(V_x, V_y, V_z, \theta, \varphi) \quad (5.2)$$

Cette approche était à l'origine de la conception du premier système du rendu basé image à une fonction plénoptique de cinq dimensions. En fait, les auteurs proposent des cartes sphériques autour des points de vue différents de la scène à capturer. Lors du rendu l'utilisateur de ce système peut générer une représentation quasi continue d'une portion de l'environnement 3D. Ce système peut générer de nouvelles scènes 3D à partir de nouvelles points de vue proches des points de capture. En restreignant l'angle de l'élévation de la caméra, ce système génère des trous noirs si le point de vue est très loin du point de capture original (voir figure 5.3)).

La problématique du rendu basé images, réside dans la difficulté de modéliser tous



(a)



(b)

FIGURE 5.3 – Le rendu des scènes 3D à partir de nouvelles points de vue. (a), (b) représentent des vues panoramiques de deux scènes 3D différentes (images extraites de [73])

les transferts lumineux dans un environnement 3D [89]. La solution proposée dans la majorité des travaux, était de réduire le nombre de paramètres de la fonction plénoptique, Shum et al ont par exemple conçu un système de rendu des objets 3D, à une fonction plénoptique de trois paramètres seulement.

$$P(V_x, V_y, \theta) \quad (5.3)$$

Le rendu dans ce système est basé sur un ensemble de mosaïques. Le dispositif de la capture utilisé est basé sur une seule caméra montée sur un plateau. La scène est capturée à des intervalles réguliers de l'espace (voir figure 5.4). La négligence de la direction verticale dans ce système a causé la distorsion verticale dans les scènes synthétisées [118].

Afin de réduire les dimensions de la fonction plénoptique, certaines méthodes comme celle de S. Chen [34], arrivent à mesurer le flux optique d'un environnement, avec une fonction plénoptique à deux dimensions. Ceci est rendu possible grâce à l'acquisition de l'environnement 3D, depuis un point de vue fixe. La méthode de Chen est utilisée dans de nombreuses applications tant elle est efficace et simple à mettre en œuvre. Le rendu dans cette méthode est basé sur une collection d'images panoramiques cylindriques, dont les paramètres V_x , V_y et V_z sont statiques. La représentation de l'échange des flux lumineux se réduit à une fonction plénoptique à

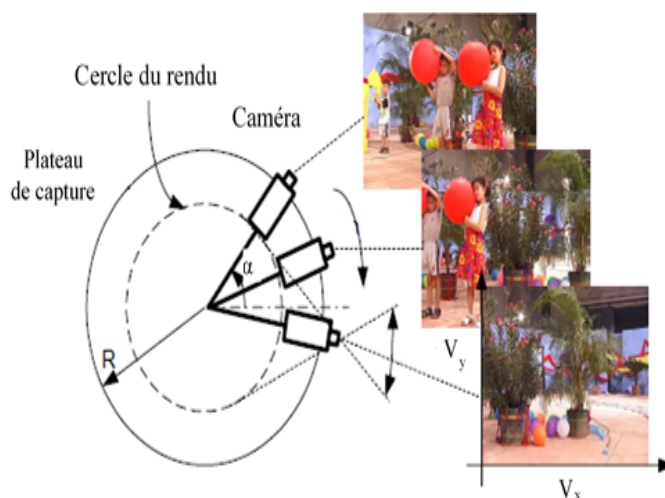


FIGURE 5.4 – *Système de capture basé sur les mosaïques centrées, (image extraite de [27])*

deux dimensions.

$$P(\theta, \varphi). \quad (5.4)$$

L'inconvénient de cette méthode vient du fait que le champ de vue vertical est limité. L'utilisateur est contraint à suivre certains chemins prédéfinis de points de vue pour visualiser l'environnement 3D [89].

R. Szeliski et H. Shum ont amélioré cette méthode grâce à la représentation de l'environnement par des images panoramiques sphériques. Aucune restriction sur les paramètres θ et φ n'est suggérée et ces deux paramètres sont entièrement échantillonnés [127].

Les méthodes citées ci-dessus sont plutôt adaptées au rendu de scènes 3D que d'objets, car elles sont basées sur un échantillonnage partiel de l'environnement. M. Levoy et P. Hanrahan ont démontré qu'une représentation intégrale d'un objet 3D, nécessite un échantillonnage dense de tous les paramètres de la fonction plénoptique, ceci offre une liberté totale à l'utilisateur de générer de nouveaux points de vue de l'espace. Afin d'interpréter ce constat, nous considérons un objet 3D placé dans un système d'acquisition panoramique. C'est-à-dire un système qui ne capture que la face de l'objet en face de la caméra, l'utilisateur dans ce cas ne peut générer de nouvelles vues de l'objet que de cette face, la face cachée de l'objet n'étant pas couverte par l'angle de vue de la caméra, ne pourra pas être rendue.

Dans leur papier intitulé *Light Field Rendering* [74], les auteurs proposent un



(a)



(b)

FIGURE 5.5 – Le rendu des scènes 3D à partir de nouvelles points de vue. (a), (b) représentent des vues panoramiques de deux scènes 3D différentes (images extraites de [73])

Le système d'acquisition du modèle 3D est constitué de deux plans en parallèle (μ, ν) et (s, t) : le premier plan (μ, ν) est composé des points de prise de vue (caméras) à des intervalles réguliers, le deuxième plan (s, t) représente l'objet. Afin d'assurer une meilleure qualité du rendu, un nombre conséquent de photographies est capturé. Le principe de cette méthode est identique à la modélisation de la fonction plénoptique [74]. Les dimensions de la fonction de représentation des rayons lumineux à l'intérieur de l'environnement, sont réduites à quatre : deux dimensions (μ, ν) correspondant au plan de capture et deux autres dimensions (s, t) correspondant au plan de projection. Le rendu des objets 3D à l'aide des light field, consiste à interpoler les plans (μ, ν) et (s, t) [74].

L'objectif du rendu à l'aide des light field est de simuler les effets de la lumière et de mesurer son champ à chaque pixel de l'objet 3D. La figure 5.6 illustre la phase de capture des light field. Dans cette figure, chaque vue du plan (μ, ν) représente la luminance émise de toutes les positions du plan (s, t) en direction de cette vue (caméra). Tandis que les images du plan (s, t) , permettent l'optimisation du rendu lumineux des positions de ce plan, ceci en focalisant toutes les caméras sur la même position du plan (s, t) [74].

Le rendu à l'aide des light field permet de s'affranchir définitivement des problèmes liés à la modélisation géométrique des objets 3D et aux occlusions de certaines parties de l'objet. De plus, il permet d'optimiser la définition de toutes les parties de

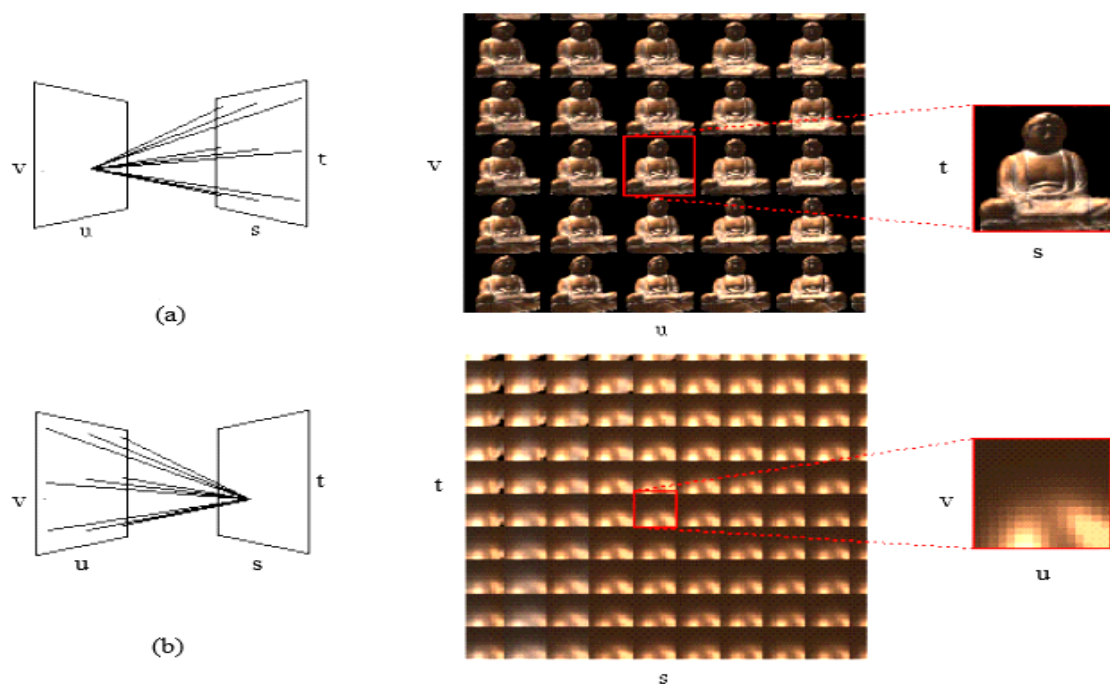


FIGURE 5.6 – Illustration de la capture des light field (image extraite de [74]). (a) représente la construction du plan (μ, ν) , (b) représente la construction du plan (s, t) .

l'objet 3D et en particulier son contour. Ces avantages qui sont offerts par les light field, représentent également les contraintes exigées par notre système d'indexation des modèles 3D. En fait, comme nous l'avons déjà mentionné dans le chapitre 4, l'objet à caractériser doit être capturé dans son intégralité et sa qualité visuelle doit être optimale. A cet effet, notre plan de capture s'est basé sur cette technique.

Les light field sont utilisés dans plusieurs applications à base d'images photo-réalistes, ils permettent le rendu des modèles 3D à très haute résolution et en temps réel, sans utiliser aucun modèle géométrique [33]. Cependant ce genre de système requiert un grand espace de stockage, les prétraitements pour la construction des light field sont très lents; l'autre inconvénient vient du fait que l'acquisition des images est faite par un dispositif très particulier, et difficile à mettre en œuvre.

En effet la très haute résolution produite par les light field nécessite un très grand nombre d'images 2D, on peut facilement dépasser 1,6 *GO* d'espace de stockage. M. Levoy et P. Hanrahan par exemple ont conçu un dispositif d'acquisition des light field, constitué de quatre plans (μ, ν) et de quatre plans (s, t) . Chaque plan (μ, ν) est composé de 16×16 à 64×64 caméras, produisant ainsi $16 \times 16 \times 4 = 1024$ vues 2D ou $64 \times 64 \times 4 = 16\,384$ vues. Sachant que chaque vue est

une image de 256×256 pixels et que chaque pixel est codé sur 4 octets, les images produites par les quatre plans (μ, ν) nécessitent un espace de stockage égale à $1024 \times 256 \times 256 \times 4 \simeq 300 MO$ (dans le cas des plans (μ, ν) de 16×16 caméras) et égale à $16\ 384 \times 256 \times 256 \times 4 \simeq 5 GO$ (dans le cas des plans (μ, ν) de 64×64 caméras). Ajoutons à cela, l'espace nécessaire au stockage de l'information produite par les quatre plans (s, t) , qui sont représentés par des matrices de 32×32 ou 128×128 [74, 89]. Un moyen efficace de la compression des light field, consiste à éliminer la redondance qui existe entre les différentes vue 2D. Le système d'acquisition des light field, expliqué ci-dessus a exploité la redondance des pixels des deux plans (μ, ν) et (s, t) pour la construction d'un dictionnaire de 16 échantillons (4 images du plan $(\mu, \nu) \times 4$ images du plan (s, t)), ce dictionnaire est ensuite utilisé dans un quantificateur vectoriel. Ceci a permis d'atteindre des taux de compression très élevé atteignant 100 :1 [74, 89].

B. Mercier a également proposé une méthode de compression des light field à l'aide de la quantification vectorielle [89]. La particularité de ce travail réside dans son dispositif d'acquisition des light field. Celui-ci est constitué de quatre plans (μ, ν) de 8 vues seulement et de quatre plans (s, t) de 384 images de 256×256 pixels. Étant donné que chaque bloc de couleur peut se répéter dans la même vue, ou dans différentes vues 2D, B. Mercier a proposé l'utilisation des dictionnaires de la quantification vectorielle, pour compresser la réordonnance inter-vue et intra-vues des light field [89].

L'objectif de la compression des light field est d'éliminer la redondance des pixels inter-vues et des pixels intra-vue. Le terme redondance des pixels inter-vus désigne la répétition des pixels entre les différentes vues du light field. La redondance intra-vue porte sur la répétition des pixels dans la même image. La résolution de l'image change d'une application à une autre, selon la bande passante de transmission, l'espace de stockage, le dispositif d'affichage. . . etc [108], ceci dit la représentation de l'objet 3D doit être scalable, pour adapter le rendu à la résolution demandée.

Dans [62, 83], les auteurs ont pris en considération cette notion de scalabilité, grâce à l'utilisation de la transformée en ondelettes. Les light field compressés à l'aide de ces méthodes présentent toutefois une certaine redondance à cause de la parallaxe. Ce terme est utilisé pour désigner la répétition du même pixel dans des vues différentes et à des positions différentes. Afin de palier à ce problème Tong et Gray ont proposé dans [130, 131] la compression de la redondance inter-vue à l'aide de la compensation de la disparité. Girod et al ont amélioré ce travail grâce à l'utilisation de la transformée en ondelettes basée région pour la compression de la

redondance intra-vue, et la compensation de la disparité pour compresser la redondance inter-vue [108].

La compensation de la disparité est initialement inspirée de la compensation du mouvement, qui est utilisée dans les codeurs de vidéo [84, 82]. Dans [108], cette méthode est considérée comme une fonction qui permet le passage d'une vue 2D en une autre. Prenons par exemple, deux vues v_1 et v_2 qui sont capturées depuis les angles de vue A_1 et A_2 . p_2 est une position de v_2 , sa projection dans v_1 est notée par p_1 . Le calcul de la valeur du pixel à la position p_2 consiste à :

- Projeter la position p_2 sur le modèle géométrique de l'objet.
- Extraire la valeur à la position p_1 à l'aide d'une interpolation bilinéaire.

La redondance résiduelle à l'intérieur de chaque vue 2D est ensuite éliminée, à l'aide de la transformée en ondelettes basée région, que nous avons déjà présenté dans le chapitre 3. Cette méthode de compression permet de réduire d'une façon significative la taille des light field. Néanmoins, on ne peut pas l'utiliser tel qu'elle est, à cause de la perte d'information et les erreurs sur le contour, qui sont engendrés par la compensation de la disparité. Notre schéma de compression est basé sur la transformée en ondelettes basée région pour éliminer la redondance intra vue.

Vu leur taille, les light field doivent être compressés avant d'être stockés. La recherche des light field, nécessite un accès direct aux pixels, grâce à la décompression des light field, qui augmente le temps d'accès à ses pixels. Afin de diminuer ce temps, l'indexation et la recherche peuvent être faites dans un niveau plus haut, ce niveau peut être le niveau des light field compressés, ou le niveau des light field transformés. Dans notre travail, nous réalisons les opérations d'indexation et de recherche des light field dans le domaine transformée. Notre schéma d'indexation est basé sur la transformation des vues 2D à l'aide de la transformée en ondelettes basée région.

Comme nous l'avons vu ci-dessus, la structure du dispositif d'acquisition des light field diffère d'un travail à un autre. Shen et al par exemple ont proposé dans [116] un système d'acquisition et de rendu des light field, constitué de vingt caméras réparties sur les coins d'un dodécaèdre (plan de capture (μ, ν)) (voir figure 5.7). Ce système propose l'indexation de 20 light field, correspondant à différentes orientations de l'objet. Grâce à l'élimination par opposition orthogonale des projections opposées, ce nombre est réduit à dix light field (voir figure 5.8). Donc le plan (s, t) comprend dix light field. Chaque caméra peut se déplacer sur les trois sommets des polygones du dodécaèdre, produisant ainsi 3×20 projections du modèle 3D. Autrement dit, chaque light field L_i ($i= 1...10$) du plan (s, t) , est représenté par 60

projections 2D et chaque objet 3D est représenté par $60 \times 10 = 600$ projections 2D.

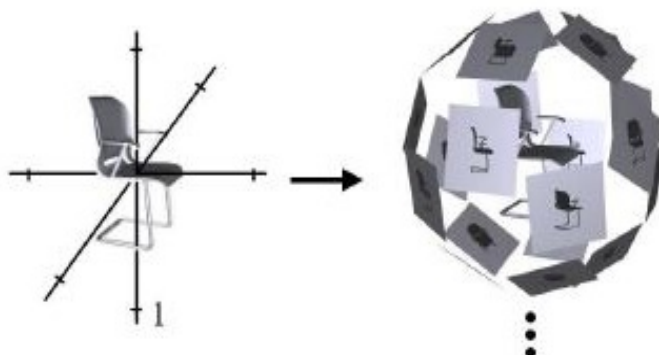


FIGURE 5.7 – *plan de capture* (μ, ν) [116]

Ce système d'acquisition produit 600 projections 2D du modèle 3D, réduisant ainsi le nombre total de vues produit par n'importe quel système d'acquisition des light field. Par exemple le mur de caméras proposé par le système d'acquisition des light field de l'université de Stanford comprend 1000 caméras, réparties sur 7 murs. Dans [116] Chaque projection est ensuite caractérisée par des indices des coefficients de fourier et des moments de Zernike.

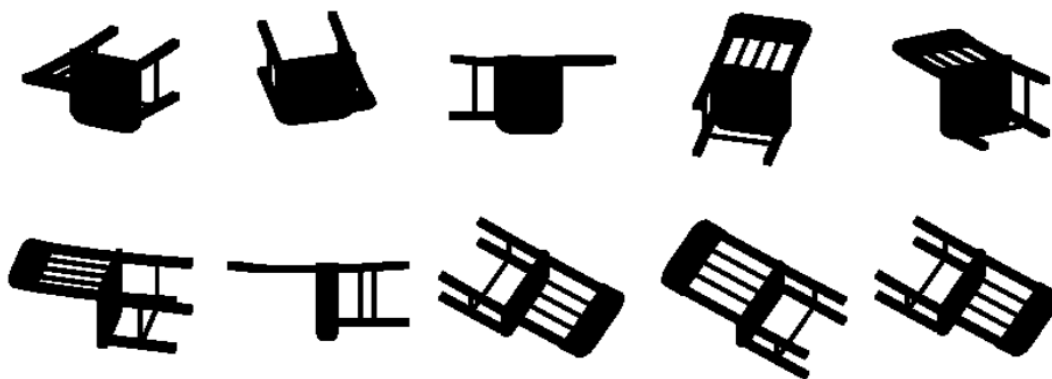


FIGURE 5.8 – *plan* (s, t) *light fields* [116]

Grâce à ce dispositif (voir figure 5.8), de très hauts niveaux de résolution des objets 3D, peuvent être atteints. Or, ce facteur n'est pas nécessairement décisif pour une indexation efficace. Un choix judicieux des vues 2D peut améliorer significativement les performances d'indexation des objets 3D. Partant de ce constat, nous proposons dans le papier *Light field retrieval in compressed domain* présenté à l'occasion de la 21^{ème} conférence internationale en Europe centrale sur l'infographie, la

visualisation et la vision par ordinateur[21], de réduire le nombre de projections à vingt-six, au lieu de soixante vues proposées dans le travail de Shen.

Dans [21], nous cherchons à caractériser les objets 3D par des vues générées par le light field, tout en utilisant un nombre réduit de projections 2D. Ceci nous a amené à ne considérer qu'un seul light field du plan (s, t) et vingt-six angles de prise de vue du plan (μ, ν) (voir figure 5.9). Notre simulation du dispositif de capture de l'objet 3D consiste en :

- Un objet placé au centre d'une sphère 3D.
- vingt-six caméras placées sur la surface de la sphère, à des intervalles réguliers.

Cette réduction est rendue possible grâce à l'alignement des objets 3D. En utilisant l'ACP de tous les objets de la même classe partagent la même pose, aucune rotation du light field n'est donc nécessaire.

Dans notre travail, nous simulons l'acquisition des light field à l'aide des fonctions de la bibliothèque OpenGL. Nous plaçons tout d'abord l'objet 3D d'extension $(.obj)$ dans un environnement 3D virtuel, cet objet est ensuite normalisé et aligné à l'aide des techniques expliquées dans le chapitre 4. La normalisation de l'objet permet de le placer dans une sphère englobante. Nous définissons ensuite vingt-six angles de prise de vue virtuels, ceux-ci permettent l'échantillonnage du plan (μ, ν) (le contour de la sphère). Chaque prise de vue est une image 2D de résolution 256×256 pixels, donc le plan (μ, ν) de notre dispositif produit $26 \times 256 \times 256 = 1,625$ Méga pixels. Sachant que chaque pixel est codé sur 4 octets, le plan (μ, ν) est échantillonné par 6,5 MO d'information. L'indexation des light field se contente des images acquises, aucun rendu du modèle 3D n'est réalisé, à cet effet, nous n'échantillonnons pas le plan (s, t) , seul l'échantillonnage du plan (μ, ν) nous suffira pour la caractérisation du light field. La comparaison entre notre dispositif et ceux des références citées ci-dessus montre clairement que la taille de l'information utilisée pour l'indexation des light field est nettement inférieure à celle du rendu. L'angle entre les différents points de prise de vue est nettement supérieur dans notre cas.

Algorithme de simulation d'acquisition des light field

Algorithme 5.1

Arguments d'entrée

O : Objet 3D maillé, normalisé et aligné selon l'axe des x.

Etape 1 : La mise de l'objet o à l'intérieur de la sphère.

Etape 2 : La pose d'une caméra virtuelle à chacun des vingt-six coins de la sphère. L'implantation de cette fonction est réalisée à l'aide de la bibliothèque OpenGL.

Sorties : Les projections 2D de l'objet 3D. La figure 5.9 représente l'environnement

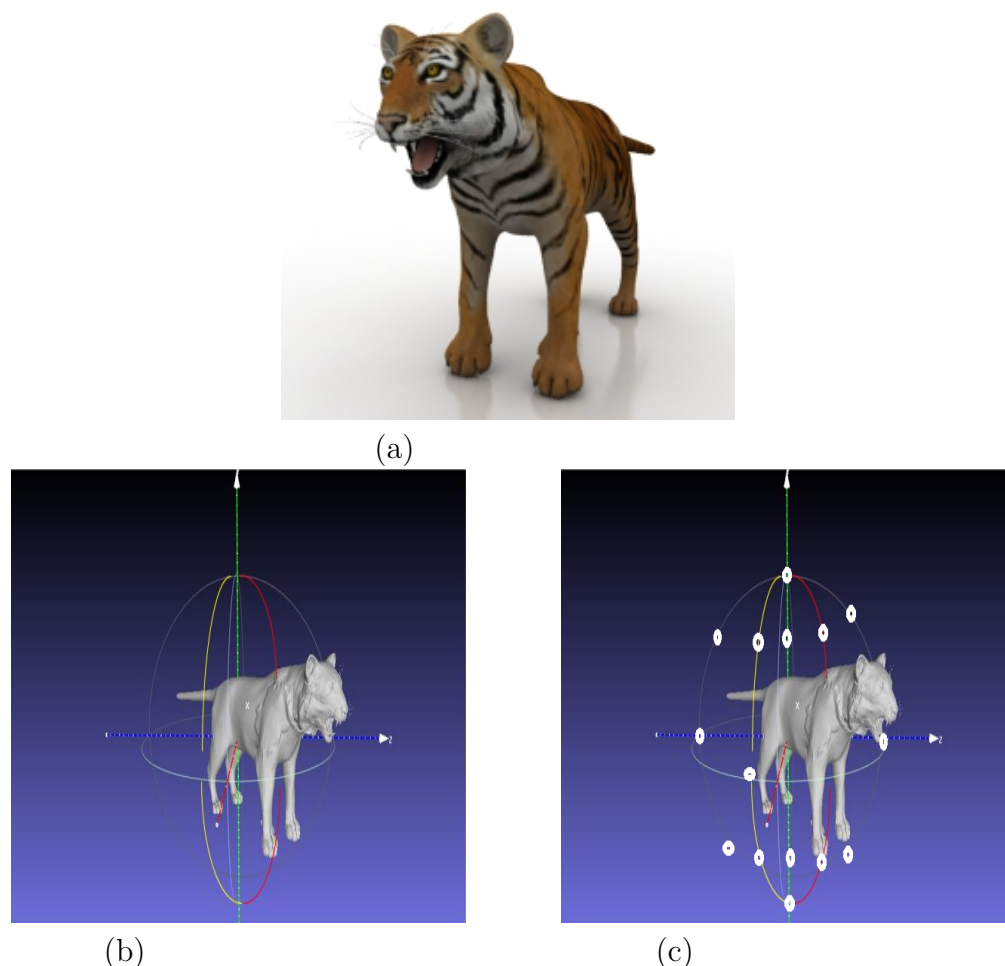


FIGURE 5.9 – *Illustration du simulateur d'acquisition des light field. (a) représente le modèle 3D à capturer. (b) représente le modèle 3D dans un environnement de capture sphérique. (c) Les anneaux blancs représentent les caméras virtuelles, qui sont réparties sur la surface de la sphère*

3D virtuel utilisé pour la capture des light field. Le plan dans notre cas est sous forme d'une sphère, les caméras de capture virtuelles sont positionnées sur quatre cercles reliant les sommets de la sphère et distancés de 45° l'un de l'autre, chaque cercle comprend six caméras, ajoutant à cela deux caméras aux deux sommets de la sphère. Le plan (s, t) dans notre cas n'est pas échantillonné.

5.2.2 Extraction des vues 2D

Notre approche d'indexation est basée sur les vues 2D constituant du light field, pour la construction de ses descripteurs. Il est possible d'obtenir les projections 2D du modèle 3D grâce à une étape d'extraction de ses vues 2D. Étant donné qu'un light field est constitué d'un nombre très conséquent de vues 2D, qui sont fortement corrélées, notre système d'indexation doit se contenter des vues les plus significatives de la sémantique du modèle 3D. Le choix des vues 2D doit impérativement prendre en compte certaines contraintes, qui garantissent une bonne correspondance entre les projections du modèle 3D et les différentes vues 2D. D'autre part le nombre de vues doit être le plus restreint que possible.

Cette extraction de vues est indispensable dans les systèmes d'indexation et de recherche de modèles 3D, du moment où une seule projection 2D n'est jamais suffisante pour représenter un modèle 3D. Elle peut être le plus souvent trompeuse, comme on peut le remarquer sur la figure 5.10. Le jeu avec les mains sur cette figure, montre clairement qu'on peut obtenir des formes différentes avec le même objet (la main). Autrement dit les influences du milieu d'acquisition peuvent avoir un effet négatif sur la sémantique de l'objet capturé. Par exemple, en cachant une partie d'un véhicule, sa forme peut ressembler à la forme d'un marteau. Afin d'éviter ce genre de situation, il faut extraire plusieurs vues et dans plusieurs directions. Cependant le plan de projection doit être le même à utiliser pour tous les modèles 3D.



FIGURE 5.10 – *Illustration des effets trompeurs qui peuvent être produits par les modèles 3D*

L'opération de projection est précédée par les opérations de prétraitement du modèle 3D du chapitre 4, ces opérations permettent d'aligner le modèle 3D dans un repère tridimensionnel. De plus le plan choisi pour la projection du modèle 3D, permet l'extraction de vingt-six vues de l'objet 3D. Chaque point de vue v_i est caractérisé par :

1. Son centre c_i . Les centres des points de prise de vue dans notre cas correspondent aux vingt-six sommets du dodécaèdre
2. Son point de visé ν_i . Ce point est le centre du repère 3D.
3. Son vecteur nord qui pointe vers la direction de la prise de vue \vec{n}_i .

Additivement à ces paramètres, nous utilisons les deux paramètres de l'échantillonnage de la fonction plénoptique (α et β). Ceux-ci dépendent des paramètres c_i , ν_i et \vec{n}_i . Par exemple : Le jeu de paramètres cités ci-dessous permet de définir les paramètres $\alpha = 0$ et $\beta = 0$:

$$c_1 = (100), \nu_1 = (000), \vec{n}_1 = (010). \quad (5.5)$$

Les caractéristiques des points de prise de vue (c_i , ν , n_i) sont utilisées pour générer le modèle 3D à l'aide de la librairie graphique 3D OpenGL. Cette librairie est donc capable d'afficher un modèle 3D à partir des caractéristiques des points de prises de vue. Grâce à cette application on peut également voir le modèle 3D de différentes directions, produisant ainsi les projections 2D selon un certain nombre de directions. Chaque caméra virtuelle utilisée pour obtenir les projections de la figure 5.12 est positionnée à l'aide de deux angles α et β . L'angle α contrôle le déplacement vertical de la caméra. L'angle β contrôle le déplacement horizontal des caméras (voir figure 5.11).

L'application de projection des objets 3D est basée sur les fonctions de la librairie OpenGL. Cette application consiste à :

- Représenter l'objet dans un environnement tri-dimensionnel.

Comme nous l'avons déjà expliqué au chapitre 2, les bases d'objets 3D dans notre travail, sont constituées de fichiers d'extension *.obj*. Les coordonnées de la triangulation des objets, sont exploitées par la fonction d'affichage, pour construire le modèle géométrique de l'objet. Le rendu consiste à coller la texture (qui est stockée dans d'autres fichiers), sur les polygones du fichier *.obj*.

- Créer une caméra virtuelle

La librairie OpenGL se dispose d'un ensemble d'outils pour la création d'une caméra virtuelle dans l'environnement 3D. Dans notre cas, nous plaçons cette caméra sur le contour de la sphère englobante, ceci à l'aide des paramètres c_i , ν et n_i . Les vues 2D capturées par cette caméra, représentent les projections perspectives de l'objet 3D. Les vingt-six projections sont ensuite obtenues, grâce au déplacement du caméra autour de l'objet, ce déplacement est contrôlé par les paramètres c_i , ν et n_i . La figure 5.12 montre le résultat de la projection de l'objet *éléphant*, à l'aide de notre algorithme de projection.

Ces vues ainsi obtenues sont ensuite caractérisées à l'aide de la méthode d'in-

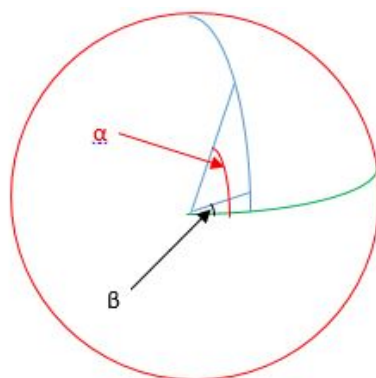


FIGURE 5.11 – Répartition des caméras sur la sphère de capture des vues 2D

dexation de vues 2D par la transformée en ondelettes basée région, qui a été détaillée dans le chapitre 3. Dans la section suivante, nous allons appliquer cette méthode aux light field, et nous montrons quelques exemples d’histogrammes obtenus à l’aide de cette méthode.

5.2.3 Caractérisation des light field

Dans la section précédente, nous avons présenté notre méthode d’extraction des vues 2D. Ces projections feront l’objet d’une caractérisation sémantique. Dans cette partie, chaque projection 2D est tout d’abord segmentée en deux régions R_1 et R_2 à l’aide de l’algorithme de la segmentation du papier [90] :

- R_1 représente l’objet ou la région d’intérêt.
- R_2 représente le fond de la projection 2D.

Contrairement à la méthode d’indexation présentée dans le chapitre 3, l’indexation des projections 2D portent seulement sur les régions d’intérêts. Le fond dans les modèles 3D n’a aucune signification.

Prenons par exemple deux objets différents Obj_1 et Obj_2 . Chaque projection 2D de cet objet fera tout d’abord l’objet d’une transformation en ondelettes. Les indices construits à partir des coefficients de cette transformée ne sont pas significatifs de la sémantique des deux objets. La caractérisation des projections dans ce cas s’est montrée peu discriminative. Ceci est dû aux régions de fond similaires dans les deux projections. La solution proposée dans ce travail est de ne transformer que les régions d’intérête des deux objets. Comme nous l’avons déjà expliqué dans le chapitre 3, cette solution consiste à segmenter les vues 2D en deux régions et d’appliquer la transformée en ondelettes basée région (SA-DWT) sur la région d’intérêt.

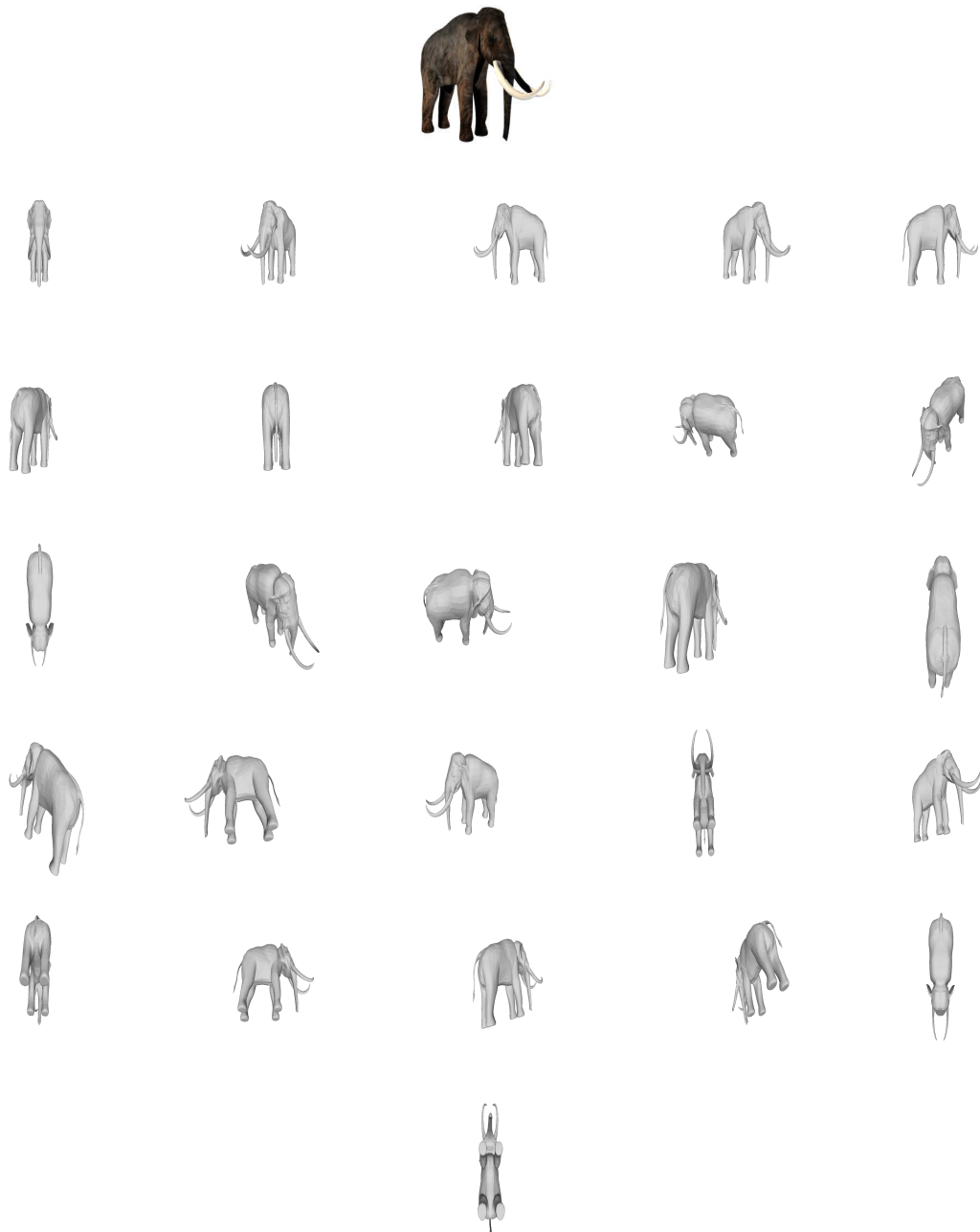


FIGURE 5.12 – Résultat de la projection 2D d'un objet 3D.

Les coefficients basse fréquence de la transformée en ondelettes sont ensuite utilisés pour la caractérisation de la couleur des projections 2D. La texture est caractérisée par les coefficients haute fréquence de la transformée en ondelettes. Le contour est caractérisé à l'aide de la méthode *DLEP*. À l'instar de la méthode d'indexation du chapitre 3, chaque projection est donc caractérisée par :

- Un indice de couleur de 400 bins.
- Un indice de texture de 3000 bins.

– Un indice de contour de 100 bins.

Chaque modèle 3D est donc caractérisé par 26 histogrammes de couleur de 400 bins chacun, 26 histogrammes de texture de 3000 bins chacun, et 26 autres histogrammes de contour de 100 bins chacun.

5.2.3.1 Caractérisation de la couleur

Les histogrammes de couleur offrent une indexation robuste aux occlusions, aux déplacements et aux rotations des formes 3D [102]. Dans notre cas, l’histogramme de l’attribut couleur représente l’occurrence des patterns couleur. Ces histogrammes sont construits à partir des coefficients de la sous bande basse fréquence de la transformée en ondelettes (5.13 (c, d)). Dans ces deux exemples, l’indice de la couleur est composé de 400 bins. Ces indices caractérisent la distribution de la primitive couleur des pixels de la projection 2D ($\alpha = 0^\circ, \beta = 0^\circ$), des objets *avion 1* et *avion 2*.

Il est connu que la sous bande basse fréquence de la transformée en ondelettes, représente l’approximation en niveaux de gris de l’image originale [17]. Après l’extraction des projections 2D de l’objet 3D, l’attribut couleur peut alors être caractérisée, par les coefficients basse fréquence de la transformée en ondelettes. Chaque projection 2D sous format JPEG est tout d’abord convertie en trois composantes Y (luminance), C_b, C_r (composantes de chrominance). Ces trois composantes sont ensuite transformées à l’aide de la transformée en ondelettes basée région. Les patterns de couleur dans notre cas, ne sont que l’association trois à trois des coefficients basse fréquence, des composantes Y, C_b, C_r , selon l’algorithme 3.1 du chapitre 3.

En dépit de toutes les mesures prises lors de la normalisation des espaces de la couleur, celle-ci peut toutefois apparaître insuffisante pour la caractérisation de tous les objets de la base. En effet, la figure 5.13 montre que l’attribut couleur pourra perturber le processus de caractérisation. Deux objets similaires pourront le plus souvent être de couleurs différentes. Quoiqu’ils appartiennent à la même classe, les objets *avion 1*, *avion 2* de la figure 5.13 ne partagent pas la même distribution de couleurs. Ceci s’est traduit par deux histogrammes de couleur partiellement différents, laissant apparaître une incapacité de l’attribut couleur à caractériser l’aspect sémantique de ces deux objets.

Dans la figure 5.13, nous avons entouré les intervalles des patterns de couleur similaires des deux histogrammes (c) et (f) par des cercles verts, les intervalles des patterns de couleur différents sont limités par des cercles bleus. La figure montre que les niveaux de couleur de la projection de l’objet (b) sont plus regroupés et plus centrés autour de quelques patterns, contrairement aux patterns de la projection

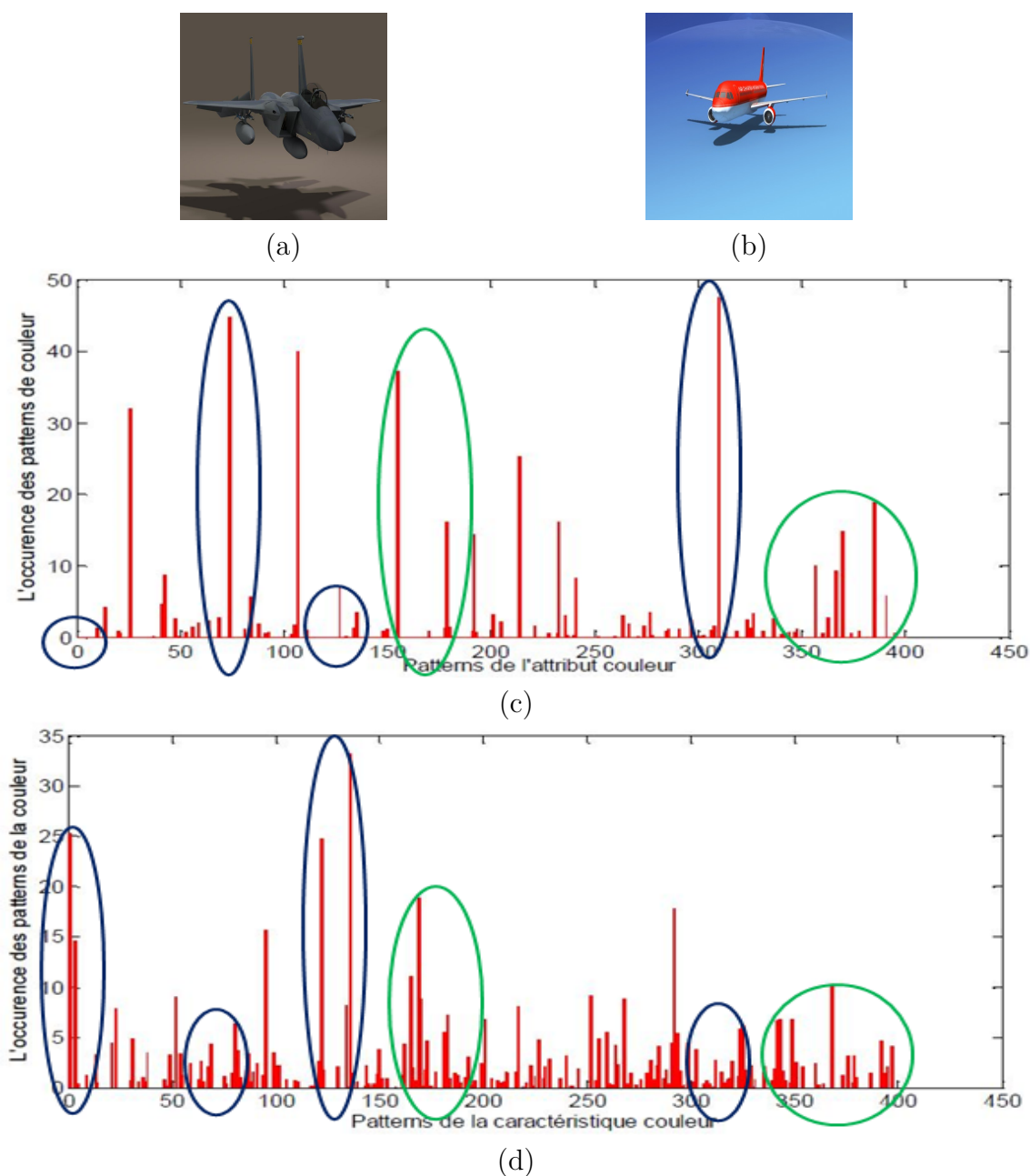


FIGURE 5.13 – Exemples d’histogrammes de couleur de la composante Y des vues 2D d’objets 3D. (a, b) représentent les objets 3D. (c, d) représentent les histogrammes de la couleur des composantes Y , des projections des objets 3D (a, b) respectivement.

de l’objet (a), qui sont plus dispersés. Ceci s’est traduit par un indice de couleur (c) moins dense que (d), autrement dit la distribution des niveaux de couleur entre les deux histogrammes est très différente. Cette analyse explique la différence qui existe entre les plages de couleur des deux histogrammes. Une des solutions de la caractérisation des objets identiques de couleurs différentes, réside dans l’emploi de la notion *espace* pour la caractérisation des objets 3D. À cet effet nous proposons

dans la section suivante, une méthode basée sur la caractérisation de l'attribut forme des objets 3D. Celle-ci étant robuste aux changements de couleur, de la luminosité et de l'ombre du milieu d'acquisition.

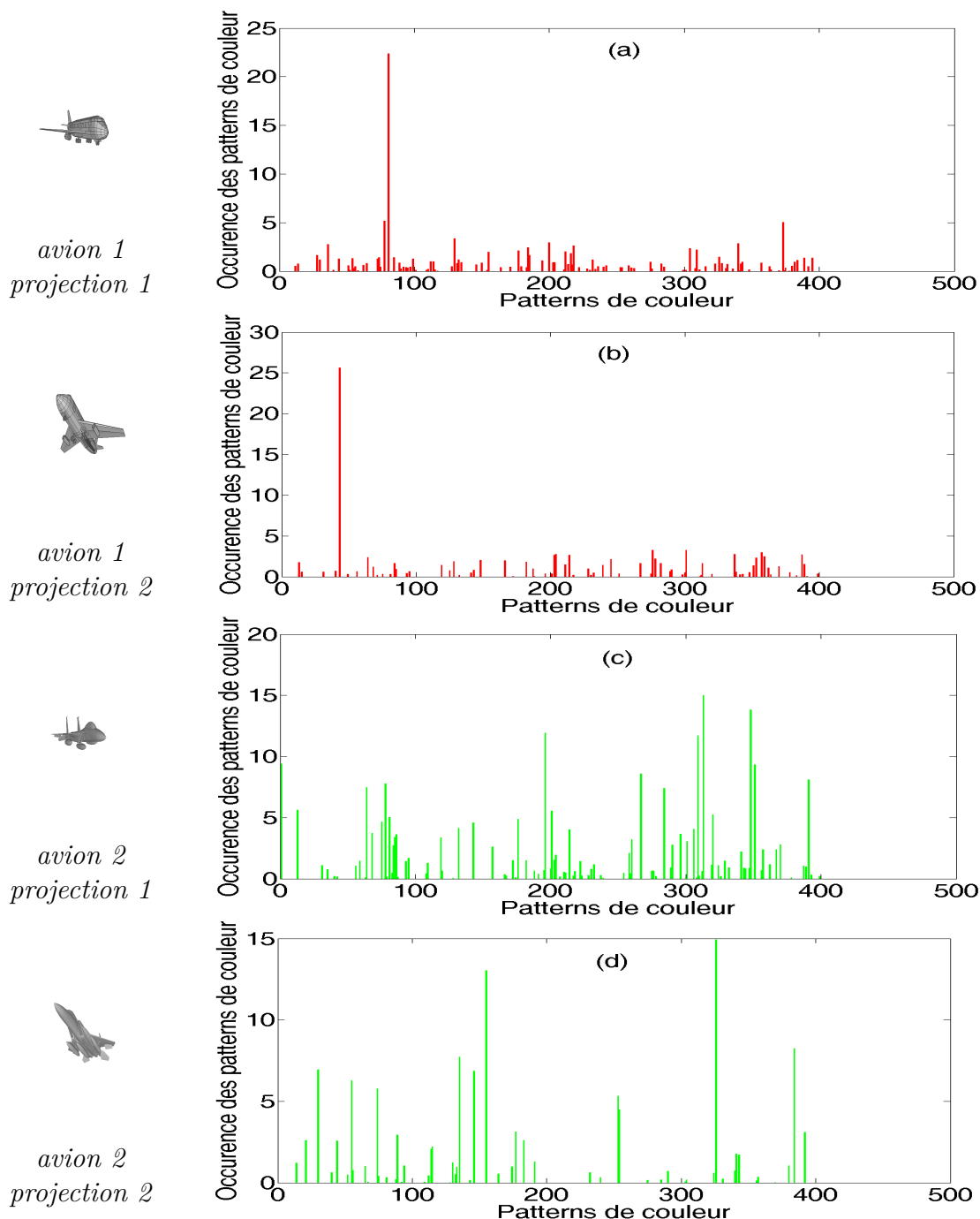


FIGURE 5.14 – Illustration de l'incapacité de l'attribut couleur à caractériser des objets de la même classe et de différentes couleurs. a, b, c et d représentent les indices couleur des projections de l'objet avion1, avion2 à $(\alpha=135^\circ, \beta=0^\circ)$ et $(\alpha=45^\circ, \beta=45^\circ)$ respectivement.

Jusqu'à présent, on a testé l'efficacité de l'indice de couleur sur des objets de mêmes classes. Dans la section précédente, nous avons vu que cet attribut arrive aussi bien que mal à traduire la ressemblance entre les objets similaires. Cette efficacité de caractérisation a vite diminué quand on a affaire à des objets similaires de différentes couleurs. Dans cette section, nous proposons également d'évaluer l'efficacité de l'indice couleur à différencier entre des objets de la même couleur, et de formes différentes. En d'autres termes des objets appartenant à différentes classes, mais qui partagent le même attribut couleur. Ainsi, pour chaque objet de la figure 5.15 nous construisons deux indices de deux projections différentes. Les indices obtenus semblent avoir un important degré de ressemblance, qui dépasse même celui des deux objets similaires. Devant cette impuissance, il est devenu nécessaire de trouver d'autres attributs, qui mettent en valeur les facteurs communs des objets similaires et qui ont une plus grande faculté de marginalisation en cas de caractérisation d'objets différents.

5.2.3.2 *Caractérisation du contour*

Dans notre travail, la caractérisation du contour consiste à étudier la variation des niveaux de gris autour de ses pixels. Comme il est illustré sur la figure 5.16, l'histogramme du contour est constitué de 100 bins pour une seule projection 2D. Cet histogramme est le moins dense en termes du nombre de patterns. La caractérisation de l'attribut contour d'une projection 2D, permet de décrire le sens de la variation des niveaux de gris sur le contour de l'objet. Après la transformation à l'aide de la SA-DWT de la projection 2D, on étudie la distribution des niveaux de gris des blocs appartenant au contour. Seuls les blocs du coefficient LL de la composante Y , sont concernés par cette analyse. En fait chaque bloc du contour est caractérisé par un pattern de quatre coefficients. Chaque coefficient est la représentation décimale d'une séquence de neuf bits, chaque bit est égal à 1 si la variation des niveaux de gris suit la même direction. Sinon on attribue un 0 dans le cas contraire. Dans cette section nous n'allons pas détailler davantage cette technique, qui est déjà expliquée dans l'algorithme 3.3 du chapitre 03. Nous nous intéressons plutôt aux facultés discriminatives et descriptives de l'indice du contour.

La figure 5.16 montre que la distribution des patterns du contour des projections des objets (a) et (b) est partiellement différente, on peut constater que les patterns appartenant à l'intervalle $[0\ 20]$ de l'histogramme (c), ont une occurrence presque nulle, tandis que ce même intervalle se distingue par de fortes occurrences dans l'histogramme d. À degrés plus ou moins importants, les occurrences des patterns de l'intervalle $[20\ 100]$, sont majoritairement comprises entre 4 et 15. À cet effet, la distance entre les occurrences (qui sont représentées par des cercles bleus)

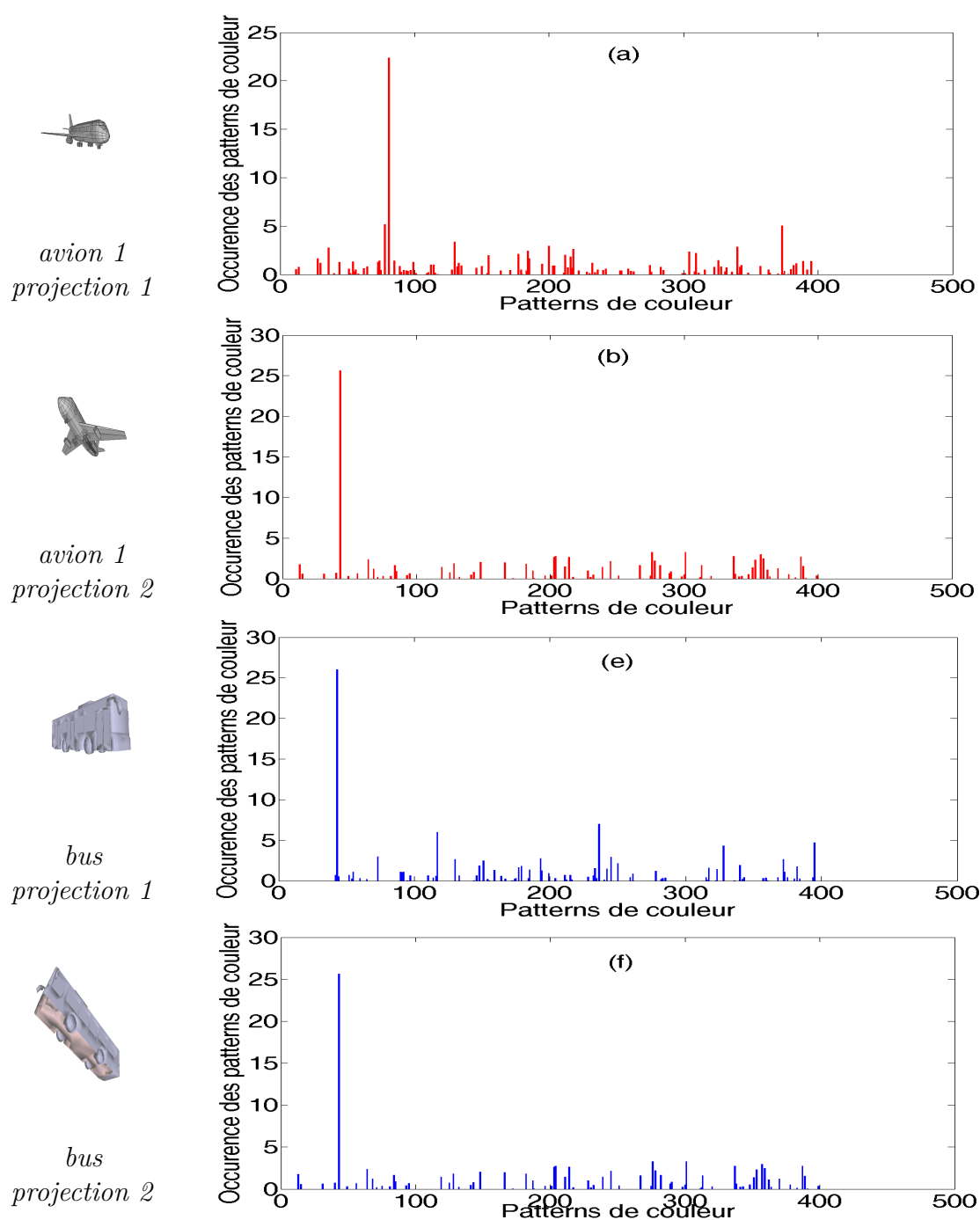


FIGURE 5.15 – *Caractérisation partielle (seulement de deux projections 2D) de deux objets de classes différentes . a, b, c et d représentent les indices couleur des projections de l'objet avion1, bus à $(\alpha=135^\circ, \beta=0^\circ)$ et $(\alpha=45^\circ, \beta=45^\circ)$ respectivement.*

dans cet intervalle, n'est pas si importante. Enfin, on peut dire que l'ajout de l'attribut contour a eu un impact positif sur les performances de classification de notre algorithme. Contrairement à l'attribut couleur, cet attribut se montre plus performant pour la caractérisation de la sémantique de l'objet. Afin d'étudier les pouvoirs

discriminatifs de cet attribut, nous proposons dans la section suivante, de caractériser la forme d'un ensemble d'objets de différentes sémantiques.

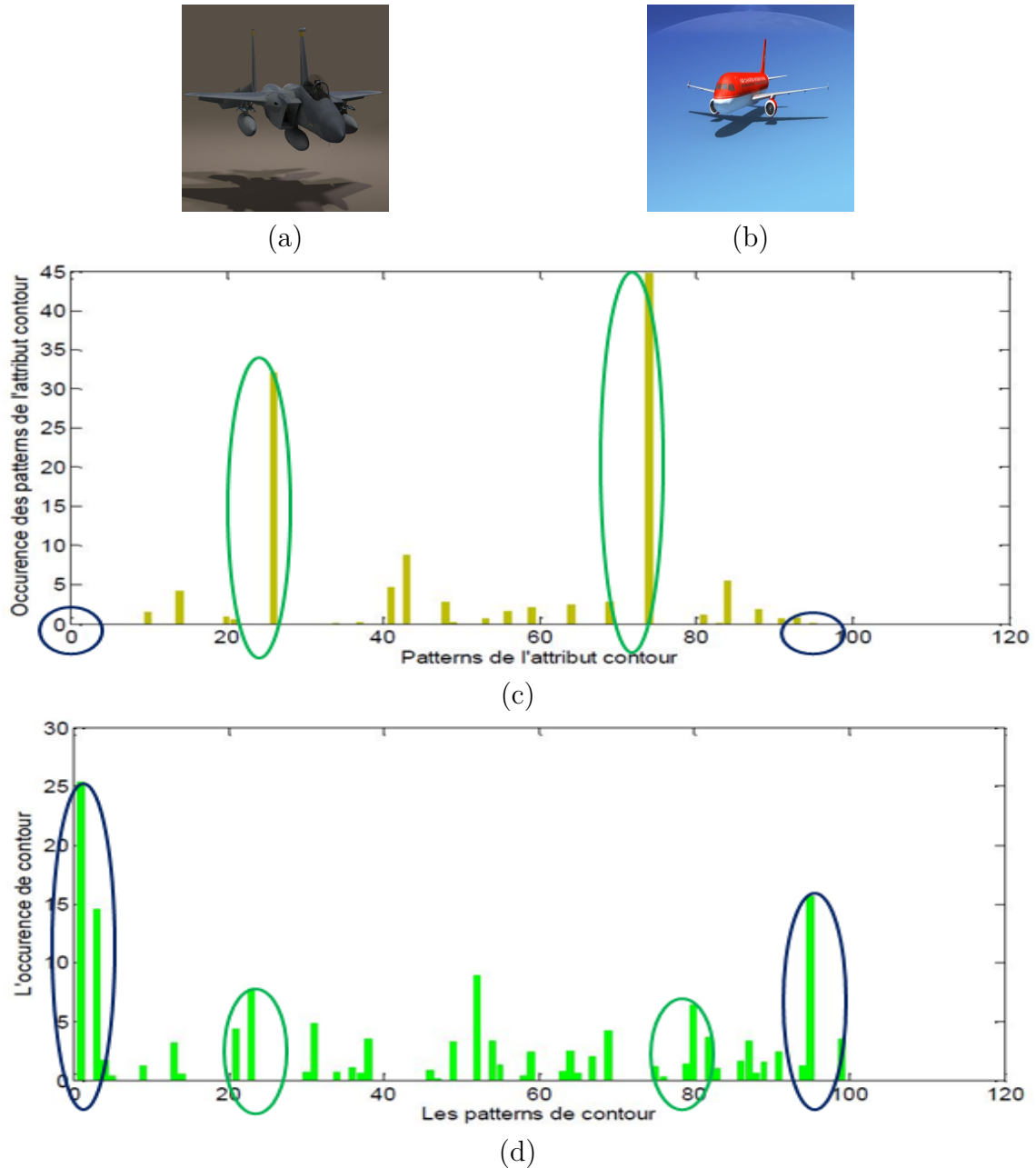


FIGURE 5.16 – Exemples d'histogrammes de contour de la composante Y des vues 2D d'objets 3D. (a, b) représentent les objets 3D. (c, d) représentent les histogrammes de contour, de la composante Y des projections des objets (a) et (b) respectivement.

Les figures 5.17 et 5.18 représentent les indices du contour de deux projections des objets :

- Avion 1.

- Avion 2.
- Bus.

Les projections ont été capturées par la caméra virtuelle, à partir des angles de vue :

- A_1 ($\alpha = 135^\circ$, $\beta = 0^\circ$) : projection 1.
- A_2 ($\alpha = 45^\circ$, $\beta = 45^\circ$) : projection 2.

Il est clair que les patterns du contour des deux projections de l'objet *avion 1* et *avion 2*, se caractérisent par une distribution similaire de leur occurrences respectives (voir figure). L'attribut contour dans cet exemple a permis de différencier entre l'objet *bus*, et les objets *avion 1* et *avion 2*, sous différentes conditions de capture (voir figure). Un autre avantage de cet attribut concerne la taille de ses histogrammes, qui ne dépasse pas les 100 bins. Ceci permet d'accélérer la recherche dans les bases images, et de simplifier les comparaisons entre ces images.

En l'absence de l'occlusion, l'attribut contour s'avère beaucoup plus discriminant que l'attribut couleur. Nous nous intéressons à présent à l'influence de l'occlusion sur la qualité discriminative de l'attribut contour. Cette analyse porte sur la comparaison de deux objets identiques de la classe *avion*. Comme le montre la figure, la projection 1 de l'objet *avion 1* est partiellement cachée par un autre objet. Ceci s'est traduit par la diminution de l'occurrence du *pattern 20*, qui est égale à 100 dans la figure et dans cette figure sa valeur est diminuée à 38. Lors de la phase de comparaison, ce changement de valeurs des occurrences conduira à une augmentation de la distance entre les deux objets *avion 1* et *avion 2*.

On peut également constater que l'histogramme du contour de la projection 2 du même objet, n'a enregistré aucun changement. À cet effet, nous pouvons dire que la comparaison entre les projections 2 des deux objets *avion 1* et *avion 2*, est plus appropriée. Ceci, nous a permis de proposer une première solution au problème de l'occlusion, qui consiste à comparer toutes les vues des deux objets et ne garder que la plus faible distance. Nous proposons également l'utilisation de la texture, qui est robuste vis-à-vis le problème de l'occlusion.

5.2.3.3 Caractérisation de la texture

L'histogramme de la caractéristique texture est la combinaison de trois histogrammes représentant la distribution des coefficients de détails dans les sous bandes *HL*, *LH* et *HH*. La construction de l'histogramme de texture des différentes projections d'un objet 3D est présentée dans le chapitre 3. La figure (5.20 (c), (d)), montre des exemples d'histogrammes de texture de deux projections identiques de deux objets de la classe *avion*. En comparaison aux attributs contour et couleur, les histogrammes de l'attribut texture représentent la même distribution des patterns de texture. Autrement dit l'attribut texture possède un important pouvoir de clas-

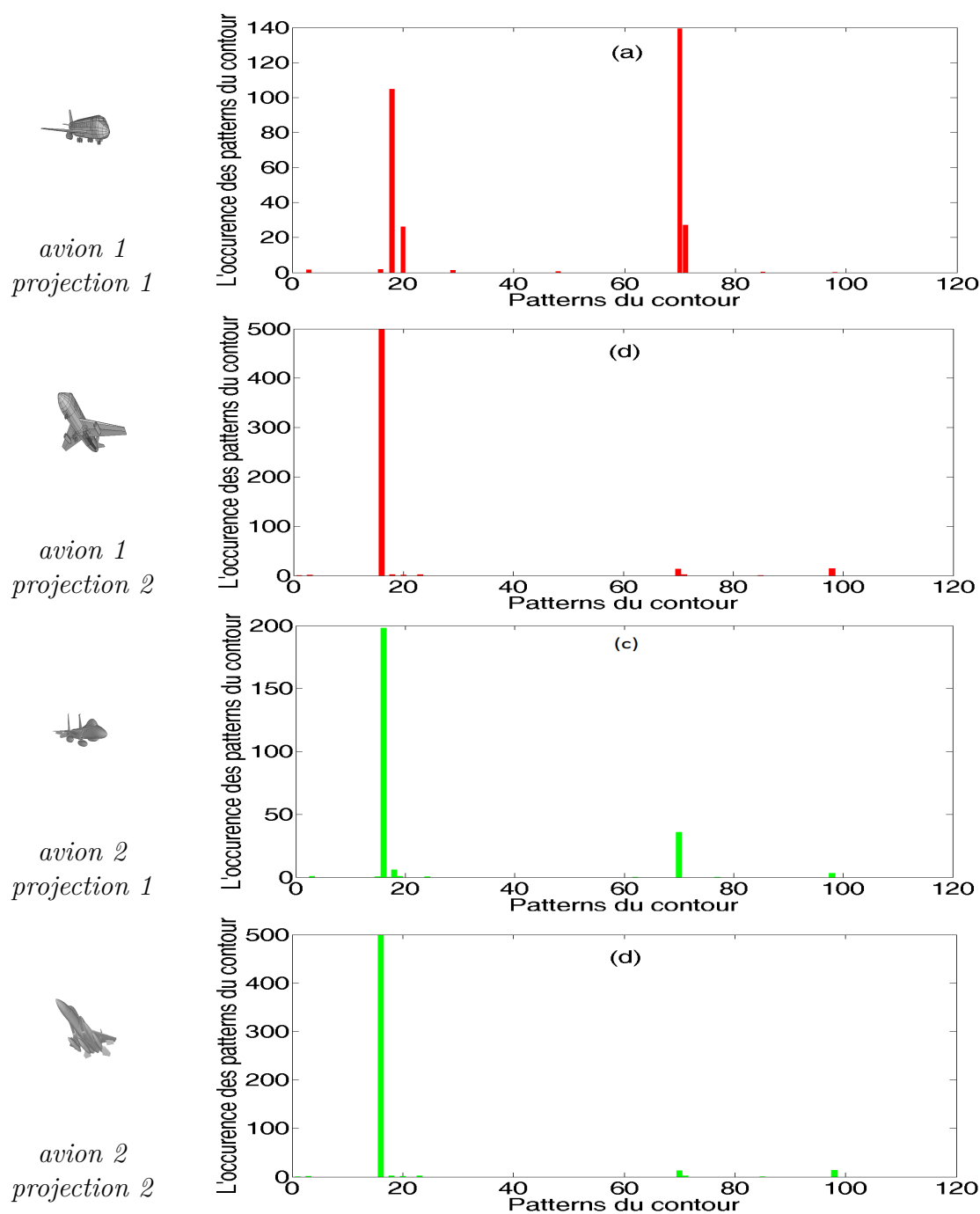


FIGURE 5.17 – Caractérisation partielle (seulement de deux projections 2D) de la classe avion à l'aide de l'attribut contour. a, b, c et d représentent les indices contour des projections avion1, avion2, des objets avion1 et avion2. à $(\alpha=135^\circ, \beta=0^\circ)$ et $(\alpha=45^\circ, \beta=45^\circ)$ respectivement.

sification.

L'histogramme de texture de la figure 5.21(a) est constitué de trois histo-

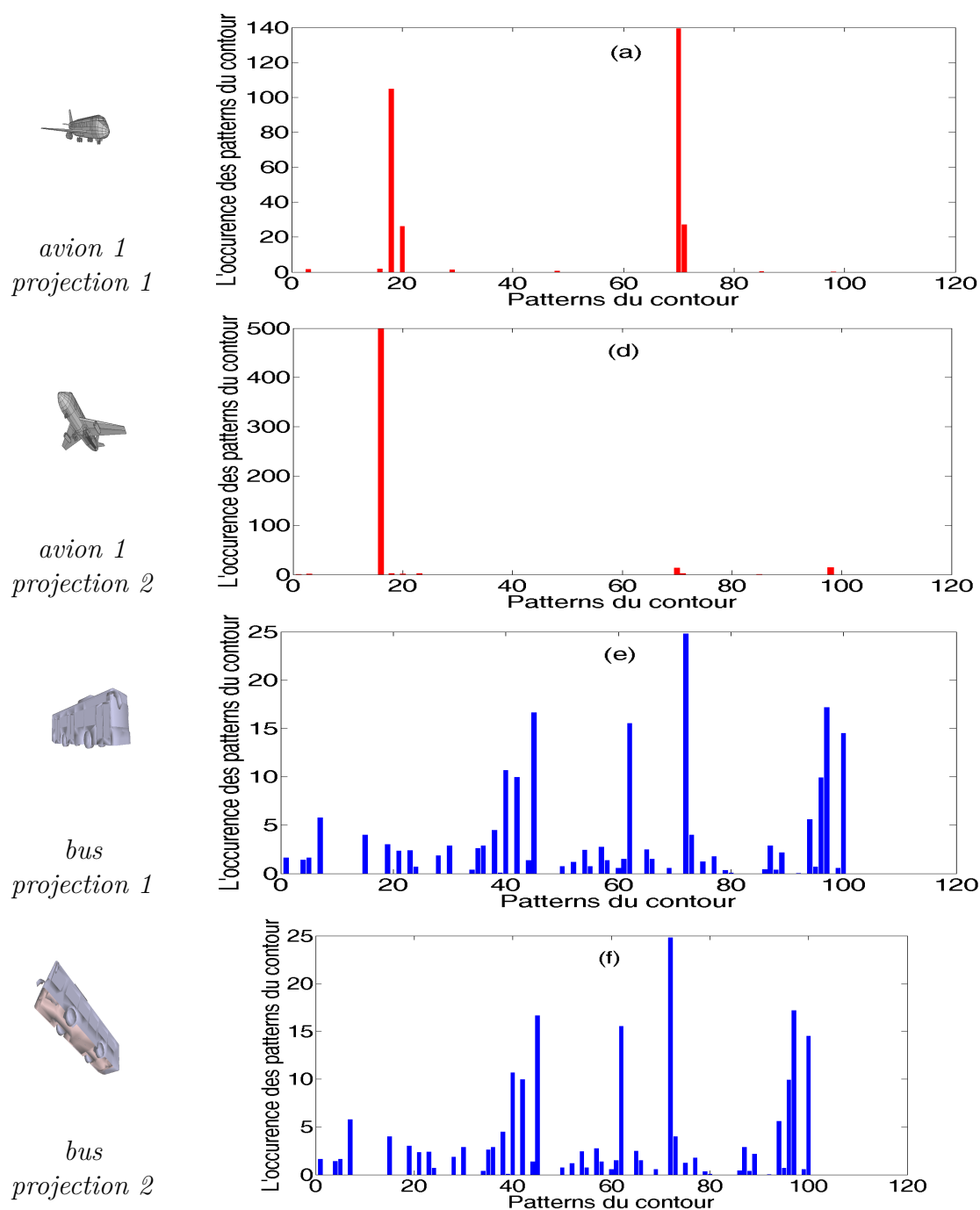


FIGURE 5.18 – Caractérisation partielle (seulement de deux projections 2D) de deux objets de classes différentes . a, b, c et d représentent les indices du contour des projections de l'objet avion1, bus à $(\alpha=135^\circ, \beta=0^\circ)$ et $(\alpha=45^\circ, \beta=45^\circ)$ respectivement.

grammes H_{iSHL} , H_{iSLH} et H_{iSHH} . La comparaison de cet histogramme aux histogrammes de texture des objets *avion 1* et *avion 2*, révèle la différence entre les objets des deux classes. Cette différence est concentrée dans les patterns compris

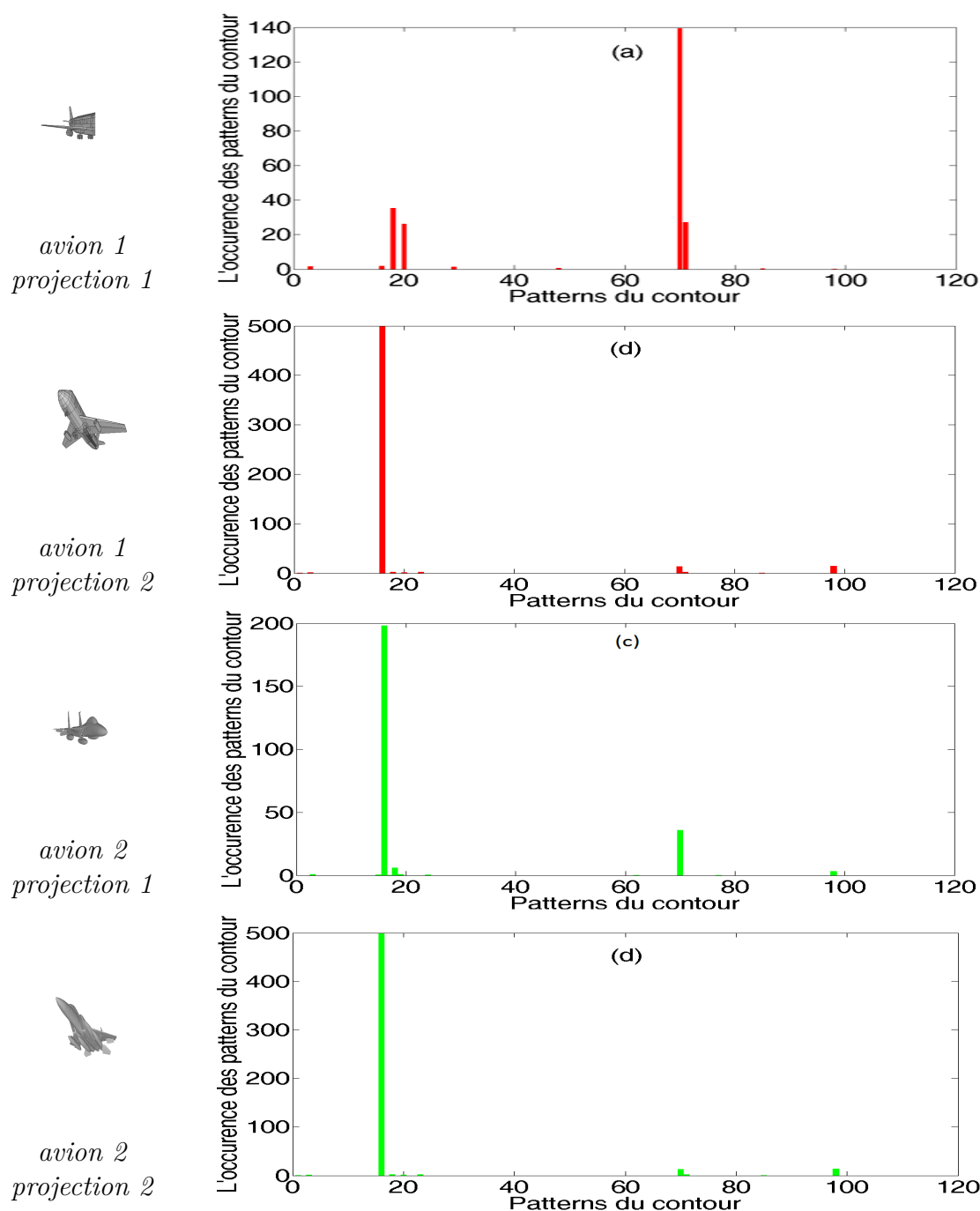


FIGURE 5.19 – Illustration de l'effet de l'occlusion sur l'indice du contour. *a*, *b*, *c* et *d* représentent les indices contour des projections avion1, avion2, des objets avion1 et avion2. à $(\alpha=135^\circ, \beta=0^\circ)$ et $(\alpha=45^\circ, \beta=45^\circ)$ respectivement.

entre 2000 et 3000, qui correspondent à l'histogramme His_{HH} . Autrement dit la différence entre les classes *bus* et *avion*, vient de l'orientation horizontale de leur détails. Ce même histogramme était le seul, à être influencé par l'occlusion d'une petite partie de l'objet *avion 1*. Néanmoins cet impact est presque négligeable. À

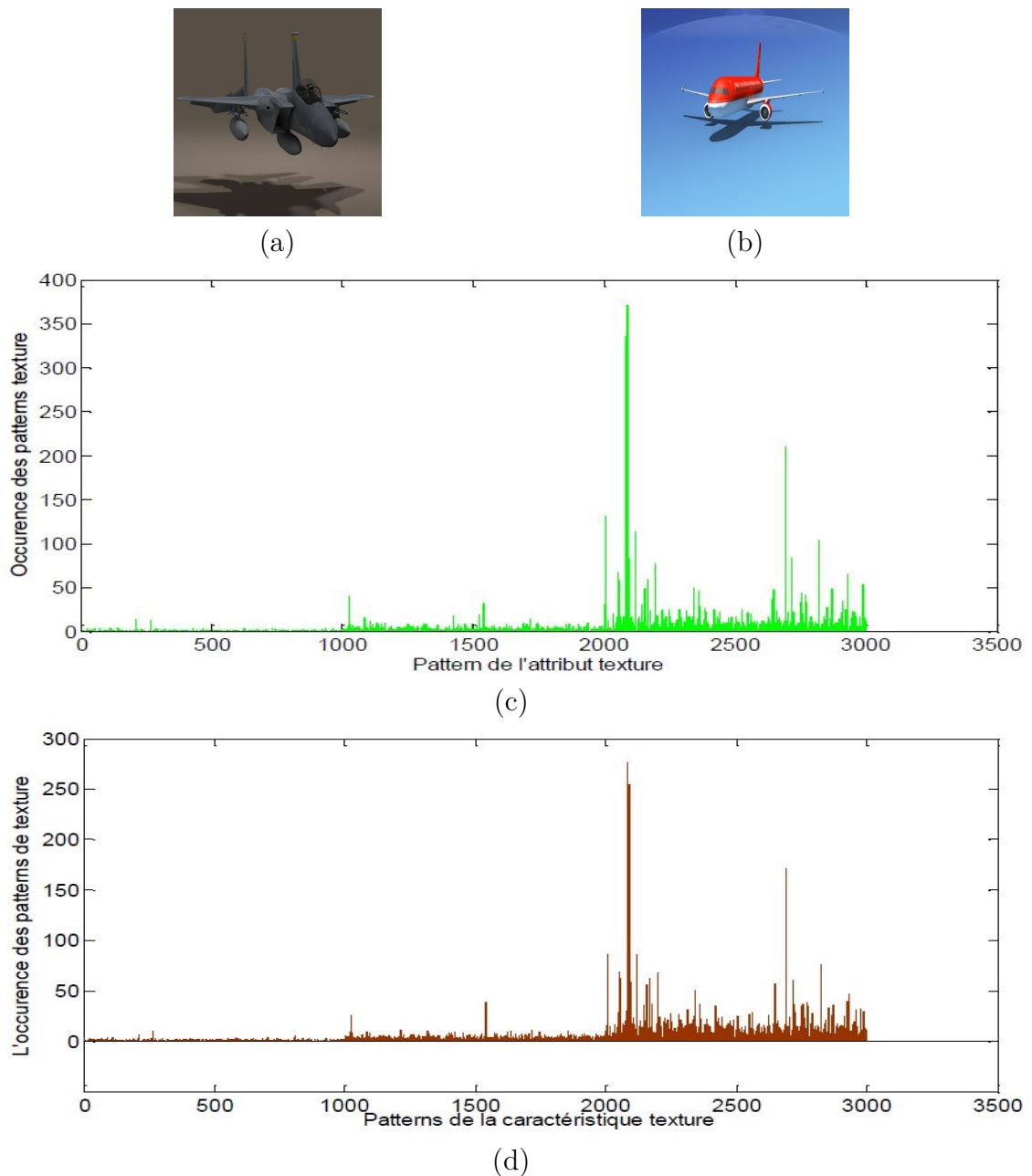


FIGURE 5.20 – *Caractérisation partielle (seulement de deux projections 2D) de la classe avion à l'aide de l'attribut texture. a, b, c et d représentent les indices texture des projections avion1, avion2, des objets avion1 et avion2. à $(\alpha=135^\circ, \beta=0^\circ)$ et $(\alpha=45^\circ, \beta=45^\circ)$ respectivement.*

cet effet, nous pouvons dire que l'histogramme de texture est plus robuste vis-à-vis l'occlusion que l'histogramme du contour.

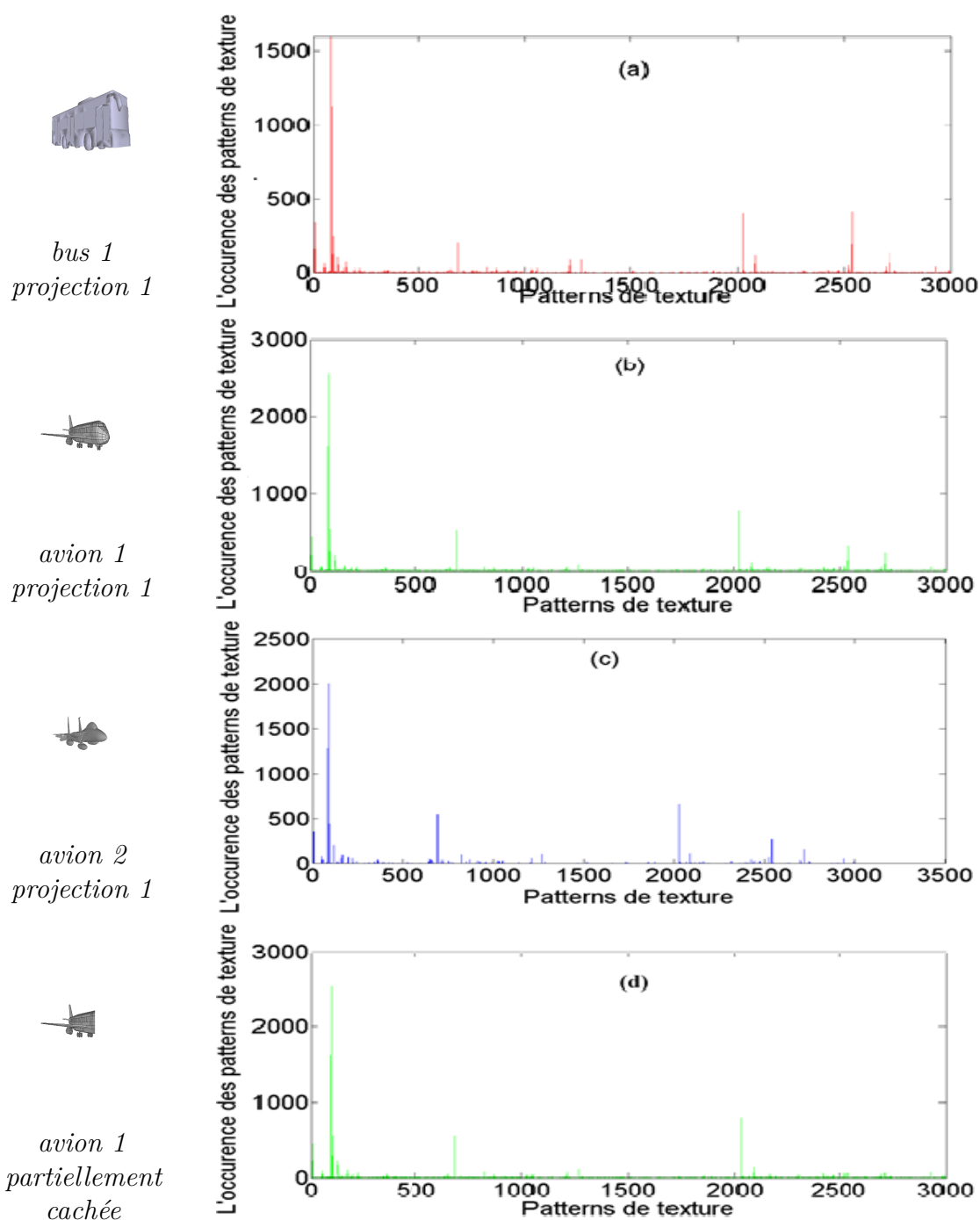


FIGURE 5.21 – *Caractérisation de l'attribut texture. a, b et c représentent les indices de l'attribut texture des projections avion1, avion2 et bus respectivement. (d) représente l'indices de l'attribut texture d'une projection partiellement cachée de l'objet avion1*

5.3 Conclusion

Dans ce chapitre nous avons présenté la méthodologie de la construction des indices de couleur, contour et de texture. Etant donné que l'indexation des vues

2D par la transformée en ondelettes a été déjà abordée dans le chapitre 3, dans ce chapitre on s'est intéressé beaucoup plus aux traitements d'images, permettant le passage du 3D aux vues 2D. À cet effet, nous avons proposé notre système de projection des modèles 3D. Le choix du nombre de caméras de capture est basé sur une étude empirique que nous allons la présenter dans le chapitre 7. Les histogrammes construits à l'aide des trois indices montrent clairement que les indices de texture offrent de bons résultats par rapport aux indices de couleur et de contour. Ces résultats vont être exploités lors de la construction de notre système de recherche des light field.

Chapitre 6

Recherche Des Light Field

Résumé : Cette partie présente une solution au problème de la recherche dans les bases de modèles 3D. La complexité de ce sujet, réside dans la lenteur de la recherche. Chaque modèle de la base des objets 3D, fera l'objet de plusieurs comparaisons. Une autre difficulté viendra cette fois-ci du choix de la métrique, qui est à la base de l'évaluation des similarités. À cet effet, une grande partie de ce chapitre est consacrée à l'accélération et à la simplification du processus de la recherche. Nous nous intéressons également au choix de la métrique de mesure de similarité.

6.1 Introduction

Les descripteurs de modèles 3D présentés dans le chapitre 05, permettent d'extraire l'information sémantique de chaque modèle. Chaque objet 3D se voit donc associé à un ensemble de descripteurs (descripteur de couleur, descripteur de contour et descripteur de texture). La recherche dans les bases d'objets 3D, est initiée par un modèle exemple ou une requête fournie par l'utilisateur au moteur de recherche. Cette image exemple peut se présenter sous plusieurs formes : une vue 2D du modèle à trouver, un dessin ou un croquis ou même un modèle tridimensionnel. Dans les moteurs de recherche par l'exemple, le choix de la requête guide complètement l'opération de la recherche dans la base d'image. Contrairement à la phase d'indexation, qui est accomplie en hors ligne, la phase de recherche est basée sur une mesure de similarité ou de dissimilarité en ligne, entre la requête et les images de la base.

L'indexation des objets 3D est une question centrale dans la conception des moteurs de recherche. Une caractérisation adéquate des objets de la base d'images, peut améliorer significativement les performances de n'importe quel système de recherche d'images. Ceci revient à dire que les phases de recherche et d'indexation de l'image sont indissociables. Au lieu de comparer toutes les images de la base une par une à l'image requête, la phase d'indexation nous offrent la possibilité de comparer seulement les signatures associées à ces images. Ceci permet de gagner en termes de temps et de complexité des calculs de comparaison et de recherche. La difficulté qui se pose pour ce type d'approches vient de la nécessité d'achever la recherche le plus rapidement possible. Or, cette contrainte est trop forte pour que l'on puisse se satisfaire des temps enregistrés par les approches d'indexation et de recherche actuelles. Comme on l'a déjà précisé dans le chapitre précédant, l'inconvénient majeur de l'indexation des objets 3D est la taille intéressante des indices. En utilisant trois histogrammes pour caractériser les projections 2D, qui sont au nombre de vingt-six pour chaque objet 3D, le temps des comparaisons est très intéressant. Ceci, nous a contraints à simplifier la procédure de la recherche et de la comparaison des indices. Ce chapitre traite également le choix de la métrique pour mesurer les distances entre les indices.

Après avoir traité la question de l'indexation des light field dans le chapitre 05, nous pouvons maintenant nous tourner vers la recherche des light field dans les bases d'objets 3D. Dans la première partie de ce chapitre, nous nous concentrons sur le processus général de la phase de recherche. On va présenter le schéma général de notre méthode de recherche dans cette section. La seconde partie de ce chapitre se focalise sur la caractérisation de la requête fournie par l'utilisateur. L'indexation de la requête commence par la normalisation et la projection de celle-ci, selon le

modèle des light field. On applique ensuite la méthode d'indexation proposée dans le chapitre 05, pour la construction des indices de la requête. Nous passons ensuite à la présentation de notre stratégie hiérarchique de recherche dans la base d'images. Enfin, nous nous intéressons à l'aspect choix de la métrique de comparaison.

6.2 Stratégie de recherche des light field

La limitation principale des moteurs d'indexation et de recherche des modèles 3D, concerne la comparaison entre les différents indices de la base et l'indice du modèle requête. Le calcul de similarité entre deux images à signature multiples peut être très lent, suivant le degré de complexité des signatures. Dans notre cas par exemple, chaque signature est composée de trois histogrammes ((Un histogramme de texture, un histogramme de couleur, et un histogramme de contour = trois histogrammes pour chaque vue de la structure des light field), chaque light field est composé de 26 vues \implies 78 histogrammes pour chaque light field ou modèle 3D). Ce nombre de signatures est à l'origine de la complexité des calculs de comparaison entre les différents indices. Alourdissant ainsi le processus de recherche.

Un deuxième inconvénient concerne cette fois-ci les signatures utilisées elles même. Deux modèles 3D peuvent avoir la même forme ou descripteur de forme, sans pour autant qu'ils aient la même sémantique. Remarquant d'emblée que cette façon d'utiliser les indices est inappropriée au processus de comparaison, Yongjian et al dans [77] ont proposé de réduire l'espace de recherche de modèles 3D, leur approche est basée sur la construction d'un ensemble de modèles significatifs. Autrement dit, ils proposent la réorganisation de la base de modèles 3D, avant de passer à l'étape d'indexation. L'algorithme proposé dans ce travail est composé de quatre parties :

- L'arbre de catégorisation des modèles 3D de la base, cet arbre est utilisé pour organiser la base d'objets 3D et pour extraire le sous ensemble d'objets significatifs.
- Le sous ensemble d'objets significatifs, qui reflète les caractéristiques sémantiques des formes de la base.
- Le fichier d'indices des vues 2D des objets 3D de la base.
- L'outil de recherche qui est composé de l'interface graphique et de l'outil de comparaison requête – modèles de la base.

La construction de l'arbre de structuration de la base, est basée sur l'algorithme [117]. Chaque nœud de cet arbre est représenté par un objet dans le sous ensemble des objets représentatifs de la base (voir figure 6.1). La notion de recherche hiérarchique est concrétisée dans ce travail, par un premier niveau de recherche basée sur le sous ensemble des objets représentatifs. L'objet résultant de cette phase est

ensuite utilisé comme requête d'une deuxième recherche dans la base d'objets 3D.

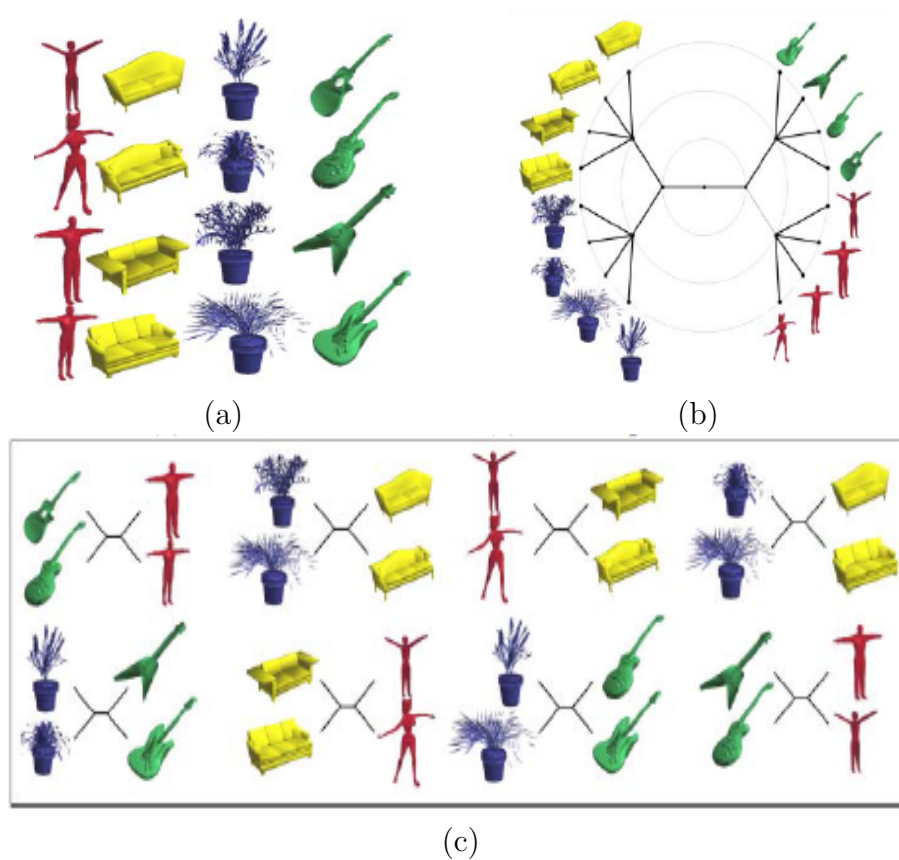


FIGURE 6.1 – Exemple de l'arbre de catégorisation et du sous ensemble d'objets représentatifs. (a) Échantillons des objets de la base. (b) L'arbre de catégorisation. (c) Regroupement des objets en sous ensembles (images extraites de [77])

dans le même sens Hau-San Wong et al ont proposé une autre méthode d'organisation de la base de modèles tridimensionnels. Les auteurs ont proposé une nouvelle méthode de recherche d'objets 3D. La taille des données de la base d'objets est réduite, grâce à l'utilisation de l'analyse en composantes principales (Kernel ACP), celle-ci est également utilisée pour extraire les caractéristiques des modèles 3D [140]. Ces caractéristiques sont ensuite utilisées, pour construire une structure hiérarchique d'indexation des objets 3D de la base. Cette étape est basée sur l'algorithme *HSOM* (*Hierarchical Self Organizing Map*) [71]. Cet algorithme permet la représentation d'une base d'objet 3D de n groupes, par un seul groupe, lequel est caractérisé par un vecteur de poids de ses objets. Lors de la phase de recherche, l'indice de la requête est comparé aux indices de ce groupe. Simplifiant ainsi le processus de comparaison [140].

Dans notre travail, nous avons proposé de simplifier le processus de comparaison, et de réduire la taille des indices. D'une manière générale, on peut résumer notre contribution dans les points suivants :

- La hiérarchisation du processus de recherche : nous proposons un processus de recherche à deux niveaux hiérarchique. Le premier niveau de comparaison est basé sur l'utilisation des indices de contour et de couleur, autrement dit la comparaison dans le premier stade est effectuée entre les indices de couleur et de contour des modèles de la base et ceux du modèle requête. À l'issue de cette étape, un premier ensemble de modèles résultants S_1 est construit. L'ensemble ainsi composé, constituera l'espace de recherche au deuxième niveau. Contrairement aux indices de la première étape, les indices de texture se distinguent par leurs tailles et leur précision de caractérisation. Afin de limiter le temps de comparaisons à l'aide de ces indices, nous proposons de réduire la zone de recherche et de comparaison. À l'issue de la première phase de recherche, l'ensemble de comparaison est réduit à trente-cinq pourcent de la taille initiale. Les indices de texture permet d'optimiser les performances de notre processus d'indexation et de recherche. L'ensemble S_2 résultant de cette étape représente le résultat final du processus de recherche.
- La sélection des vues 2D caractéristiques de modèle 3D demeure un problème de recherche ouvert. Ainsi, un certain nombre d'approches ont tenté d'échantillonner l'espace d'acquisition du modèle 3D. Quelques fonctions d'échantillonnage ont été jugées très denses, produisant ainsi un nombre intéressant de vues 2D pour chaque modèle 3D. D'autres utilisent un degré d'échantillonnage moins important, réduisant ainsi le nombre de vues 2D. En absence de critère précis pour le choix de la fonction de l'échantillonnage des modèles 3D, nous avons opté dans le chapitre précédent pour un échantillonnage à 26 vues. Au deuxième niveau du processus de recherche, aucune restriction n'est prise en compte. Pour chaque modèle, 26 vues sont prises en compte pour la construction de l'indice du modèle.

La figure 6.2 représente l'architecture générale de notre approche de recherche hiérarchique des modèles 3D. Nous y distinguons deux niveaux hiérarchiques de recherche.

Ce schéma traite deux types de requêtes. Les requêtes 2D ne nécessite pas d'étapes de normalisation et de projection, la requête initiée par l'utilisateur est directement segmentée en deux régions (*Le fond et la région d'intérêt*), les indices de couleur de contour et de texture sont ensuite construits à partir des coefficients de la transformée en ondelettes basée région de la région d'intérêt. Les indices des projections de la requête 3D sont construits de la même façon. Contrairement aux requêtes 2D, la construction des indices est précédée par : 1) La normalisation de la requête 3D. 2)

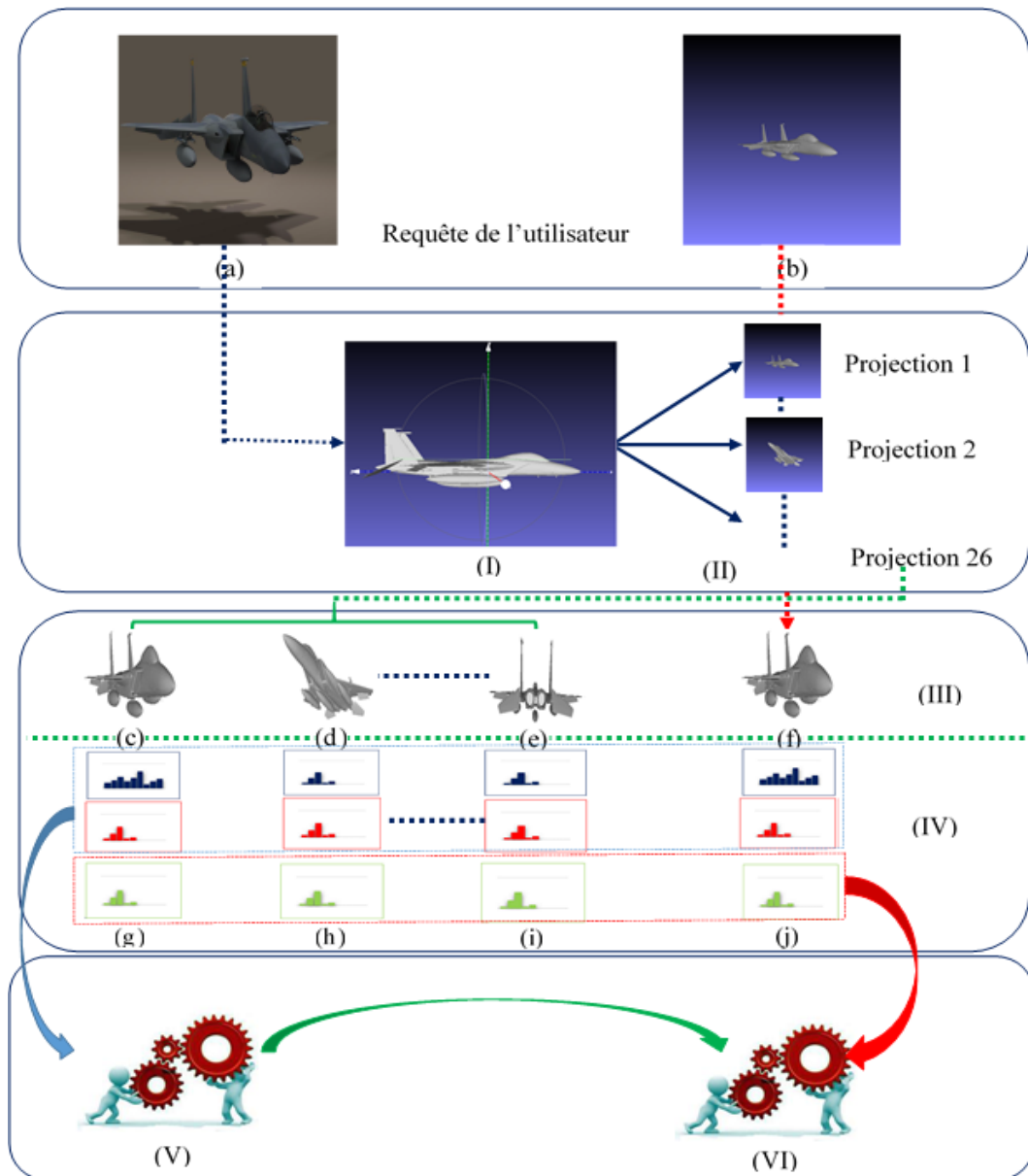


FIGURE 6.2 – Schéma général de la recherche des light field. (a), (b) représentent une requête 3D, 2D respectivement. (c), (d) et (e) représentent les régions d'intérêt des projections 1, 2 et 26 respectivement de la requête a. (f) représente la région d'intérêt de la requête b. (g), (h) et (i) représentent les indices des projections de la requête a. (j) représente les indices de la requête b. Les histogrammes bleus, rouges et verts représentent les indices de couleur, de contour et de texture respectivement. (I), (II) représentent respectivement la phase de la normalisation et de la projection de la requête (a). (III) représente la phase de l'extraction de la région d'intérêt de la requête. (IV) représente la SA-DWT (régions d'intérêt) des requêtes et la construction de leur indices. (V) représente le premier niveau de comparaison. (VI) représente le deuxième niveau de comparaison.

L'extraction des projections 2D de cette requête.

6.3 Visualisation des résultats

La recherche de modèles 3D consiste à comparer et à afficher les objets similaires à la requête. Dans ce type de traitement, la difficulté est de définir le nombre d'images à afficher. Comment jugeons qu'une image de la base est similaire à la requête ou non. La recherche par l'exemple repose sur la mesure de la similarité entre la requête et les modèles 3D de la base par l'intermédiaire de leurs signatures. Chaque modèle étant indexé par un ensemble d'indices, une distance $\{d\}_i^{i=1..n}$, (n représente le nombre de modèles dans la base d'images) est associée à chaque modèle $\{M\}_j^{j=1..n}$. Le résultat de la recherche à visualiser, correspond donc à l'ensemble des modèles à faibles distance d . Le nombre de modèles pertinents dans ces méthodes est défini par l'utilisateur. Ce nombre peut également être dynamique, en affichant tous les modèles, dont la distance à la requête est inférieure à un seuil défini au préalable. Une étude exhaustif des techniques de recherche, montre que la plupart des méthodes sont basées sur le premier mode d'affichage. À cet effet, dans notre travail, nous estimons comme pertinents les 12 premiers modèles résultant de la phase de comparaison. Les distances de ces modèles à la requête sont les plus faibles.

L'affichage des résultats, consiste à retourner à l'utilisateur les modèles à faibles distances. Ces modèles sont affichés suivant un ordre croissant de distances : le premier modèle dans la liste est celui dont la distance à la requête est la plus faible. A l'issue de cette section, il parait très clair que la mesure de distance joue un rôle crucial dans le processus de comparaison. La section suivante traite ce sujet, un état de l'art non exhaustif y est présenté.

6.4 Processus de comparaison

Le processus de comparaison fait l'interface entre l'afficheur des résultats et la base des descripteurs. Ce processus fonctionne sur un principe répétitif ; l'opération de calcul de la similarité est répétée pour tous les indices de la base. L'état de l'art des méthodes de comparaison des modèles 3D, montre que la définition d'une fonction de calcul de similarité est le sujet principal de la plupart des travaux. Cette fonction notée (SIM) permet de quantifier la similarité et d'associer un score à chaque modèle de la base. La forme de l'indice joue également un rôle primordial dans la définition de la fonction (SIM). Étant donné que la plupart des travaux emploient des indices sous forme d'histogrammes ou de distribution de probabilité ; la fonction

de comparaison prendra généralement la forme d'un outil statistique. Dans cette section, nous allons présenter les différentes fonctions de calcul de similarité. Nous détaillons également notre processus de comparaison.

6.4.1 Conventions

La fonction SIM est une relation linéaire, qui peut associer à chaque élément de l'ensemble de départ O (l'ensemble des indices de l'objet), une valeur positive unique d (représentant une distance) [114] :

$$SIM : O \times O \rightarrow R^+ \quad (6.1)$$

Cette fonction doit vérifier les critères suivants :

- Identité : $\forall p \in O, SIM(p, p) = 0$.
- Positivité : $\forall (p_1, p_2) \in (O \times O), SIM(p_1, p_2) > 0$. Avec $p_1 \neq p_2$.
- Symétrie : $\forall (p_1, p_2) \in (O \times O), SIM(p_1, p_2) = SIM(p_2, p_1)$.
- Inégalité triangulaire : $\forall (p_1, p_2, p_3) \in (O \times O \times O), SIM(p_1, p_3) \leq SIM(p_1, p_2) + SIM(p_2, p_3)$.

6.4.2 Normes de Minkowski

Les normes de Minkowski, aussi connues sous le nom de normes L_N sont données par :

$$Sim(his_r - his_c) = \left(\sum_i |his_{ri} - his_{ci}|^N \right)^{\frac{1}{N}} \quad (6.2)$$

Ou

- his_r représente l'histogramme de la requête.
- his_c représente l'histogramme du modèle candidat.

N représente le degré de la norme de Minkowski, les normes 1, 2 et ∞ sont les plus utilisées. Les normes de Minkowski sont très utilisées dans les systèmes de comparaison. Ceci est motivé par la simplicité à les mettre en œuvre, et leur temps d'exécution.

6.4.3 Distance de Bhattacharyya

La distance entre deux histogrammes his_r et his_c dans cette méthode, est définie comme la somme du produit des racines aux carrés des éléments des deux histogrammes. Le résultat est ensuite normalisé en le soustrayant de 1 [12] :

$$SIM(his_r, his_c) = 1 - \sum_i \sqrt{his_r} \sqrt{his_c} \quad (6.3)$$

6.4.4 Distance de Earth Mover

Souvent les indices (histogrammes) à comparer sont de dimensions différentes. Cela nous ramène à définir l'ensemble d'éléments à comparer entre les deux histogrammes au moindre cout. C'est dans ce sens qui est apparue l'approche de calcul de distance d'Earth Mover (EMD). Soient deux histogrammes his_r et his_c à comparer avec :

- $his_r = h_{r1}, h_{r2}, h_{r3} \dots h_{rn}$: l'histogramme du modèle requête de n éléments h_i .
- $his_c = h_{c1}, h_{c2}, h_{c3} \dots h_{cm}$: l'histogramme du modèle candidat de m éléments h_j .

Cette approche vise à minimiser le cout de la transformation de l'histogramme his_r en histogramme his_c , sous les contraintes suivantes :

$$E_{ij} \geq 0 \quad (6.4)$$

Avec E_{ij} représente l'ensemble des éléments h_i de l'histogramme his_r à transporter vers les éléments h_j de l'histogramme his_c .

$1 \leq i \leq n$ et $1 \leq j \leq m$.

$$\sum_{i=1}^n E_{ij} \leq \alpha_i \quad (6.5)$$

$\{\alpha_i\}$ représente l'ensemble des poids des éléments h_i .

$$\sum_{i=1}^m E_{ij} \leq \beta_j \quad (6.6)$$

$\{\beta_j\}$ représente l'ensemble des poids des éléments h_j .

$$\sum_{i=1}^n \sum_{j=1}^m E_{ij} = \min\left(\sum_{i=1}^n \alpha_i, \sum_{j=1}^m \beta_j\right) \quad (6.7)$$

La formule utilisée pour le calcul de la similarité est donnée par :

$$Sim(his_r, his_c) = \frac{\sum_{i=1}^n \sum_{j=1}^m d_{ij} E_{ij}}{\sum_{i=1}^n \sum_{j=1}^m E_{ij}} \quad (6.8)$$

d_{ij} représente la distance entre les éléments d'ordre i et j , appartenant aux deux histogrammes his_r et his_c respectivement. Cette méthode présente l'avantage de restreindre l'espace de comparaison. Cependant sa mise en œuvre reste très complexe. A cet effet elle est plus adaptée aux histogrammes de petites dimensions [112].

6.4.5 Divergence du χ^2

Les approches statistiques de l'analyse de l'image, considéré celle-ci comme une variable aléatoire. Chaque valeur d'une caractéristique des pixels a une probabilité de réalisation donnée. L'ensemble de ces probabilités constitue un vecteur de caractéristiques. La mesure de la similarité consiste donc à déterminer si les réalisations correspondant à deux images, sont issues de la même distribution de probabilité [46].

Soient his_r, his_c les histogrammes caractérisant le modèle requête, et le modèle candidat respectivement. Dans le cas où nous considérons l'ensemble des valeurs de ces deux histogrammes, comme une distribution de probabilité. La distance selon Hetzel et al [56], peut être calculée par la formule suivante :

$$Sim(his_r, his_c) = \sum_i \frac{(his_{ri} - his_{ci})^2}{(his_{ri} + his_{ci})} \quad (6.9)$$

6.4.6 L'intersection d'histogrammes

Cette approche est basée sur le travail de Swain et Ballard [126] pour l'indexation couleur de l'image. Etant donné que les histogrammes représentent la forme la plus utilisée pour la représentation des indices de l'image ; l'idée de Swain et Ballard était de développer des méthodes de comparaison propres aux histogrammes. Cette méthode consiste à calculer le nombre de pixels ayant les mêmes caractéristiques. L'intersection de deux histogrammes se définit donc comme la somme des intersections, chaque intersection correspond au nombre de pixels communs entre les deux histogrammes. Mathématiquement ceci peut être traduit par la formule suivante :

$$Sim(his_r, his_c) = \sum_{i=1}^N \min(h_{ri}, h_{ci}) \quad (6.10)$$

N représente le nombre d'éléments des deux histogrammes his_r et his_c . Afin de normaliser cette formule, les auteurs ont proposé de diviser le rapport exprimé dans la formule 7.6 par la somme des valeurs de l'histogramme du modèle requête :

$$Sim(his_r, his_c) = \frac{\sum_{i=1}^N \min(h_{ri}, h_{ci})}{\sum_{i=1}^N h_{ri}} \quad (6.11)$$

Selon la formule 6.11, deux objets similaires auront un résultat proche de 1. Si le résultat de ce rapport est proche du zéro, les deux modèles sont fortement différents. Ce rapport peut également être utilisé pour mesurer la dissimilarité, selon la formule suivante :

$$dism(his_r, his_c) = 1 - inter(his_r, his_c) \quad (6.12)$$

6.4.7 Distance de Hausdorff

Cette méthode consiste à : 1) comparer chaque élément de l'histogramme de la requête à l'ensemble des éléments de l'histogramme du modèle candidat. On garde la plus faible distance pour chaque élément de la requête. 2) calculer la valeur maximale des distances obtenues. Cette distance mesure la similarité entre les deux histogrammes. Ceci peut être exprimé par la formule suivante :

$$Sim(his_r, his_c) = \max_{h_r \in his_r} \{ \min_{h_c \in his_c} \{ d(his_r, his_c) \} \} \quad (6.13)$$

$d(h_r, h_c)$ représente la distance entre les éléments h_r et h_c .

6.5 Comparaison des light field

Un light field peut être indexé par un ensemble d'histogrammes correspondant à ses différentes vues 2D. Ceci a conduit Ding-Yun C et al [32] à proposer une approche de mesure de similarité plus en accord avec la structure particulière des light field. Dans le cas où l'on cherche à comparer deux light field (L_r, L_c), l'auteur propose d'additionner les similarités issues de toutes les vues 2D, pour calculer la similarité entre L_r et L_c . Le premier aspect traité dans ce travail est la définition du système de caméras. Ce système doit produire la plus faible distance entre L_r et L_c , en additionnant les similarités entre les différentes vues.

Dans cette approche, chaque modèle est placé au centre d'un système de vingt caméras (figure c). Afin de simplifier les calculs et de gagner en temps de recherche, Ding-Yun C et al ont choisi que ce système soit sous la forme d'un dodécaèdre. Sachant que cette forme est constituée d'un ensemble de polygones de trois sommets pour chacun, lors de la recherche, chaque caméras se déplace sur les trois sommets, produisant ainsi 60 angles de vue. Pour un point de vue donné ν_i $i = 1 \dots 60$, la distance entre les deux modèles est égale à la somme des distances entre les 10 vues 2D (résultant de la rotation de l'objet autour de son point de gravité) des deux modèles, cette distance peut être exprimée par la formule suivante :

$$d_{\nu_i} = \sum_{j=1}^{10} d(L_{1j}, L_{2j}) \quad (6.14)$$

$i=1 \dots 60$. L_{1j} représente la vue 2D du premier modèle dans la pose j . L_{2j} représente la vue 2D du deuxième modèle dans la pose j .

Cette approche continue à tourner le système de caméras dans un processus répétitif, à chaque itération une distance d_{ν_i} est calculée. A la fin de ce processus,

on définit la distance entre les deux modèles, comme étant la plus faible distance de l'ensemble des distances $\{d_{\nu_i}\}$:

$$d_{L_1, L_2} = \min\{d_{\nu_i}\} \quad (6.15)$$

$i = 1 \dots 60$. Comme nous pouvons le constater dans la figure 6.3, le système de ca-

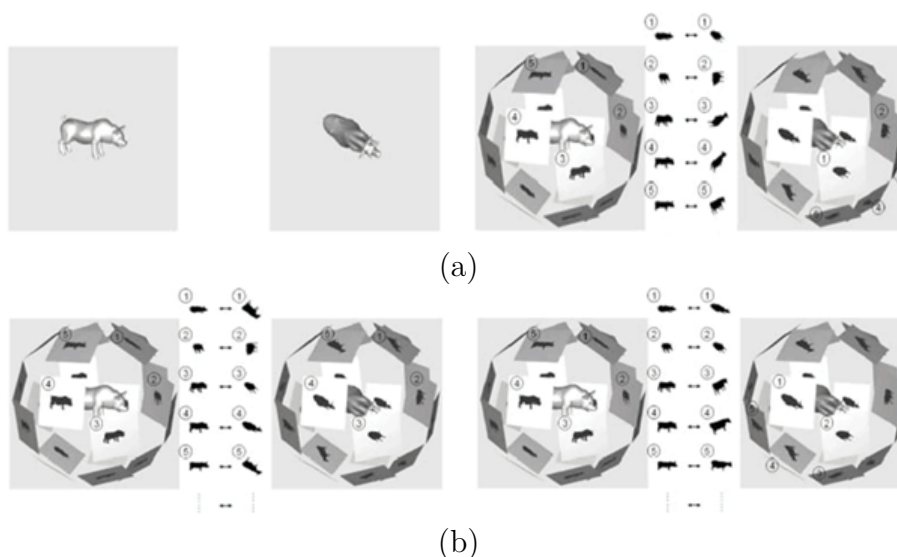


FIGURE 6.3 – Illustration du système de comparaison des light field (proposé dans l'approche light field descriptor LFD). (a) représente les deux modèles à comparer, la projection des deux objets dans un système de caméras à la forme d'un dodécèdre. (b) La comparaison des vues 2D dans deux systèmes de caméras différents.

méras illustré dans la figure b, offre la meilleure similarité entre les deux objets. A cet effet, le calcul de la similarité entre les deux modèles dans cet exemple consiste à additionner les distances associées aux 10 light field. Ce processus de recherche offre à cette approche une robustesse vis à vis les variations de l'orientation du modèle. Cependant le temps de calcul est loin d'être optimal. Dans cette optique, nous proposons une normalisation des modèles 3D, avant toute opération de comparaison ou de recherche. Afin d'éviter de faire tourner le modèle dans un processus répétitif, nous proposons de réorienter tous les modèles dans le même sens dans un repère unique. Contrairement à l'approche proposée par Ding-Yun C et al, nous plaçons les modèles à comparer au centre d'un système de caméras statique, aucune rotation de ce système n'est réalisée. Nous calculons ensuite la distance entre les deux modèles (voir figure 6.4). Cette figure représente la recherche des objets similaires à une requête 3D, dans le cas des requêtes 2D, celles-ci sont comparées successivement aux vingt-six projections du modèle candidat de la base.

La figure 6.4 montre que l'alignement des poses des deux objets 3D, a per-

mis de réduire le nombre de comparaisons. Chaque projection du premier objet est comparée directement à la projection correspondante dans le deuxième objet. Ce processus comprendra donc vingt-six comparaisons. La distance entre les deux objets est calculée par l'équation 6.15, avec $i = 1 \dots 26$.

Dans la suite de ce chapitre, nous nous intéressons à la comparaison des vues 2D des light field. Nous utilisons l'une des métriques proposées précédemment. Nous essayons également d'adapter cette métrique à nos histogrammes, qui peuvent parfois être la combinaison de plusieurs histogrammes.

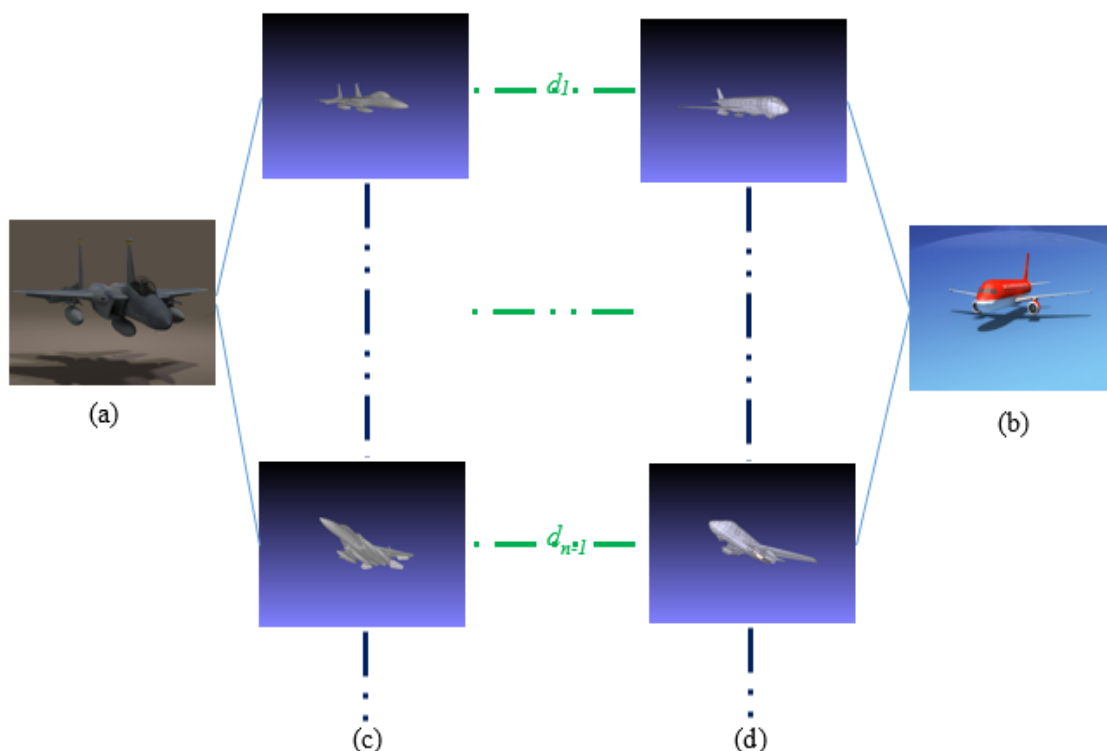


FIGURE 6.4 – Comparaison de deux objets 3D. (a), (b) représentent deux objets 3D. (c), (d) représentent des échantillons des ensembles de projections 2D des deux objets (a) et (b) respectivement. d_1 , d_{n-1} représentent les distances entre la première et la $(n-1)$ ^{ème} projection des deux objets (a) et (b)

6.5.1 Comparaison des vues des light field

Dans le cas de modèles 3D aux descripteurs multiples, la caractérisation est traduite par la construction d'histogrammes issus de différents espaces de caractéristiques (la forme, la texture, la couleur... etc). Chaque histogramme exprime la variation des caractéristiques par une méthode différente. L'échelle des histogrammes, les centres des repères employés ne se ressemblent que rarement. Ainsi la fusion de ce genre de descripteurs se révèle toujours comme une opération complexe et assez dé-

licate. Deux solutions ont été proposées dans la littérature : 1) la concaténation des histogrammes en un seul histogramme, avant de procéder aux comparaisons. Dans cette méthode les histogrammes concaténés doivent être normalisés, cette opération consiste à mettre à l'échelle les valeurs des différents histogrammes. 2) la fusion des distances après une comparaison séparée des différents histogrammes. Cette opération doit être précédée par une étape de normalisation des distances des différents histogrammes.

Notre méthode de comparaison des vues 2D, se repose sur les indices employés et le niveau hiérarchique de comparaison. Au premier niveau de comparaison, nous adoptons un modèle hiérarchique séparant les deux attributs la *couleur* et le *contour*. Contrairement à cette approche, la comparaison des projections 2D au deuxième niveau de comparaison, repose sur une description mono-vectorielle de la projection 2D, cette notion est concrétisée par la concaténation des trois histogrammes de texture H_T^i , (i représente le niveau de décomposition de la transformée en ondelettes basée région de la projection 2D, tel que $1 \leq i \leq 3$) en un seul histogramme de texture H_T .

Comme nous l'avons vu au chapitre V, notre méthode d'indexation des light field, décrit chaque projection par trois histogrammes de : couleur (H_C), de contour (H_E) et de texture (H_T). Au premier niveau de comparaison, nous nous limitons à la comparaison des histogrammes de couleur et de contour. Cette comparaison est basée sur la distance de Hetzel (équation 6.9). Les distances des histogrammes de couleur et de contour s'écriront donc sous la forme suivante :

$$d(H_{C(r)}, H_{C(c)}) = \sum_{j=1}^{400} \frac{(H_{C(r)}^j - H_{C(c)}^j)^2}{(H_{C(r)}^j + H_{C(c)}^j)} \quad (6.16)$$

$d(H_{C(r)}, H_{C(c)})$ représente la distance entre deux histogrammes de couleur de deux projections données.

$H_{C(r)}, H_{C(c)}$ représentent les histogrammes de l'attribut couleur de deux projections des objet requête et candidat respectivement.

$$d(H_{E(r)}, H_{E(c)}) = \sum_{j=1}^{100} \frac{(H_{E(r)}^j - H_{E(c)}^j)^2}{(H_{E(r)}^j + H_{E(c)}^j)} \quad (6.17)$$

$d(H_{E(r)}, H_{E(c)})$ représente la distance entre deux histogrammes du contour de deux projections données.

$H_{E(r)}, H_{E(c)}$ représentent les histogrammes de l'attribut contour de deux projections des objet requête et candidat respectivement.

Chaque objet 3D de la base est donc distingué par vingt-six distances de couleur et vingt-six autres distances de contour. Afin que les valeurs de ces distances soient homogènes, nous proposons de normaliser les vingt-six distances à l'aide de la formule suivante :

$$d_{Nor} = \frac{d - \min\{d_i\}_{i=1}^{26}}{\max\{d_i\}_{i=1}^{26} - \min\{d_i\}_{i=1}^{26}} \quad (6.18)$$

La distance entre deux projections au premier niveau, est le résultat de la fusion des distances de la couleur et du contour, selon la formule suivante :

$$d(H_{(r)}, H_{(c)}) = \alpha \times d_{Nor}(H_{C_{(r)}}, H_{C_{(c)}}) + (1 - \alpha) \times d_{Nor}(H_{E_{(r)}}, H_{E_{(c)}}). \quad (6.19)$$

La valeur du coefficient de pondération est définie dans la partie expérimentale. La distance entre deux objets est la distance minimale des vingt-six distances, qui existent entre les différentes projections. À l'issue de ce premier niveau, seulement 38% des objets de la base, qui sont retenus pour construire l'espace de recherche du deuxième niveau de comparaison. Les objets sont classés selon un ordre croissant de leur distances.

Le raffinement de la pertinence à la requête des objets résultats du premier niveau, est basé sur un deuxième niveau de comparaison. En effet, nous allons comparer les histogrammes de texture ($H_{T_{(c)}}$) des objets de la base à l'histogramme de texture de la requête ($H_{T_{(r)}}$), suivant un processus itératif de comparaisons. La comparaison dans ce niveau, reste identique à celle utilisée dans le premier niveau de comparaison. La particularité réside dans la concaténation qui précède le calcul de la distance. Cette combinaison peut être exprimée par la formule suivante :

$$H_T = [H_T^1 H_T^2 H_T^3] \quad (6.20)$$

H_T^1 , H_T^2 , H_T^3 représentent les histogrammes de texture aux premier, deuxième et troisième niveau de décomposition de la transformée en ondelettes basée région.

Les deux objets requête et candidat sont ensuite comparés, à l'aide des distances ($d(H_{T_{(r)}}, H_{T_{(c)}})$) entre les histogrammes de texture de leurs vingt-six projections respectives. Ces distances sont calculées à l'aide de la formule 6.16 en utilisant l'histogramme de texture à la place de l'histogramme de couleur. Les distances obtenues sont ensuite normalisées à l'aide de la formule 6.18. Contrairement au premier niveau de comparaison, aucune fusion des distances n'est effectuée dans ce deuxième niveau. Ceci dit la distance entre les attributs de texture de deux objets, est égale à la distance minimale des distances de texture des vingt-six projections. Le résultat de la recherche à ce deuxième niveau est les douze premiers objets 3D,

qui sont les plus proches de la requête en terme de distance.

6.6 Conclusion

La recherche de l'objet est une étape décisive dans notre méthode. À cet effet nous avons prêté beaucoup d'attention aux choix de la distance de comparaison et à l'architecture du système de recherche. Vu la complexité des indices employés par le système d'indexation des objets 3D, nous avons proposé l'organisation des comparaisons selon le degré de complexité des indices employés. Le premier degré de notre système de recherche est basé sur des comparaisons entre les indices de couleur et de contour. La taille de ces indices permet de balayer la base de modèles 3D en un temps très court. À l'issue de cette étape la taille de la base de données se réduit considérablement. Cette deuxième base constitue l'espace de recherche au deuxième niveau ou les comparaisons sont basées sur les indices de texture. Les résultats en terme de temps montrent que la hiérarchisation de notre système de recherche d'objets 3D est significativement améliorée.

Chapitre 7

Simulations et résultats

Résumé : Dans ce dernier chapitre, nous présentons les simulations des approches proposées. Nous y présentons également les résultats obtenus sur les diverses bases d'images. Ces résultats justifient nos choix théoriques et argumentent nos propos. Finalement, nous présentons une étude comparative de nos résultats avec quelques méthodes de référence dans le domaine de l'indexation et de la recherche d'objets 3D.

7.1 Introduction

Les travaux de cette thèse ont respecté certaine méthodologie. Celle-ci consiste à mener les expériences et les tests en parallèle avec l'étude théorique des approches. Chaque contribution proposée a été accompagnée d'une simulation, on a ainsi effectué les simulations suivantes : une première simulation concernant l'indexation des vues 2D par la transformée en ondelettes basée région. Celle-ci consiste en l'exploitation des descripteurs décrits dans le chapitre trois dans un système de recherche de modèle 2D. Nous avons également effectué diverses simulations pour fixer les valeurs des coefficients de pondération présentés précédemment. Ces coefficients vont ensuite être maintenus pour l'ensemble des simulations.

Nous nous sommes ensuite intéressés à la simulation de la méthode de la normalisation des modèles 3D. Les chapitres cinq et six ont été accompagnés de simulations d'indexation et de recherche de modèles 3D. Dans cette partie nous nous sommes intéressés à différents aspects. À l'instar du travail fait dans le cas des images 2D, nous avons commencé par l'évaluation des performances en utilisant les différents indices des vues 2D. Nous évaluons ensuite les performances de notre système de recherche de modèles 3D, pour différentes métriques de calcul de la similarité. Une de nos contributions concerne l'accélération du temps de recherche, à cet effet nous proposons l'évaluation de notre système en terme du temps. Finalement, une étude comparative avec d'autres méthodes de la littérature a été proposée.

7.2 Bases de données

Les simulations de l'indexation et de la recherche d'images 2D, ont été effectuées sur deux bases d'images issues de la base d'images Corel [6]. Dans les tests sur les modèles 3D, on a utilisé deux bases : la base d'objets Princeton Shape Benchmark (PSB) [7] et la base NTU 3D modèles benchmark [8].

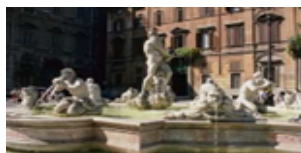
7.2.1 La Base d'images Corel

Les images dans cette base sont réparties sur 23 catégories d'images, communément appelées classes, contenant 1000 images chacune : bus, cheval, africain, immeubles, plage, dinosaure, éléphant, rose, montagne, plats, chien, lézard, mannequins, coucher de soleil, plage, cascade, monuments, navire, ski, dessert, papillon, nuit, forêt. La base de Wang [9] regroupe 1000 images 256*384 pixels de la base Corel, qui sont réparties sur 10 classes :bus, cheval, africain, immeubles, plage, dinosaure, éléphant, rose, montagne, plats. Une deuxième base d'images issue de la base Corel est utilisée dans ce chapitre [6]. Celle-ci est constituée de 5000 images

réparties sur 50 classes : . Dans la suite de ce chapitre, nous désignons la base de Wang par le terme DB1, et la dixième base 5000 par le terme DB2.



Plage



Monument



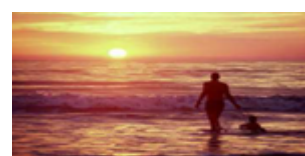
Eléphant



Montagne



Immeuble



Coucher de soleil



Voiture



Plat



Rose

FIGURE 7.1 – *Echantillons d'images de la base de Corel*

7.2.2 La base Princeton Shape Benchmark (PSB)

La base de modèles 3D Princeton Shape Benchmark (PSB) est constituée des images collectées par l'équipe du *Princeton Shape Retrieval and Analysis Group (PSRAG)*. Cette base est considérée comme une référence pour les spécialistes de l'indexation et la recherche d'objets 3D. Elle est composée de deux bases : La *Train database* qui est utilisée pour l'apprentissage. Elle est composée de 907 objets, regroupés en 90 classes. La deuxième partie du PSB est destinée aux tests, elle est connue sous l'appellation du *Test Database*. Cette partie contient 907 objets répartis sur 92 classes. Des échantillons de différentes classes de modèles 3D contenus dans la base PSB sont représentés dans la figure 7.2.

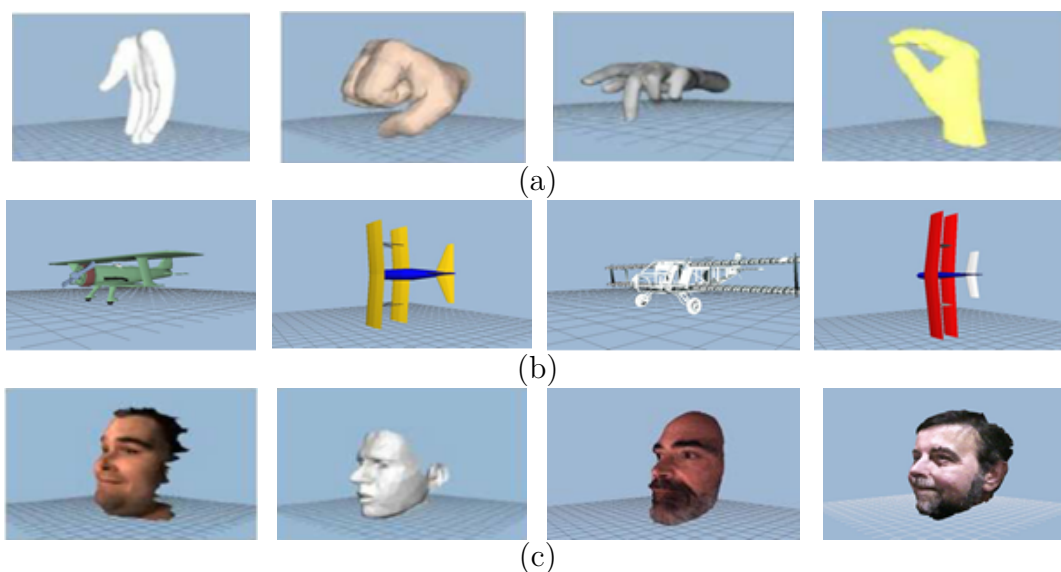


FIGURE 7.2 – *Echantillons de modèles 3D de la base PSB. (a) des échantillons d'objets 3D de la classe main. (b) des objets 3D de la classe avion. (c) des objets 3D de la classe visage.*

7.2.3 La base NTU 3D modèles benchmark

La base NTU est constituée des images collectées de différents sites web internet. Elle est publiée par un groupe de chercheurs de l'université de Taïwan, pour la communauté scientifique du domaine de l'indexation et de la recherche de modèles 3D. Cette base comprends 549 modèles 3D répartis par le collectionneur en une vérité terrain de 47 classes thématiques : véhicules, hommes, avions... etc. 1284 autres modèles sont regroupés dans diverses classes. Pour des raisons de robustesse, le concepteur déconseille l'utilisation de cette deuxième partie pour les tests. Donc, cette base contient 1 833 objets 3D.

7.3 Paramètres d'évaluation

L'évaluation des méthodes d'indexation et de recherche d'images est possible, grâce à l'utilisation des méthodes de classification des images et objets. Le principe de ces méthodes d'évaluation est de compter comme correcte chaque image ou modèle appartenant à la même classe que la requête. Nos expérimentations impliquent l'utilisation de trois paramètres d'évaluation :

- Courbe Rappel/Précision

Cette courbe est basée sur deux critères : Le rappel et la précision. Le paramètre rappel mesure le rapport entre le nombre d'images pertinentes retrouvées et le nombre

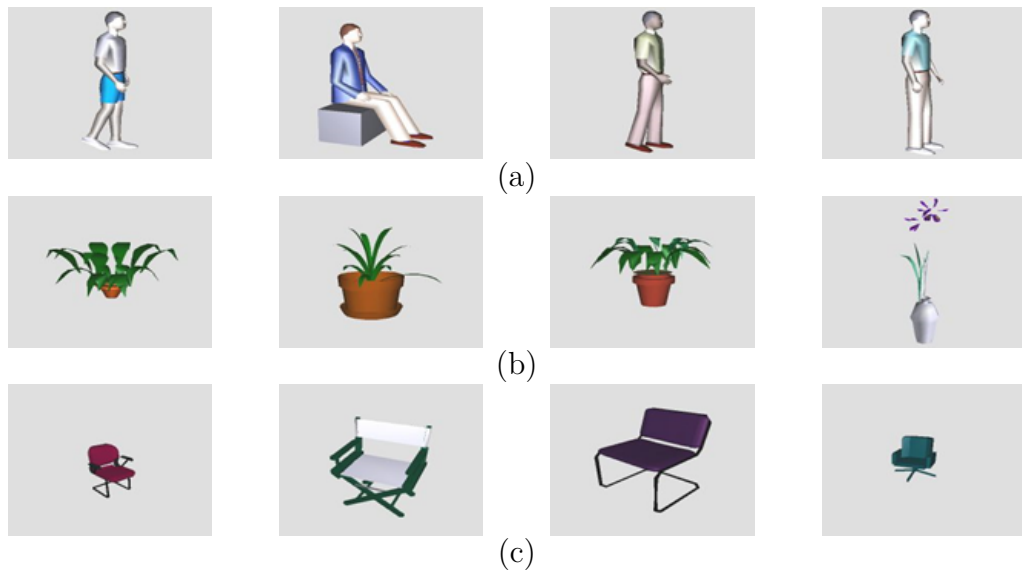


FIGURE 7.3 – Échantillons de modèles 3D de la base NTU. (a) des échantillons d'objets 3D de la classe homme. (b) des objets 3D de la classe plante. (c) des objets 3D de la classe fauteuil.

d'images pertinentes dans la base [114] :

$$Rappel = \frac{\text{nombre d'images pertinentes retrouvées}}{\text{nombre total d'images pertinentes}} \quad (7.1)$$

Le paramètre précision permet de calculer le rapport entre le nombre d'images pertinentes retrouvées, et le nombre total des images :

$$Précision = \frac{\text{nombre d'images pertinentes retrouvées}}{\text{nombre d'images retrouvées}} \quad (7.2)$$

– Taux de pertinence Moyen (ARR)

Comme nous l'avons vu dans le chapitre 6, les images retournées par un moteur de recherche sont affichées dans un ordre décroissant de similarité. Les premières images dans cette liste sont les plus similaires à la requête. Le taux de pertinence RR est défini comme étant le rapport entre le nombre d'images pertinentes retrouvées et le nombre total d'images pertinentes dans toute la base, pour un ensemble de k images retrouvées.

$$RR = \frac{\text{nombre d'images pertinentes retrouvées}}{\text{nombre d'images pertinentes}} \quad (7.3)$$

un deuxième paramètre est également défini pour calculer la valeur moyenne de l'ensemble des taux de pertinence pour k images retournées. Ce paramètre est appelé le taux de pertinence moyen (ARR).

7.4 Expérimentations sur l'indexation et la recherche d'images 2D basée région

Comme nous l'avons vu dans les chapitres 5 et 6, les objets 3D dans notre travail se présentent toujours sous la forme d'un ensemble d'images 2D. D'autres parts, notre système traite les requêtes 2D. À cet effet, nous commençons nos simulations par l'évaluation des méthodes d'indexation et de recherche des images 2D. Nous nous focalisons plus particulièrement, dans cette partie sur l'évaluation de notre méthode d'indexation d'images basée région (proposée dans le chapitre 03). Cette évaluation est un point délicat, car le processus d'indexation et de recherche d'images dépend de différents paramètres (le nombre d'itérations de la transformée SA-DWT, le nombre de *bins* des différents histogrammes.etc). À cet effet, nous proposons de mener un ensemble d'expériences pour fixer les valeurs optimales de ces paramètres. Dans un second temps, nous étudions l'influence de l'introduction de la notion basée région, en comparant notre méthode aux méthodes d'indexation basées contenu. Nous concluons cette partie par une comparaison de notre méthode à quelques méthodes d'indexation d'images 2D basées région.

7.5 Protocole de recherche des vues 2D

Notre protocole d'expériences suit les étapes suivantes :

1. Les images de nos bases d'images sont tout d'abord segmentées en deux régions chacune : la région d'intérêt et la région du fond.
2. Chaque image est convertie en composantes Y , C_b et C_r .
3. L'application de la transformée en ondelettes basée région sur l'ensemble des images de la base d'images.
4. La construction de trois histogrammes caractérisant les attributs couleur, texture et contour (H_C , H_T , H_E) pour chaque image. Cette étape est réalisée à l'aide des algorithmes de la construction de descripteurs décrits dans le chapitre 3.
5. Nos bases d'images sont constituées de différentes classes d'images. Pour chaque manipulation, nous considérons l'image requête I_r , nous évaluons la similarité entre l'image requête et les images de la base d'images. Cette évaluation est définie par la distance χ^2 comme suit :

$$d(H_r, H_c) = \sum_{i=1}^L \frac{(H_q(i) - H_c(i))^2}{H_q(i) + H_c(i)} \quad (7.4)$$

$H_r(i)$ et $H_c(i)$ représentent les $i^{\text{ème}}$ valeurs de l'histogramme des images requêtes et candidates respectivement.

L représente le nombre de *bins* ou la longueur de l'histogramme.

Afin de combiner la distance de calcul de la distance finale, nous devons normaliser les distances en appliquant la formule expliquée dans la section () :

$$d_{nor}(H_r, H_c) = \frac{d(H_r, H_c) - \min(d(H_r, H_c))}{\max(d(H_r, H_c)) - \min(d(H_r, H_c))} \quad (7.5)$$

\min et \max calculent les valeurs maximale et minimale des distances entre les histogrammes des différentes images. À l'issue de cette étape, on aura trois distances d_{HC} , d_{HT} et d_{HE} pour chaque image candidate.

6. Calcul de la distance d_{H_r, H_c} pour chaque image de la base à l'aide de la formule suivante :

$$d(H_r, H_c) = \alpha \times d_{HC} + (1 - \alpha) \times [\beta \times d_{HT} + (1 - \beta) \times d_{HE}] \quad (7.6)$$

Nous évaluons ensuite notre algorithme pour différentes valeurs des paramètres de pondération α et β : Les k premières images constituent le résultat de notre système d'indexation et de recherche de l'image.

7.5.1 Niveau de décomposition de la transformée SA-DWT

Une importante question à laquelle nous avons du répondre avant de commencer nos expérimentations, est le nombre d'itérations nécessaire de la SA-DWT. À cet effet, nous avons commencé par un ensemble d'expériences sur la base d'images DB_1 , afin de définir la valeur optimale de N (le nombre d'itérations de la SA-DWT), pour laquelle notre approche d'indexation enregistre les meilleures performances. La courbe du paramètre ARR en fonction du nombre d'images retrouvées, montre que l'augmentation du nombre d'itération de la SA-DWT avait un impact positif sur les performances de notre approche. On peut ainsi constater que pour un nombre d'itérations $N \geq 3$, les performances exprimées par le paramètre ARR ont atteint la valeur optimale ($ARR = 0,98$ pour 60 images retrouvées voir figure 7.4). Dans la suite de ce chapitre, nous utilisons une SA-DWT à trois itérations.

7.5.2 Taille des histogrammes

- Taille de l'histogramme de texture

Nous avons présenté dans le chapitre III, la méthode de construction de l'histogramme de texture. Celle-ci s'appuie sur la combinaison des histogrammes HL , LH et HH . Chaque histogramme correspond à une sous bandes des coefficients de détails de la transformée en ondelettes. La taille de ces histogrammes (qui est exprimée en nombre de *bins*) est très intéressante. Au-delà d'une certaine longueur les valeurs de

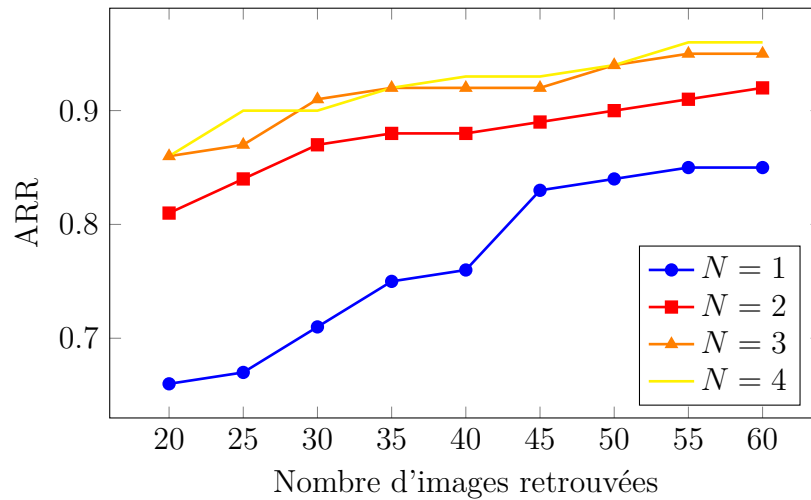


FIGURE 7.4 – Évaluations des performances du système d'indexation d'images pour différents nombre d'itération de la SA-DWT.

$bins$ deviennent assez faibles et insignifiante.

Nous évaluons ici les performances de la méthode de l'indexation de l'image tout en variant le nombre de $bins$ de l'histogramme de texture.

Les courbes de la figure 7.5, montrent que 3000 $bins$ de l'histogramme de texture donnent de très bons résultats par rapport aux autres longueurs. On va garder cette mesure empirique pour le reste de cette section.

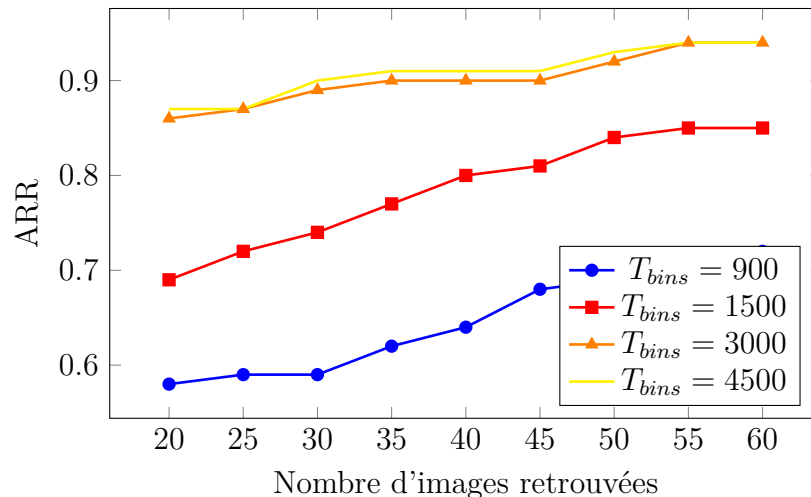


FIGURE 7.5 – Évaluations des performances du système d'indexation d'images par rapport au nombre de bins de l'histogramme de texture.

– Taille de l'histogramme de couleur

En ce qui concerne le nombre de *bins* de l'histogramme de couleur, nous pouvons voir sur la figure 7.6 que 400 *bins* sont suffisants pour atteindre un résultat optimal. Au delà de cette valeur les performances se convergent nettement.

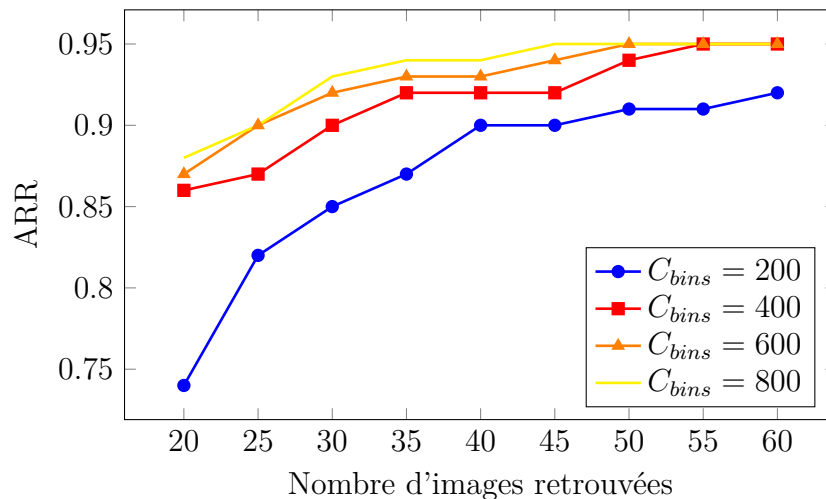


FIGURE 7.6 – Évaluations des performances du système d'indexation d'images par rapport au nombre de bins de l'histogramme de couleur.

– Taille de l'histogramme de contour

Les résultats de la Figure 7.7 montrent clairement que l'histogramme de contour est le moins volumineux par rapport aux autres. En effet, si nous comparons les performances pour $E_{bins} = 100$ et ceux obtenus avec 200 *bins*, nous nous rendons compte que la différence en terme de performance est presque nulle.

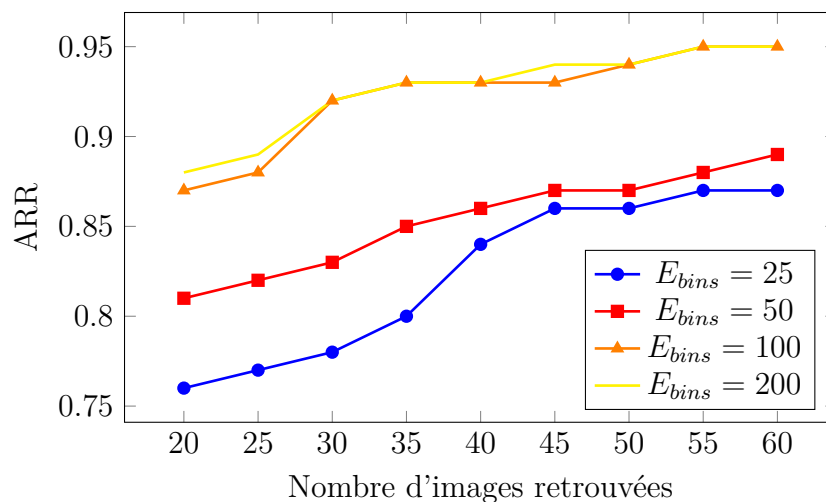


FIGURE 7.7 – Évaluations des performances du système d'indexation d'images par rapport au nombre de bins de l'histogramme de contour.

7.5.3 Calcul des paramètres de pondération α et β

Afin de définir les paramètres de pondération α et β de l'équation 7.6, nous effectuons deux tests sur la base DB1. Les paramètres retrouvés vont être maintenues pour tous les tests de notre travail. Comme il est illustré sur la figure 7.8, nos tests consistent à varier ces paramètres afin d'évaluer les performances de notre méthode pour chaque valeur de pondération. Nous pouvons ainsi remarquer que les meilleurs résultats sont obtenus pour de faibles valeurs de β ($\beta_{optimal} = 4,5$) et de grandes valeurs de α ($\alpha_{optimal} = 5,5$). Autrement dit pour une forte pondération des indices de l'attribut contour et de l'attribut couleur, par rapport à l'attribut texture.

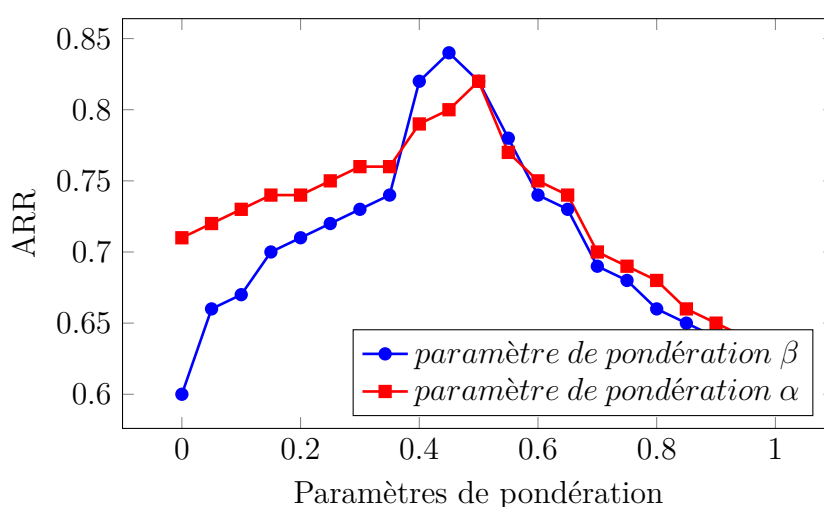


FIGURE 7.8 – Définition des valeurs des paramètres de pondération.

7.5.4 Comparaison de la méthode proposée aux méthodes d'indexation basées région

– Première expérience

Nous proposons ici une comparaison de notre méthode à la méthode proposée dans [139]. Nous évaluons en particulier le gain de performances apporté par l'utilisation de la description de contour. Les résultats de la figure 7.9 montrent que :

- L'utilisation du fond de l'image dans le cas de la classe dinosaure avait un effet positif sur les performances de l'algorithme, parce que toutes les images de cette classe partagent le même fond de l'image.
- Les résultats qui sont les plus dégradés sont ceux obtenus pour les classes émeuble, plage et éléphant. Nous pouvons ainsi remarquer que l'utilisation de notre méthode offre une nette amélioration dans les résultats enregistrés dans ces classes.

L'amélioration des performances en utilisant notre méthode est due à l'emploi d'une caractérisation plus étendue des contours. Ceux-ci sont décrits dans les directions 0° , 45° , 90° et 135° , contrairement à la méthode du papier [139], qui ne capture le contour que dans les directions 0° , 45° et 90° .

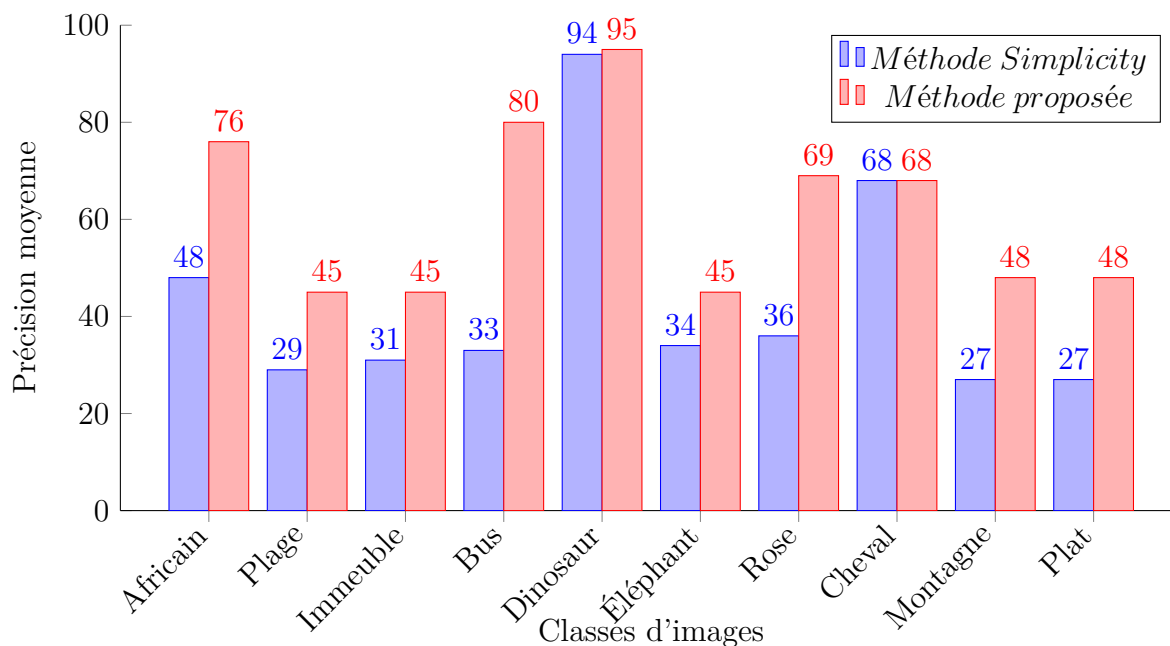


FIGURE 7.9 – Comparaison de notre méthode d'indexation basée région à la méthode *Simplicity* à l'aide du paramètre *Précision moyenne*.

– Deuxième expérience

Nastev et al ont proposé un système d'indexation basé sur la description locale de l'information de contour. Les descripteurs sont basés sur l'utilisation des coefficients de la transformée en ondelettes des fenêtres décomposant l'image en un ensemble de blocs. L'inconvénient de cette approche, provient du fait que la fenêtre utilisée pour la décomposition de l'image, peut contenir des parties de différentes régions de l'image. Elle peut contenir des parties du fond de l'image et d'autres parties de la région d'intérêt [94]. Contrairement à cette méthode, notre approche offre une décomposition exacte de l'image et une parfaite limitation de la description locale des régions de l'image. À cet effet, nous pouvons observer que notre méthode est plus performante que la méthode de Nastev et al. La comparaison illustrée par la figure 7.10 est basée sur des tests sur la base d'images DB1.

– Troisième expérience

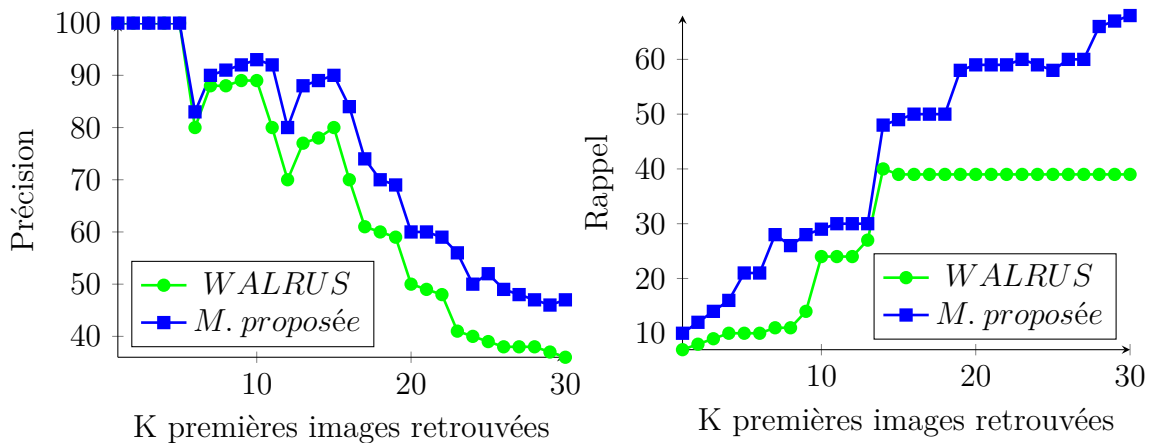


FIGURE 7.10 – Comparaison de notre méthode d’indexation basée région à la méthode de WALRUS en terme des paramètres de précision et de rappel.

Afin de palier aux inconvénients de l’utilisation de la transformée en ondelettes pour l’indexation d’images basée région. Monirul Islam et al [145] ont proposé d’employer la transformée en curvelet de l’image, qui permet de capturer les différentes orientations de la variation de texture. Cette expérience est effectuée sur une base de 5000 images (DB2). La figure montre les améliorations apportées par l’utilisation de notre méthode, par rapport à l’indexation basée sur la transformée de curvelettes. Étant donné que la transformée de curvelettes, ne peut être appliquée que sur des régions rectangulaires, son application dans le cadre de l’indexation d’images, est réalisée après une conversion des régions de formes irrégulières en régions de formes régulières. À cet effet, l’information contenue dans la région à transformer ne correspond pas exactement à l’information contenue dans la région à caractériser.

– Quatrième expérience

Les premiers résultats expérimentaux sur la base d’images DB1 montrent que notre méthode a pu dépasser les performances enregistrées par les autres méthodes d’indexation, en terme de précision. Le paramètre précision a été calculé avec $k = 20$ (k désigne le nombre d’images retrouvées). Cet histogramme montre que notre méthode a la capacité de retrouver presque parfaitement l’ensemble des classes, à l’exception des classes *plage*, *montagne*, *éléphant*, où nous n’avons pas dépasser les 80% de précision. Ceci est dû à la composition des images de ces classes, qui est très hétérogène et de textures très différentes, rendant leur distinction par rapport aux autres classes très difficile. Par contre, notre méthode procure une meilleure caractérisation locale des régions dinosaures, bus et rose et enregistre de meilleurs résultats par rapport aux approches *RCBIR* et *Fuzzy Club*. On peut également remarquer que les approches *IRM* et *histogramme géométrique* semblent a priori être mal adaptées à la

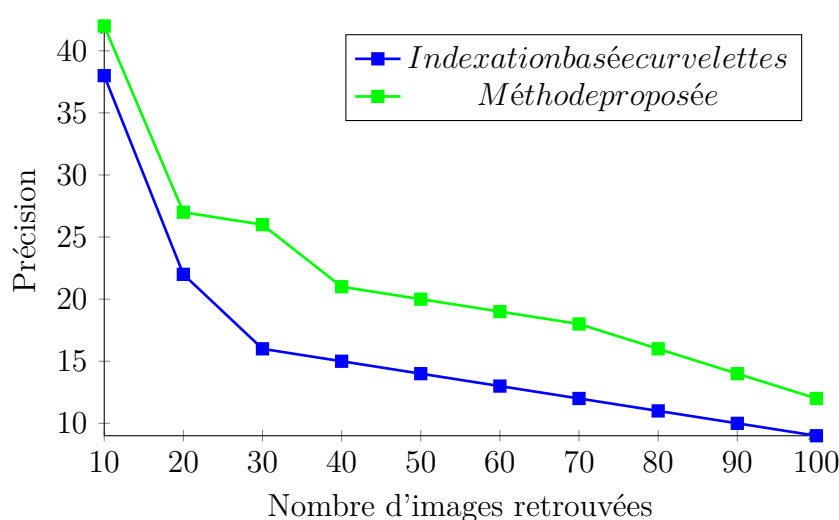


FIGURE 7.11 – Comparaison de la méthode proposée à la méthode basée sur les curvelettes en terme de précision.

caractérisation des attributs texture et couleur dans l'ensemble des classes (figure 7.12).

7.6 Expérimentations sur l'indexation et la recherche des light field

Dans cette partie, nous désignons la base d'objets NTU par le terme DB3 et DB4 une partie de la base PSB. DB4 est la sélection de 1829 objets 3D, qui sont répartis sur les différentes classes de la base PSB. Le but des expériences dans cette section est de :

1. Mettre en évidence quel est le nombre de projections 2D des light field idéal à leur caractérisation, qui permettra de réaliser le compromis performances d'indexation/temps d'exécution.
2. Étudier l'apport des méthodes de la normalisation et de l'alignement des objets 3D, aux performances de notre méthode d'indexation. Ceci en terme de précision et du temps de calcul.
3. Mettre en évidence les avantages apportés par l'introduction de la notion de la recherche hiérarchique.
4. Comparer la méthode proposée aux méthodes d'indexation d'objets 3D.

Nous avons proposé dans le chapitre 04 les opérations de normalisation et d'alignement d'objets 3D, précédant toutes les opérations d'indexation et de recherche. Le

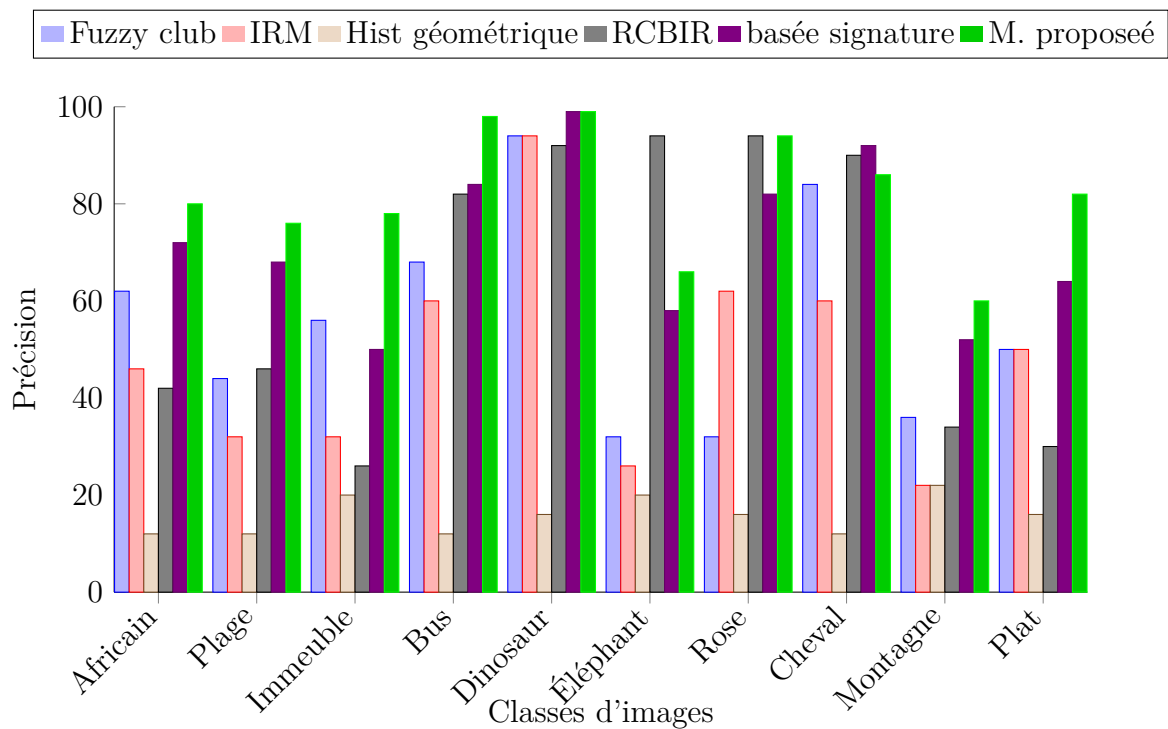


FIGURE 7.12 – Comparaison de notre méthode d'indexation basée région aux méthodes d'indexation basées régions en terme du paramètre précision.

but de ce travail est de rassembler les objets de la même classe autour des mêmes conditions de pose et de projection dans un repère 3D. Grâce à la normalisation et à l'alignement des modèles 3D, il est possible de se contenter d'une seule comparaison. Étant donné que les objets partagent la même pose, on peut s'en passer de la rotation itérative des objets lors de la phase de comparaison. Dans ce cadre, nous présentons dans ce chapitre les résultats expérimentaux, visant à comprendre l'influence de ces contributions sur les performances de l'indexation.

L'une de nos contributions dans le chapitre 05 est la réduction du nombre de projections 2D des light field à vingt-six projections. Nous allons montrer que grâce à des choix judicieux nous avons pu atteindre les mêmes performances avec un nombre réduit de projections. Ci-dessus nous allons évaluer notre méthode avec différents nombre de projections et nous allons montrer que nous pouvons atteindre les mêmes objectifs avec seulement vingt-six projections.

Enfin, notre dernière contribution concerne la phase de la recherche. La hiérarchisation de cette étape a permis de gagner en terme du temps de calcul et même en terme de performances d'indexation.

Les expériences sur les images 2D dans la section précédente nous a permis de déterminer :

- La dimension des histogrammes de couleur, de texture et de contour.
- Les paramètres de pondération α et β .
- Le nombre d'itérations de l'algorithme SA-DWT.

Dans les expériences qui suivent, nous allons maintenir ces résultats (à l'exception des paramètres de pondération), pour caractériser les projections 2D. L'indexation des projections 2D concerne seulement la région d'intérêt, aucune pondération n'est donc utilisée entre la région d'intérêt et le fond de l'objet. Notre algorithme d'indexation comprends deux niveaux de comparaison. À cet effet, la distance entre les indices au premier niveau, est basée sur la fusion pondérée des distances des indices de couleur et de contour. Ceci à l'aide de la formule suivante :

$$d(H_{(r)}, H_{(c)}) = \omega \times d_{Nor}(H_{C_{(r)}}, H_{C_{(c)}}) + (1 - \omega) \times d_{Nor}(H_{E_{(r)}}, H_{E_{(c)}}). \quad (7.7)$$

$d_{Nor}(H_{E_{(r)}}, H_{E_{(c)}})$ représente la distance normalisée entre deux histogrammes du contour de deux projections données.

$d_{Nor}(H_{C_{(r)}}, H_{C_{(c)}})$ représente la distance normalisée entre deux histogrammes de la couleur de deux projections données.

Dans la section suivante, nous allons réaliser des tests pour définir la valeur du paramètre de pondération ω . Contrairement au premier niveau de comparaison, au deuxième niveau la distance est calculée directement, sans aucune fusion entre les distances.

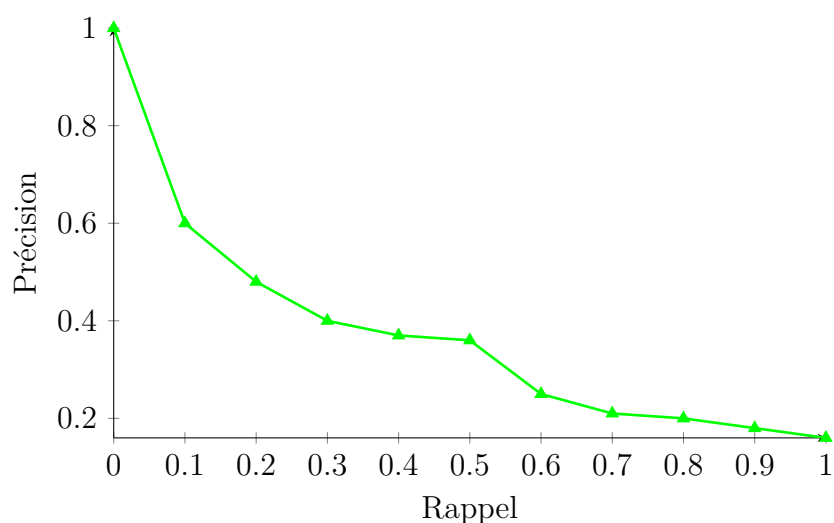
Dans la dernière partie de ce chapitre, nous allons comparer notre méthode aux méthodes d'indexation d'objets 3D :

- La méthode d'indexation par les light field (LFD).
- La méthode d'indexation par les harmoniques 3D.

7.6.1 Détermination du paramètre de pondération ω

Notre schéma de recherche des light field, est basé sur une fusion pondérée des distances entre les histogrammes de couleur, et les distances des histogrammes de contour. Afin de définir la valeur optimale du paramètre de pondération ω (cité ci-dessus), nous évaluons les performances de notre algorithme sur la base d'objets DB3, pour différentes valeurs de ce paramètre.

La courbe *précision/rappel*, de la figure 7.13 montre que les performances de notre algorithme (pour des valeurs de ω inférieures à 0.4), est en relation proportionnelle avec l'augmentation de la valeur du paramètre de pondération. Cette

FIGURE 7.13 – Détermination de la valeur de pondération ω

phase ascendante des performances, est suivie par une régression de la courbe *précision/rappel*. Ce résultat révèle que les performances de notre algorithme sont optimales, pour un paramètre de pondération ω égal à 0.4, ceci peut également être expliqué par la forte capacité de caractérisation de l'attribut contour, comparé à l'attribut couleur.

7.6.2 Expérimentation sur l'alignement

Comme nous l'avons déjà expliqué dans le chapitre 04, la comparaison entre les projections de deux light field différents dépend essentiellement de leurs poses et leur emplacements (leurs conditions d'acquisition) dans le repère 3D. La comparaison aveugle de deux ensembles de projection peut facilement amener à des résultats complètement erronés. Deux objets similaires de différentes poses produisent des projections 2D complètement différentes aux mêmes angles de vue, induisant ainsi à des distances très importantes. La solution à ce problème est d'amener tous les objets de la même classe dans le même repère 3D avec la même pose à l'aide de l'ACP.

Afin de montrer la contribution de la normalisation et de l'alignement des objets 3D, nous avons effectué deux expériences, la première sans aucune normalisation et la seconde avec des objets normalisés et alignés. Les résultats obtenus sont illustrés dans la figure 7.14. On peut facilement constater la nette amélioration en terme de performances entre les deux cas de figure.

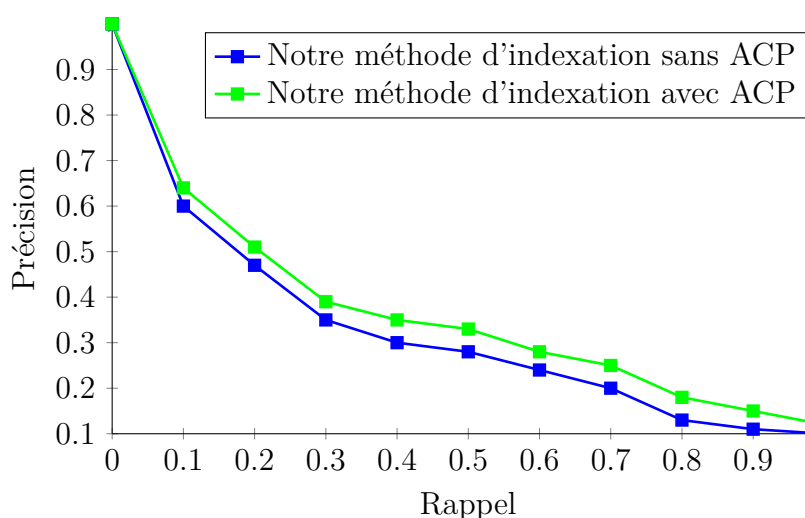


FIGURE 7.14 – Comparaison des performances de la méthode proposée avec et sans ACP et normalisation.

7.6.3 Expérimentation sur le nombre de vues

Les objets 3D rendus par les light field sont parfaitement résolus, grâce à un dispositif de capture qui peut aller jusqu'à 1000 caméras. Quoique cette qualité soit minutieusement recherchée par n'importe quel algorithme d'acquisition de modèles 3D, les approches d'indexation peuvent se contenter d'une qualité assez réduite à condition qu'elle soit capable de capturer les attributs statistiques de l'objet. Dans le chapitre 05, nous avons montré que le nombre des vues 2D peut être réduit à seulement vingt-six vues. Afin d'évaluer les performances de notre algorithme pour des configurations différentes de capture de modèles 3D, nous traçons la courbe *précision/rappel* pour différents nombres de vues. Notre expérience consiste à diviser la sphère de capture à un ensemble de cercles. Ceux-ci sont placés à un angle θ l'un de l'autre. Donc la sphère de projection est divisée en $(360^\circ / \theta)$ cercles, chaque cercle est composé de y points de prise de vue, y est égale à 5. En total le nombre de points de prise de vue est calculé par la formule suivante :

$$\text{Nombre de points de prises de vue} = (360/\theta \times y) - ((360/\theta - 1) \times 2). \quad (7.8)$$

Les résultats de la figure 7.15 montrent une nette dégradation des performances pour les faibles nombres de projections 2D. Deux ensembles se sont formés, un premier ensemble regroupant les algorithmes de projection avec un angle θ compris entre 5° et 45° , cet ensemble enregistre de plus faibles performances, comparées à celles enregistrées par le deuxième groupe, dont l'angle θ est compris entre 45° et 5° . On peut également constater la convergence de la courbe *précision/rappel* pour θ en

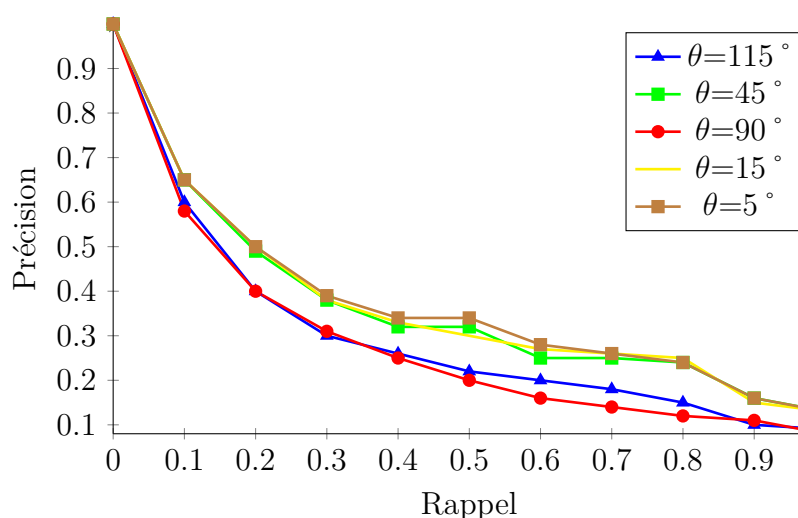


FIGURE 7.15 – Évaluation de la méthode proposée pour différentes valeurs de l'angle θ

dessous de 45° . De ce fait la méthode proposée est basée sur un angle de projection θ , qui est égal à 45° .

7.6.4 Expérimentations sur le processus de la recherche

Nous avons proposé dans le chapitre 06 une optimisation visant à accélérer le processus de la recherche des objets 3D. Celle-ci est basée sur la hiérarchisation de la recherche sur deux niveaux. Dans le premier niveau la comparaison est limitée aux indices de couleur et de contour. Les k premiers objets de cette phase constituent l'entrée du deuxième niveau de comparaison. L'objectif théorique de cette architecture est d'accélérer la recherche, en évitant des comparaisons très lentes des descripteurs de texture, dont la taille est égale à 3000 bins par histogramme.

Afin de prouver la faisabilité de ces notions théoriques, nous proposons dans la figure 7.16 les courbes *précision/rappel* de notre algorithme au premier et deuxième niveau de recherche. Les tests dans cette section sont réalisés sur la base d'objets 3D *DB1*, les résultats obtenus montrent l'amélioration apportée par l'utilisation de cette architecture. En effet, en comparant les performances au premier niveau de la recherche à celles du deuxième niveau, nous nous rendons compte que le deuxième niveau de cette architecture permet d'optimiser les résultats enregistrés au premier niveau.

Dans la figure 7.17 l'algorithme de recherche semble être plus performant au deuxième niveau de comparaison. Ceci est dû au fait que la requête recherchée par-

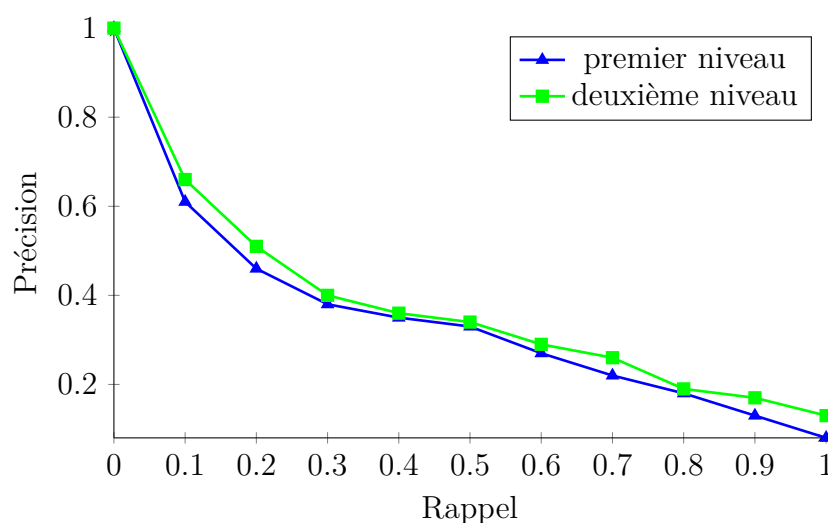


FIGURE 7.16 – Comparaison des performances de notre méthode au premier et deuxième niveau de recherche, à l'aide de la courbe Rappel\Précision.

tage les attributs couleur et forme avec des objets appartenant à d'autres classes. La deuxième requête 7.18 se distingue par une texture caractéristique de l'objet *poisson*, à cet effet une nette amélioration est enregistrée pour cette deuxième requête.

En comparaison avec les figures 7.17(a) et 7.18(a), les figures 7.17(b) et 7.17(b) montre que l'utilisation des indices de la texture a permis d'éliminer les effets négatifs de l'attribut couleur. On peut facilement remarquer que notre algorithme est devenu plus distinctif de l'objet poisson blanc. Cela s'explique par la particularité de la texture de ce genre d'objet par rapport aux autres classes. Dans ces figures, nous montrons les 06 premiers objets retrouvés à chaque niveau. Au premier niveau la liste des 6 premiers objets retrouvés contient des objets similaires à la requête et d'autres qui ne le sont pas. Ceci est dû au fait qu'ils partagent le même attribut couleur. Une nette amélioration est apparue au deuxième niveau. À cet effet, nous pouvons dire que la hiérarchisation du processus de recherche a permis d'optimiser les performances de notre algorithme en terme de pertinence des objets retrouvés par rapport à la requête. Une deuxième amélioration est attendue de cette approche, celle-ci concerne l'accélération du processus de recherche. Nous proposons donc, d'étudier l'impact temporel de la hiérarchisation de la recherche dans la section suivante.

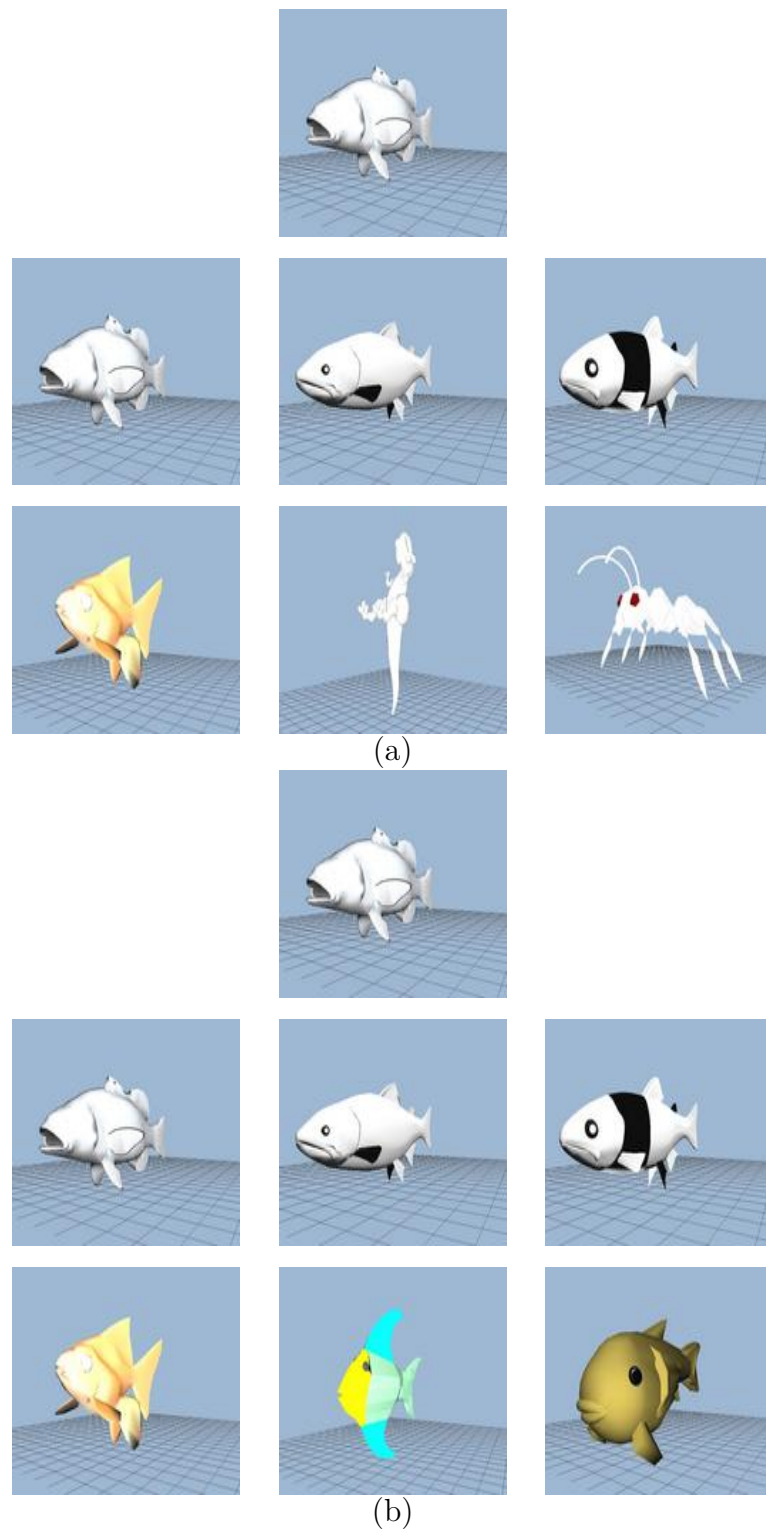


FIGURE 7.17 – Résultat du processus de la recherche. (a) représente les six premiers objets 3D retrouvés au premier niveau de la recherche. (b) représente les six premiers objets 3D retrouvés au deuxième niveau de la recherche.

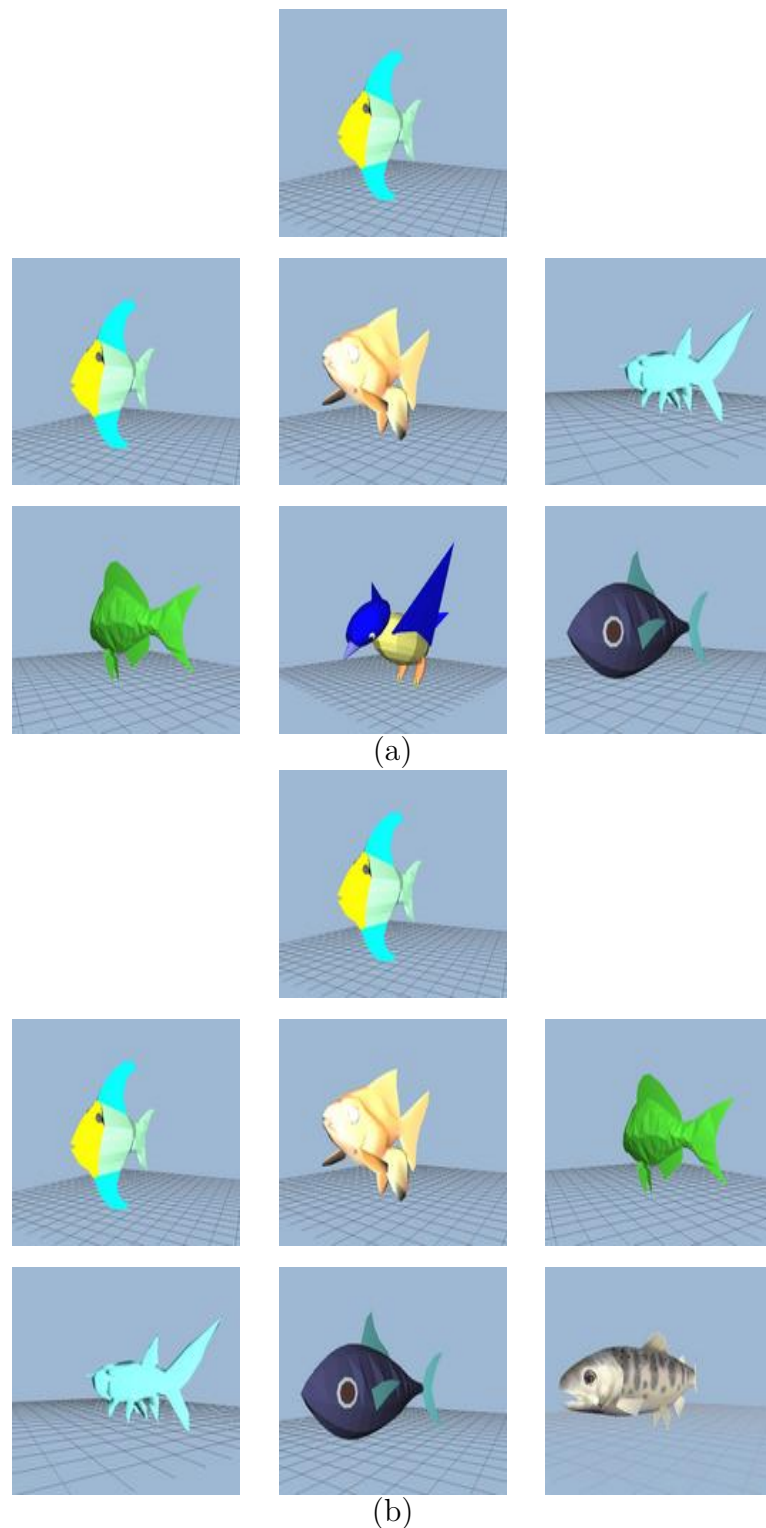


FIGURE 7.18 – Comparaison des performances de l’algorithme de recherche aux deux niveaux de comparaison. (a) représente les six premiers objets 3D retrouvés au premier niveau de la recherche. (b) représente les six premiers objets 3D retrouvés au deuxième niveau de la recherche

7.6.5 Accélération et robustesse du processus de comparaison

Les méthodes d'indexation et de recherche des objets 3D, sont confrontées à différentes contraintes, celles-ci concernent plus particulièrement :

- Le temps de recherche de la requête dans la base d'objets 3D.
- La taille de la mémoire du système, qui est utilisée lors de la phase de recherche.
- La robustesse du résultat de la recherche vis-à-vis les changements de pose.

Ces contraintes étaient derrière la proposition de nos contribution, des chapitres 04 et 06, qui portent sur :

- La normalisation et l'alignement des objets 3D, pour réduire le nombre de comparaisons.
- La hiérarchisation du processus de comparaison.

L'objectif de cette section est l'étude expérimentale de l'impact de ces contributions, sur le temps et la robustesse de la phase de comparaison.

7.6.5.1 Temps de comparaison des objets 3D

Afin d'évaluer l'impact de nos contributions sur le temps de comparaison, nous proposons la mesure de ce temps, pour différentes configurations. Le système utilisé est un PC de processeur Intel Core 2 Duo, il possède 2 GO de RAM.

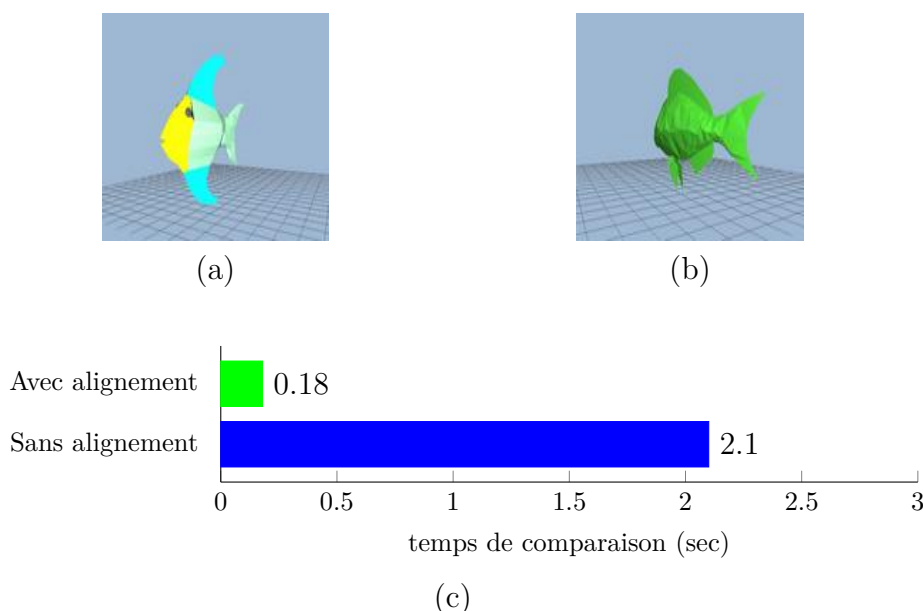


FIGURE 7.19 – Impact de l'alignement et la normalisation des objets sur la durée de la phase de comparaison. (a), (b) les objets à comparer. (c) la durée de comparaison entre les objets (a) et (b)

La figure 7.19 montre que l'alignement et la normalisation des objets 3D, a permis d'accélérer la comparaison de 1.98 sec.

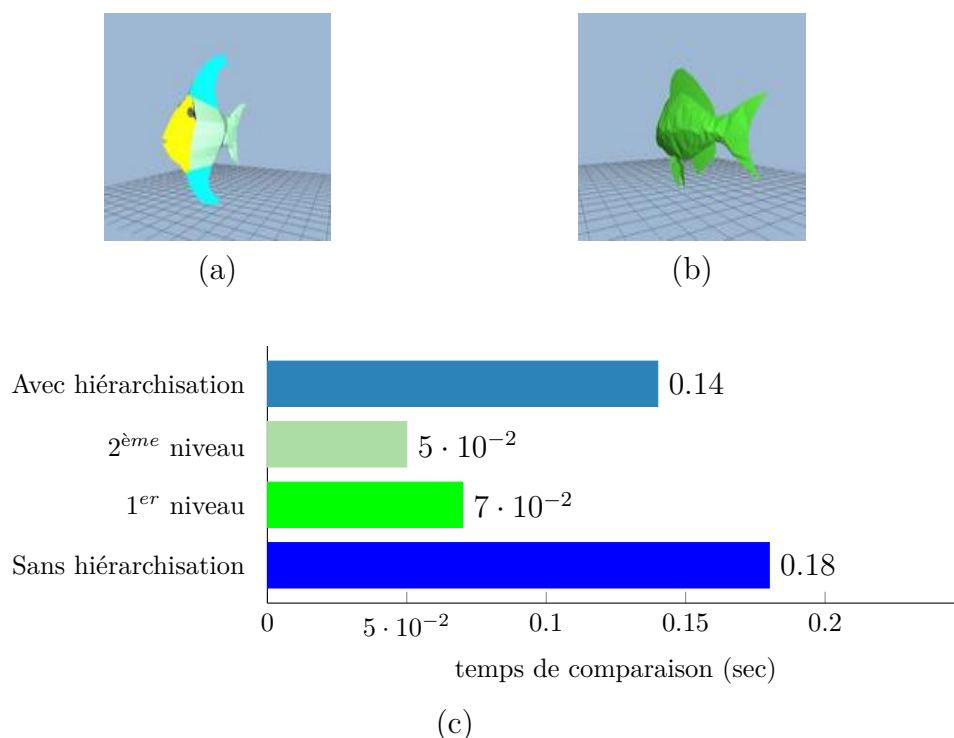


FIGURE 7.20 – Impact temporel de la hiérarchisation sur la phase de comparaison. (a), (b) les objets à comparer. (c) la durée de comparaison entre les objets (a) et (b)

La figure 7.20 montre que la comparaison au premier niveau est plus lente qu'au deuxième niveau. La hiérarchisation du processus de comparaison, a permis toutefois de gagner une durée de 0.06 sec en temps de comparaison.

7.6.5.2 Robustesse aux déformations

Afin d'évaluer le degré de robustesse de notre algorithme. Nous effectuons un ensemble de tests sur 100 objets, sélectionnés aléatoirement de la base DB2. Le but de ces tests, est de mesurer la variation des distances entre l'objet d'origine et les objets déformés. Les déformations abordées dans cette section sont :

- La rotation.
- Le bruit.

Tout d'abord, nous commençons par créer une base de tests, on mesure la distance maximale, et la distance moyenne entre les différents objets. On déforme ensuite ces objets, en changeant leurs poses et en leur ajoutant du bruit. Enfin, nous mesurons la variation des distances maximale et moyenne. La table 7.1 montre une certaine variation des distances suite aux déformations des objets 3D. Concernant le change-

	Rotation	Bruit
Distance maximale	0.865	1.124
Distance moyenne	0.412	0.652

TABLE 7.1 – Variation des distances entre les objets suite à leur déformations

Paramètre	Valeur optimale
Longueur de l'indice de la couleur H_C	400
Longueur de l'indice du contour H_E	100
Longueur de l'indice de la texture H_T	3000
Nombre d'itérations de la SA-DWT	3
Nombre de vues 2D	26
Niveaux de comparaison	2

TABLE 7.2 – Paramètres optimaux de l'algorithme d'indexation et de recherche des light field

ment de la pose, cette variation vient pour confirmer que l'application de l'ACP, ne garantit pas un alignement parfait des objets de la base. Comparée aux autres distances entre les différentes signatures de la base, l'influence de cette variation reste tout de même faible. Donc, on peut conclure que notre méthode est relativement robuste aux déformations des objets 3D.

7.6.6 Étude comparative

Les résultats dans les sections ci-dessus nous ont permis de déterminer la configuration la plus performante de notre algorithme d'indexation des light field. Les paramètres de cette configuration sont tous regroupés dans la table 7.1. Nous allons à présent comparer notre méthode sur les base d'objets 3D :

1. DB3.
2. DB4.

Afin de prouver l'efficacité de notre méthode, nous mettons en place un ensemble de tests expérimentaux. Ceux-ci consistent à exécuter la méthode LFD, la méthode d'indexation à l'aide des harmoniques sphériques, ainsi que notre méthode. Les performances de ces méthodes, sont ensuite comparées en terme de :

- L'efficacité du processus d'indexation et de recherche des objets 3D.
- La durée de la comparaison entre les objets 3D.

– La robustesse des résultats vis-à-vis les déformations des objets 3D.
La courbe de précision-rappel a été tracée pour 12 premiers objets retrouvés.

7.6.6.1 *Éxpérimentations sur la base d'objets 3D DB3*

Nous nous intéressons tout d'abord, à la comparaison de notre méthode à la méthode d'indexation à l'aide des descripteurs des light field (communément appelé LFD). Cette méthode est connue pour être parmi les plus performantes des méthodes d'indexation d'objet 3D basée vues 2D. Nous rappelons également, que cette méthode est basée sur la représentation light field des objets 3D.

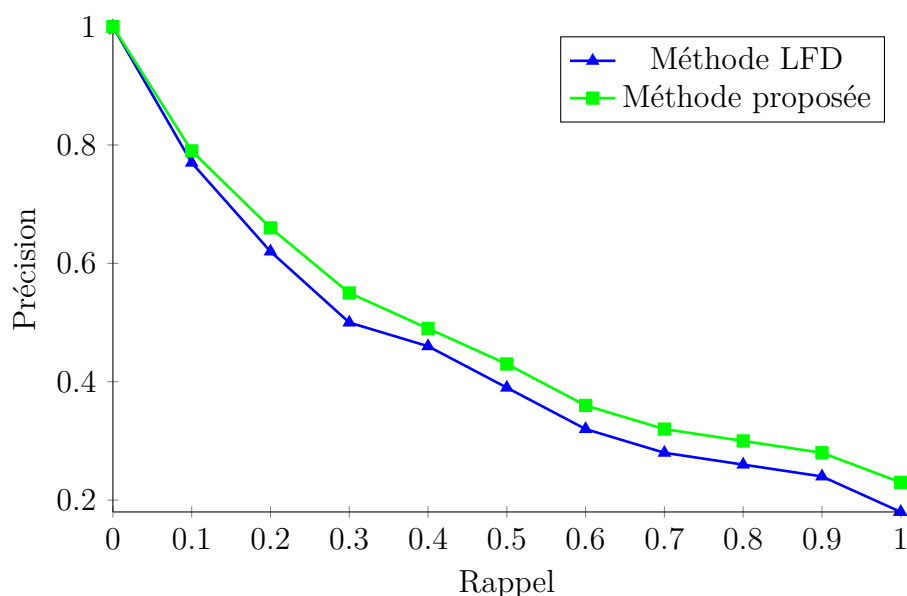


FIGURE 7.21 – *Comparaison des performances de notre méthode à la méthode LFD, à l'aide de la courbe Précision-Rappel.*

La figure représente 7.21 l'évolution du paramètre précision en fonction du paramètre rappel. Selon la méthode utilisée, l'augmentation du rappel est accompagnée par une régression des valeurs du paramètre précision. Cette dégradation est plus importante dans le cas de la méthode LFD. La courbe précision-rappel montre clairement la supériorité des performances de notre méthode, par rapport à la méthode LFD.

7.6.6.2 *Éxpérimentations sur la base d'objets 3D DB4*

De toutes les méthodes d'indexation basées caractéristiques géométriques, que nous avons présentée, la méthode d'indexation à l'aide des harmoniques sphériques

(HS), semblent être la plus performante. Nous avons donc choisi de nous comparer à cette méthode.

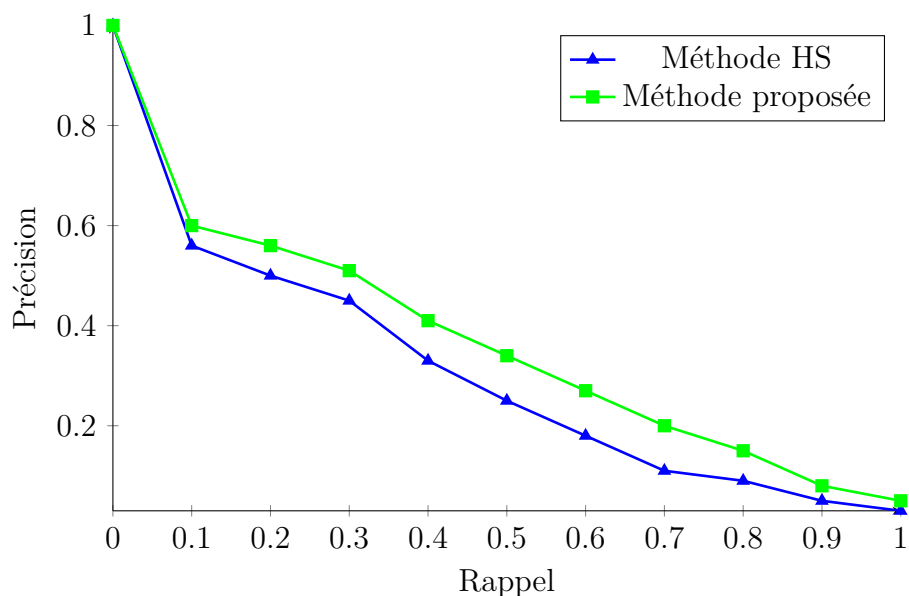


FIGURE 7.22 – Comparaison des performances de notre méthode par rapport à la méthode HS, à l'aide de la courbe Précision-Rappel.

La courbe précision-rappel de la figure 7.22, indique que notre méthode retourne les meilleurs résultats par rapport à la méthode HS. On note qu'il y'a tout de même une certaine convergence des performances, sur l'intervalle des valeurs du paramètre rappel.

Les tests que nous avons menés, montrent l'efficacité de notre méthode par rapport à la méthode LFD et la méthode HS. Cependant, afin d'avoir une vision plus approfondie, nous proposons d'autres tests qui sont basés sur chacune des classes de la base d'objets DB4. L'objectif de ces tests est d'étudier sous différentes contraintes les performances des trois méthodes. Ces contraintes portent sur la variation de la pose des objets au sein même de la même classe. La deuxième contrainte est imposée par la forte ressemblance qui existe entre un ensemble de classes différentes.

La comparaison des trois méthodes est présentée dans la figure 7.23. Notons que les résultats proviennent de nos propres travaux et des évaluations de Saupe [115] et de Chen [32]. La comparaison sur les classes de la base d'objets DB2 montre l'efficacité de ces méthodes à caractériser les objets de la classe *avion* et de la classe *voiture*, par rapport aux objets de la classe *épée* et de la classe *bouteille*. Les descripteurs des

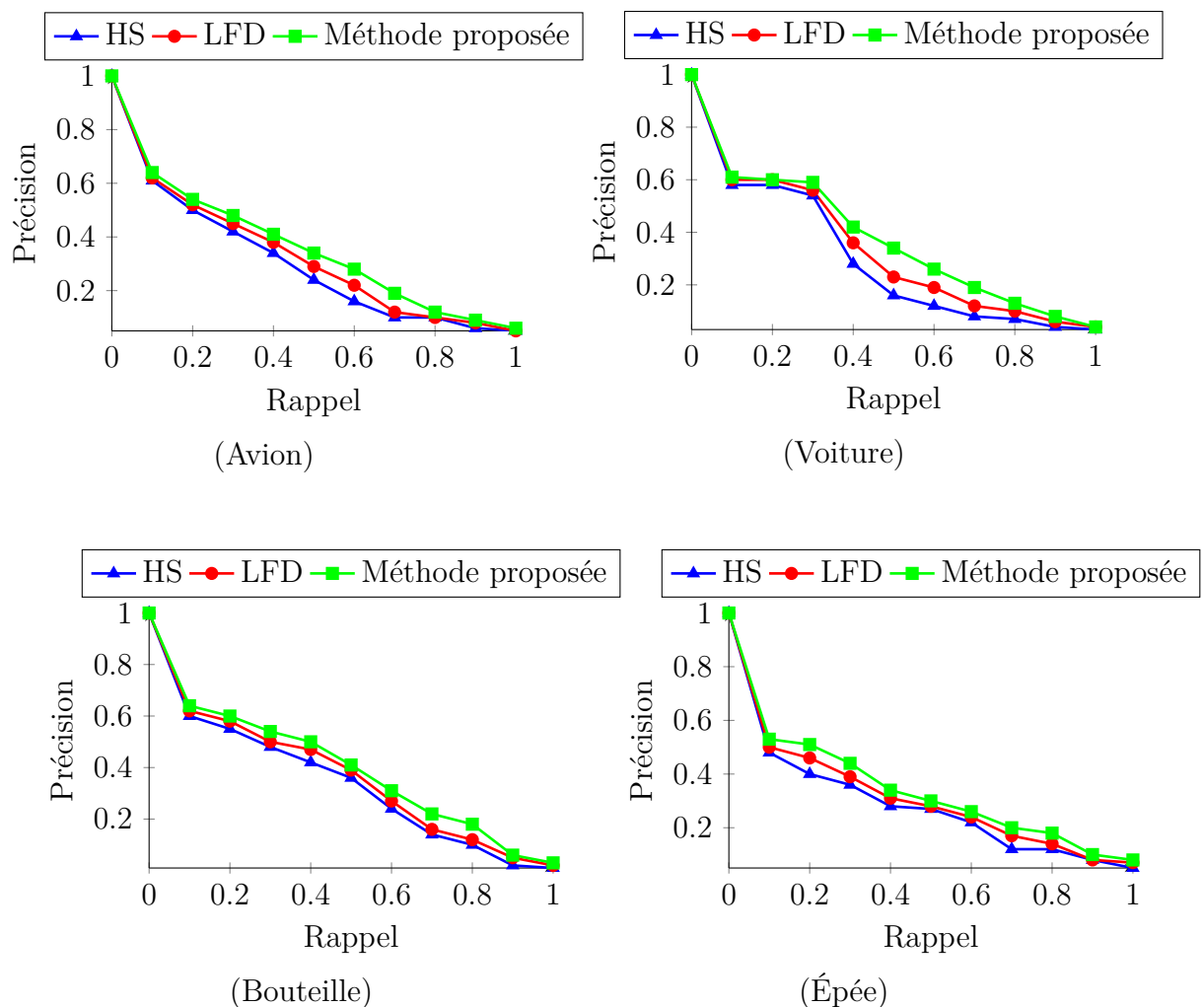


FIGURE 7.23 – Comparaison des méthodes LFD, HS et la méthode proposée à l'aide de la courbe précision-rappel

harmoniques sphériques, sont principalement utilisés pour résoudre les problèmes invariants par rotation. Ceci est confirmé par les résultats obtenus dans les classes *avion* et *voiture*, où le degré de la variation de la pose est plus important que celui des classes *bouteille* et *épée*. Cependant, ces performances se dégradent rapidement lorsqu'il est question de distinguer entre des classes, dont le degré de ressemblance inter-classe est très important. Ceci correspond aux classes *épée* et *bouteille*, qui se distinguent par des formes moins discriminatives que les classes *avion* et *voiture*. La caractérisation basée vue 2D réduit significativement ce degré de ressemblance, en capturant l'attribut forme sous différents angles de capture. Ceci s'est traduit par une nette amélioration des performances obtenues à l'aide de notre méthode et à l'aide de la méthode LFD.

7.6.6.3 Comparaison de la robustesse vis-à-vis les déformations des objets

Afin de comparer les trois méthodes (méthode proposée, LFD et HS) en terme de leur robustesses vis-à-vis la déformation des objets 3D, nous avons mis en place un test sur 100 objets de la base DB2. Ce test consiste à calculer la variation des distances maximales et minimales entre les différents objets, suite à leur déformations. Le tableau 7.3, montre que la méthode LFD est plus robuste aux rotations des objets, que les autres méthodes, ceci est dû au nombre important de poses des objets 3D, qui sont prises en compte par cette méthode lors de la phase de comparaison. Néanmoins, la supériorité de la méthode LFD est quasiment négligeable. En terme de robustesse vis-à-vis du bruit, notre méthode semble être la plus robuste comparée aux méthodes LFD et HS.

	Rotation			Bruit		
Méthode	M. proposée	LFD	HS	M. proposée	LFD	HS
Distance maximale	0.865	0.534	0.882	1.124	1.140	1.166
Distance moyenne	0.412	0.365	0.432	0.652	0.698	0.734

TABLE 7.3 – Comparaison de la robustesse de la méthode proposée aux méthodes LFD et HS

7.6.6.4 Temps de comparaison des objets 3D

Nous avons introduit dans le chapitre 6 une nouvelle structure de recherche des light field. L'objectif de cette architecture, est de réduire de l'espace de recherche grâce à une première comparaison à l'aide des descripteurs de couleur et de contour. La comparaison entre les descripteurs de texture au deuxième niveau est ainsi accélérée. Comparée à la méthode LFD, notre méthode est basée sur un nombre réduit de projections 2D des light field, cela permet également d'économiser le temps nécessaire aux comparaisons. Afin d'illustrer l'impact temporel de ces deux contributions, nous proposons de calculer le temps de comparaison de deux objets à l'aide des trois méthodes citées ci-dessus. La figure 7.24 montre que l'utilisation de notre méthode a permis d'accélérer la comparaison entre les deux objets, de 27% par rapport à la méthode LFD. L'alignement de la pose de la méthode HS lui a permis de dépasser les performances temporelles de la méthode LFD.

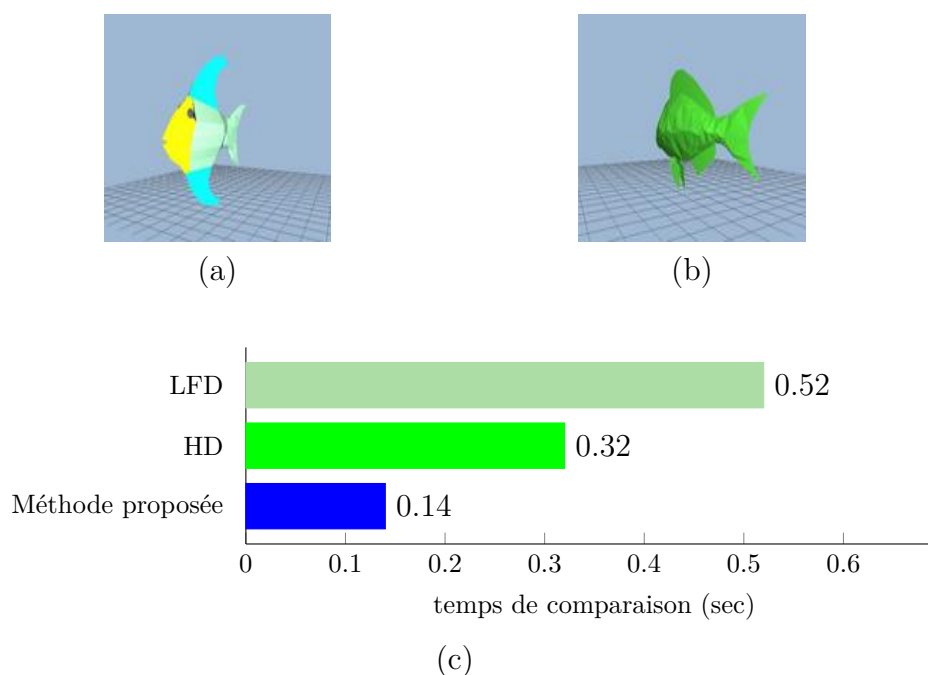


FIGURE 7.24 – *Comparaison temporelle entre la méthode proposée, LFD et HS. (a), (b) les objets à comparer. (c) la durée de comparaison entre les objets (a) et (b)*

7.7 Conclusion

Dans ce chapitre, nous avons présenté les résultats sur des bases d'images 2D et sur les bases de modèles 3D. Les deux volets abordés dans notre thèse à savoir l'indexation et la recherche des requêtes 2D et 3D, ont été évalués dans ce chapitre. Les résultats obtenus montrent que Notre méthode enregistre de bons résultats sur les deux types de requêtes.

Dans ce chapitre, nous avons également validé les études théoriques de nos contributions, grâce à un ensemble de tests. Ces tests nous ont permis également de paramétrer nos approches et de fixer la taille des différentes variables de notre algorithme d'indexation et de recherche des objets 3D. D'autre part, nous avons mis en évidence l'amélioration de la robustesse de notre méthode vis-à-vis la rotation et la déformation des objets 3D, grâce à leur normalisation et leur d'alignement.

Un intérêt particulier dans ce chapitre est porté à l'accélération du processus de recherche. Dans ce contexte, nous avons proposé deux contributions. La première porte sur la réduction du nombre de comparaisons. Afin de fixer le nombre maximal de projections nécessaires à la caractérisation des light field, nous avons évalué notre algorithme pour différent nombre de projections. La deuxième contribution a

porté sur la hiérarchisation du processus de recherche, celle-ci a été validée par des tests de comparaison temporelle des comparaisons. Ces propriétés de robustesse et de réduction de nombre de projections, sont importantes puisque cela signifie que le nombre de projections peut être réduit, sans perte importante de performances.

Enfin, nous confrontons notre méthode à deux méthodes d'indexation et de recherche de modèles 3D. Les résultats obtenus ont confirmé la supériorité de notre méthode en terme de pertinence des résultats à la requête, en terme de robustesse vis-à-vis les déformations de l'objet et enfin en terme de temps de comparaison.

Conclusion générale

Vu l'importante extension de l'utilisation des nouvelles technologies de l'information et de la communication, les algorithmes de traitement et de communication de l'information sont devenus de plus en plus indispensables. L'information utilisée et échangée peut être du texte, du son, de l'image ou même de la vidéo. Etant l'image est l'information la plus expressive, la plupart des outils de communication utilisés actuellement se basent sur l'échange et la visualisation d'images 2D, 3D ou même de type vidéo. A cet effet, la recherche dans d'outils informatiques qui seront capables de manipuler les images, et de rechercher dans les gigantesques bases d'images connaît un essor et un intérêt particulier.

Rechercher une image dans une base revient souvent comparer son contenu par rapport aux autres images stockées dans la base. Ce contenu peut être en effet organisé sous différentes formes : il peut être organisé sous forme d'histogramme représentant la répartition des intensités de pixels dans l'image, les moments servent aussi de formes de représentation des primitives de l'image. Donc, indexer une image consiste généralement à décrire l'image par des mots clés, ou par des signatures représentant son contenu.

Quoique la plupart des moteurs de recherche se basent actuellement sur l'annotation textuelle de l'image, il s'est avéré que cette technique est très difficile à achever sur des bases de millions d'images. L'éventualité était de développer des méthodes d'indexation automatiques à partir du contenu de l'image, c'est ce qu'on appelle communément les méthodes d'indexation de l'image basée contenu. Ces indices ne sont en réalité que des mesures statistiques effectuées sur les primitives de l'image, tel que la couleur, la texture, ou même la forme.

Le but de ce travail de thèse était de développer une méthode d'indexation et de recherche de modèle 3D, l'originalité de notre travail réside dans la façon avec laquelle on a traité ce sujet, en considérant l'indexation de modèles 3D, comme un sujet d'indexation de vues 2D ; grâce à cette approche, on a pu minimiser la taille du modèle à indexer, ainsi que le nombre de vues. Les deux contraintes traités par notre

méthode, ont porté sur l'accélération du processus de recherche et l'optimisation de la pertinence des résultats aux requêtes.

Dans ce contexte, on a étudié et proposé plusieurs solutions. La première solution porte sur la réduction du nombre de vues 2D à caractériser. Nous avons également proposé d'aligner tous les objets de la base pour réduire le nombre de comparaisons de chaque objet 3D. Une dernière contribution a porté sur la hiérarchisation du processus de recherche et la réduction de l'espace de recherche.

Enfin, nous avons validé nos résultats, grâce à des tests sur les bases d'images 2D et de modèles 3D. Nous avons également comparé nos résultats à plusieurs méthodes d'indexation d'images 2D et d'objets 3D.

Les perspectives de ce travail de thèse seraient d'exploiter d'autres informations de la compression des light field lors de la phase d'indexation. Comme nous l'avons expliqué aux chapitres 2 et 5, la redondance inter-vue peut être compressée à l'aide de la compensation de la disparité. Les distances inter-blocs calculées par cette technique, peuvent être à la base d'un quatrième descripteur des light field. Notre système d'indexation est basé sur un ensemble d'indices, qui sont construits à partir des coefficients de la transformée en ondelettes. Dans notre travail, nous avons utilisé des filtres prédéfinis de la transformée en ondelettes CDF 9/7, de cette façon un aspect très intéressant n'a pas été traité par notre travail, ceci peut être fixé comme perspective. Cet aspect porte sur l'étude et la conception des filtres adaptés à notre système d'indexation des light field. Enfin, dans la partie expérimentale nous avons testé nos algorithmes d'indexation des objets 3D sur des bases de milliers d'objets. Nos perspectives dans ce domaine est d'adapter nos algorithmes pour qu'ils puissent s'exécuter sur des bases plus larges.

Annexe 1

Les coefficients des filtres de synthèse de la transformée en ondelettes Le filtre de l'analyse passe bas G :

$$\begin{aligned}G(1) &= 0.03782845550726. \\G(2) &= -0.02384946501956. \\G(3) &= -0.11062440441844. \\G(4) &= 0.37740285561283. \\G(5) &= 0.85269867900889. \\G(6) &= 0.37740285561283. \\G(7) &= -0.11062440441844. \\G(8) &= -0.02384946501956. \\G(9) &= 0.03782845550726.\end{aligned}$$

Le filtre de l'analyse passe bas Q :

$$\begin{aligned}Q(1) &= 0.06453888262870. \\Q(2) &= -0.04068941760916. \\Q(3) &= -0.41809227322162. \\Q(4) &= 0.78848561640558. \\Q(5) &= -0.41809227322162. \\Q(6) &= -0.04068941760916. \\Q(7) &= 0.06453888262870.\end{aligned}$$

Annexe 2

Exemple de la structure d'un objet .obj

nom de l'objet.obj

#

v 0.0 0.0 0.0

v 10.0 11.0 1.0

v 22.1 12.0 0.0

v_t 0.25 0.0

v_t 0.5 0.0

v_t 0 0.25

v_n 0.0 0.0 1.0

v_n 0.0 0.0 -1.0

v_n 0.0 1.0 0.0

f 1/11/2 7/14/2 5/12/2

f 1/11/2 3/13/2 7/14/2

f 1/7/6 4/4/6 3/3/6

v représentent les coordonnées cartésiennes des sommets des formes triangulaires.

v_t représentent les coordonnées des textures.

v_n représentent les normales des formes triangulaires.

f représentent les indices des formes triangulaires.

Annexe 3

Considérons un ensemble de points P d'un objet O , nous cherchons à construire une sphère minimale S (d'origine c et de rayon r), qui enveloppe cet objet, tel que :
 $S = \{P \in O, \|p - c\| \leq r\}$

La construction de la sphère minimale consiste à minimiser la sphère initiale S_0 en itérant les deux opérations suivantes :

- Le déplacement du centre c .
- L'élimination des sommets négatifs.

L'arrêt de ce processus itératif est conditionné par la formation d'une forme convexe, autour du centre de la sphère. A cette étape, la formule citée ci-dessus est vérifiée.

Bibliographie

- [1] <http://displayblocks.org/light-field-archive/>.
- [2] <https://www.apprendre-a-dessiner.org/>.
- [3] http://www.3dvf.com/forum/3dvf/Ateliers/4/occlusion-premier-sujet_60_1.htm.
- [4] <http://amp.ece.cmu.edu/Droiects/3DModelRe~ev>.
- [5] http://www.vlfeat.org/matlab/vl_kmeans.html.
- [6] <http://sites.google.com/site/dctresearch/Home/content-based-image-retrieval>.
- [7] <http://shape.cs.princeton.edu/benchmark/>.
- [8] <http://3d.csie.ntu.edu.tw/~dynamic/benchmark/>.
- [9] <http://wang.ist.psu.edu/docs/related/>.
- [10] E. H. ADELSON ET J. R. BERGEN, *The plenoptic function and the elements of early vision*, in *Computational Models of Visual Processing*, 1991, p. 3–20.
- [11] D. ADLER ET D. MURDOCH, *rgl : 3D visualization device system (OpenGL)*, 2014.
- [12] F. J. AHERNE, N. A. THACKER ET P. I. ROCKETT, *The bhattacharyya metric as an absolute similarity measure for frequency coded data*, *Kybernetika*, 34 (1998), p. 363–368.
- [13] Z. ALMERAJ, C. S. KAPLAN ET P. ASENTE, *Patch-based geometric texture synthesis*, in *Workshop on Computational Aesthetics*, Anaheim, California, USA, 2013. Proceedings, 2013, p. 15–20.
- [14] T. F. ANSARY, M. DAOUDI ET J. VANDEBORRE, *A bayesian 3-d search engine using adaptive views clustering*, *IEEE Trans. Multimedia*, 9 (2007), p. 78–88.
- [15] O. K. AU, C. TAI, H. CHU, D. COHEN-OR ET T. LEE, *Skeleton extraction by mesh contraction*, *ACM Trans. Graph.*, 27 (2008), p. 44 :1–44 :10.
- [16] ———, *Skeleton extraction by mesh contraction*, *ACM Trans. Graph.*, 27 (2008), p. 44 :1–44 :10.

- [17] C. BAI, *Analyse d'images pour une recherche d'images basée contenu dans le domaine transformée*, Thèse doctorat, 2013.
- [18] C. BAI, J. ZHANG, Z. LIU ET W. ZHAO, *K-means based histogram using multiresolution feature vectors for color texture database retrieval*, *Multimedia Tools Appl.*, 74 (2015), p. 1469–1488.
- [19] L. BELHALLOUCHE ET K. BELLOULATA, *Indexation d'images par la transformée fractale*, *International Conference on Image and Signal Processing and their Applications*, 1 (2009), p 22–29.
- [20] L. BELHALLOUCHE, K. BELLOULATA ET K. KPALMA, *3d models retrieval in the compressed domain*, *Signal, Image, Geometry, Modelling, Approximation Conference*, 1 (2012), p 04–08.
- [21] —, *Light field retrieval in compressed domain*, *Winter School of Computer Graphics (WSCG). 21st International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'*, 21 (2013), p 4.
- [22] —, *A new approach to region based image retrieval using shape adaptive discrete wavelet transform*, *International Journal of Image, Graphics and Signal Processing (IJIGSP)*, 08 (2016), p. 1–14.
- [23] L. BELHALLOUCHE ET J. F. LAYERLE, *Reconnaissance de visages à l'aide des modèles actifs d'apparence*, *International Conference On Electronics Engineering*, (2013).
- [24] K. BELLOULATA, L. BELHALLOUCHE, A. BELALIA ET K. KPALMA, *Region based image retrieval using shape-adaptive DCT*, *IEEE China Summit International Conference on Signal and Information Processing*, 08 (2014), p. 470–474.
- [25] G. BERND, *Fast and robust smallest enclosing balls*, in *Proceedings of the 7th Annual European Symposium on Algorithms*, 1999, p. 325–338.
- [26] S. BIASOTTI, S. MARINI, M. SPAGNUOLO ET B. FALCIDIENO, *Subpart correspondence by structural descriptors of 3d shapes*, *Computer-Aided Design*, 59 (2006), p 1002–1019.
- [27] K. S. BING, L. YIN, T. XIN ET S. HEUNG-YEUNG, *Image-based rendering*, *Found. Trends. Comput. Graph. Vis.*, 2 (2006), p. 173–258.
- [28] H. BLUM, *A Transformation for Extracting New Descriptors of Shape*, in *Models for the Perception of Speech and Visual Form*, W. Wathen-Dunn, éd., MIT Press, Cambridge, 1967, p. 362–380.
- [29] E. BRIBIESCA, *Measuring 3-d shape similarity using progressive transformations*, *Pattern Recognition*, (1996).
- [30] R. CAREY, *The virtual reality modeling language explained*, *IEEE MultiMedia*, 5 (1998), p. 84–93.

- [31] C.-L. CHANG, X. ZHU, P. RAMANATHAN ET B. GIROD, *Light field compression using disparity-compensated lifting and shape adaptation*, IEEE Transactions On Image Processing, 15 (2006), p 793–806.
- [32] D. CHEN, X. TIAN, Y. SHEN ET M. OUHYOUNG, *On visual similarity based 3d model retrieval*, Comput. Graph. Forum, 22 (2003), p. 223–232.
- [33] S. E. CHEN, *Quicktime vir-an image-based approach to virtual environment navigation*, In ACM SIGGRAPH Computer Graphics, 29 (1995), p 29–38.
- [34] —, *Quicktime VR : an image-based approach to virtual environment navigation*, in Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1995, Los Angeles, CA, USA, August 6-11, 1995, 1995, p. 29–38.
- [35] W.-C. CHEN, J.-Y. BOUGUET, M. H. CHU ET R. GRZESZCZUK, *Efficient representation and hardware rendering of surface light fields*, In ACM SIGGRAPH Computer Graphics, 15 (2002), p 447–456.
- [36] C. CHOUDHURY, Y. TARUNY, A. RAJWADEY ET S. CHAUDHURI, *Low bit-rate compression of video and light-field data using coded snapshots and learned dictionaries*, International Workshop on Multimedia Signal Processing (MMSp), 15 (2015), p. 4673–4678.
- [37] D. COMANICIU ET P. MEER, *Mean shift : A robust approach toward feature space analysis*, IEEE Trans. Pattern Anal. Mach. Intell., 24 (2002), p. 603–619.
- [38] E. COPSON, *Metric Spaces*, Cambridge Tracts in Mathematics, Cambridge University Press, 1968.
- [39] D. R. H. P. S. D., *Pattern classification*, in second edition. New York : Wiley., 2001.
- [40] R. DATTA, D. JOSHI, J. LI ET J. Z. WANG, *Image retrieval : Ideas, influences, and trends of the new age*, ACM Comput.Surv, 40 (2008), p 5 :1–60.
- [41] D. EDMUNDSON, G. SCHAEFER ET M. E. CELEBI, *Robust texture retrieval of compressed images*, in 19th IEEE International Conference on Image Processing, ICIP 2012, Lake Buena Vista, Orlando, FL, USA, September 30 - October 3, 2012, 2012, p. 2421–2424.
- [42] O. EGGER, *Region representation using nonlinear techniques with applications to image and video coding*, Thèse de doctorat, (1997).
- [43] O. EGGER, F. BOSSEN ET T. EBRAHIMI, *Shape-adaptive wavelet transform for zerotree coding*, in Proc. Eur.Workshop Image Analysis and Coding for TV, HDTV and Multimedia Application, (1996), p 201–208.
- [44] J. FAN, Y. GAO, H. LUO ET G. XU, *Automatic image annotation by using concept-sensitive salient objects for image content representation*, in SIGIR

- 2004 : Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Sheffield, UK, July 25-29, 2004, 2004, p. 361–368.
- [45] J. P. FARRUGIA, *Modèles de Vision et Synthèse d'Images*, Thèse doctorat, 2002.
- [46] J. FOURNIER, *Indexation d'images par le contenu et recherche interactive dans les bases généralistes*, Thèse doctorat, Université de Cergy-Pontoise, France, 2002.
- [47] H. FU, D. COHEN-OR ET G. D. ET ALLA SHEFFER, *Upright orientation of man-made objects*, ACM SIGGRAPH papers, (2008).
- [48] T. FUNKHOUSER, M. KAZHDAN, P. SHILANE, P. MIN, W. KIEFER, A. TAL, S. RUSINKIEWICZ ET D. DOBKIN, *Modeling by example*, ACM SIGGRAPH papers, (2004).
- [49] T. FUNKHOUSER, P. MIN, M. KAZHDAN, J. CHEN, A. HALDERMAN ET D. D. ET D. JACOBS, *A search engine for 3d models*, ACM Transactions on Graphics, (2003).
- [50] T. A. FUNKHOUSER, P. MIN, M. M. KAZHDAN, J. CHEN, A. HALDERMAN, D. P. DOBKIN ET D. P. JACOBS, *A search engine for 3d models*, ACM Transactions on Graphics, (2003).
- [51] N. GAGVANI ET D. SILVER, *Parameter-controlled volume thinning.*, CVGIP : Graph. Models Image Process., 48 (1999), p 149–164.
- [52] —, *Parameter-controlled volume thinning*, Graphical Models and Image Processing, 61 (1999), p. 149–164.
- [53] J. GALVEZ ET M. CANTON, *Normalization and shape recognition of three dimensional objects by 3d moments*, Pattern Recognition, (1993).
- [54] Y. GAO ET Q. DAI, *View-based 3d object retrieval : Challenges and approaches*, IEEE MultiMedia, 21, p. 52–57.
- [55] M. GARLAND ET P. S. HECKBERT, *Surface simplification using quadric error metrics*, in Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1997, Los Angeles, CA, USA, August 3-8, 1997, 1997, p. 209–216.
- [56] G. HETZEL, B. LEIBE, P. LEVI ET B. SCHIELE, *3d object recognition from range images using local feature histograms*, in 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), with CD-ROM, 8-14 December 2001, Kauai, HI, USA, 2001, p. 394–399.
- [57] HEYDEN, ANDERS ET M. POLLEFEYS, *Multiple view geometry.*, Emerging Topics in Computer Vision, 90 (2005), p. 180–189.

- [58] M. HILAGA, Y. SHINAGAWA, T. KOMURA ET T. L. KUNII, *Topology matching for fully automatic similarity estimation of 3d shapes.*, In SIGGRAPH 2001, 49 (2001), p 203–212.
- [59] ———, *Topology matching for fully automatic similarity estimation of 3d shapes*, in Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2001, Los Angeles, California, USA, August 12-17, 2001, 2001, p. 203–212.
- [60] W. HUANG, Y. GAO ET K. L. CHAN, *A review of region-based image retrieval*, Signal Processing Systems, 59 (2010), p. 143–161.
- [61] S. I. IBRAHIM, A. ABUHAIBA, A. RUBA ET A.SALAMAH, *efficient global and region content based image retrieval*, international journal of image, graphics and signal processing, 5 (2012), p. 38–46.
- [62] I. IHM, S. PARK ET R. K. LEE, *Rendering of spherical light fields*, in 5th Pacific Conference on Computer Graphics and Applications (PG '97), 13-16 October 1997, Seoul, Korea, 1997, p. 59–68.
- [63] K. P. JASMINE ET P. R. KUMAR, *Color and local maximum edge patterns histogram for content based image retrieval*, International Journal of Intelligent Systems and Applications(IJISA), 6 (2014), p. 66–74.
- [64] S. JAYANTI, Y. KALYANARAMAN, N. IYER ET K. RAMANI, *Developing an engineering shape benchmark for CAD models*, Computer-Aided Design, 38 (2006), p. 939–953.
- [65] N. JHANWAR, S. CHAUDHURI, G. SEETHARAMAN ET B. ZAVIDOVIQUE, *Content based image retrieval using motif cooccurrence matrix*, Image Vision Comput., 22 (2004), p. 1211–1220.
- [66] F. JING, M. LI, H. ZHANG ET B. ZHANG, *An efficient and effective region-based image retrieval framework*, IEEE Trans. Image Processing, 13 (2004), p. 699–709.
- [67] S. B. KANG, R. SZELISKI ET P. M. ANANDAN, *The geometry-image representation tradeoff for rendering*, Proceedings of International Conference on Image Processing, Vancouver, BC, Canada, 2 (2000), p 13–16.
- [68] P. KAUFF ET K. SCHUUR, *Shape-adaptive DCT with block-based DC separation and dc correction*, IEEE Trans. Circuits Syst. Video Techn, 08 (1998), p. 237–242.
- [69] M. KAZHDAN, *An approximate and efficient method for optimal rotation alignment of 3d models*, IEEE Transactions on Pattern Analysis and Machine Intelligence, (2007).
- [70] R. KOENEN, *Overview of the mpeg-4 standard*, ISO/IECJTC/SC29/WG21, (1999).

- [71] T. KOHONEN, *Self-Organizing Maps*, 2001.
- [72] D. LAKSHMI, A. DAMODARAM, M. SREENIVASA ET J. LAL, *Content based image retrieval using signature based similarity search*, Indian J. Science and Technology, 1 (2008), p. 80–92.
- [73] M. LEONARD ET B. GARY, *Plenoptic modeling : An image-based rendering system*, in Proceedings of the 22Nd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '95, 1995, p. 39–46.
- [74] M. LEVOY ET P. HANRAHAN, *Light field rendering*, ACM SIGGRAPH, 39 (2006), p. 46–55.
- [75] J. LI, J. WANG ET G. WIEDERHOLD, *Integrated region matching for image retrieval*, in in Proc. 2000 ACM Multimedia Conf., Los Angeles, vol. 5, 2000, p. 147–156.
- [76] S. LI ET W. LI, *Shape-adaptive discrete wavelet transform for coding arbitrarily shaped texture*, Proceedings of SPIE - The International Society for Optical Engineering, (1997).
- [77] Y. LIAN ET X. SHEN, *Image-based hierarchical method of 3d model retrieval*, in Proceedings of the 2014 International Conference on Virtual Reality and Visualization, ICVRV, 2014, Shenyang, China, August 30-31, 2014, 2014, p. 160–165.
- [78] S. LIN, Y. GUO, Y. LIANG, Q. CHEN ET Y. WU, *3d model retrieval based on skeleton*, in 10th IEEE International Conference on Networking, Architecture and Storage, NAS 2015, Boston, MA, USA, August 6-7, 2015, 2015, p. 321–325.
- [79] J. J. LITTLE, *Extended gaussian images, mixed volumes, shape reconstruction*, in Proceedings of the First Annual Symposium on Computational Geometry, Baltimore, Maryland, USA, June 5-7, 1985, 1985, p. 15–23.
- [80] Y. LIU, D. S. ZHANG, G. LU ET W. Y. MA, *A survey of content-based image retrieval with high-level semantics*, ACM Comput.Surv, 40 (2007), p 262–282.
- [81] L. MA ET R. C. STAUNTON, *Optimum gabor filter design and local binary patterns for texture segmentation*, Pattern Recognition Letters, 29 (2008), p. 664–672.
- [82] M. A. MAGNOR, P. EISERT ET B. GIROD, *Model-aided coding of multi-viewpoint image data*, in Proceedings of the 2000 International Conference on Image Processing, ICIP 2000, Vancouver, BC, Canada, September 10-13, 2000, 2000, p. 919–922.
- [83] M. A. MAGNOR, A. ENDMANN ET B. GIROD, *Progressive compression and rendering of light fields*, in Proceedings of the 2000 Conference on Vision Modeling and Visualization (VMV-00), Saarbrücken, Germany, November 22-24, 2000, 2000, p. 199–204.

- [84] M. A. MAGNOR ET B. GIROD, *Data compression for light-field rendering*, IEEE Trans. Circuits Syst. Video Techn., 10 (2000), p. 338–343.
- [85] S. MAHMOUDI ET M. DAOUDI, *3d models retrieval by using characteristic views*, in 16th International Conference on Pattern Recognition, ICPR 2002, Quebec, Canada, August 11-15, 2002., 2002, p. 457–460.
- [86] O. MARQUES, L. M. MAYRON, G. B. BORBA ET H. R. GAMBÀ, *Using visual attention to extract regions of interest in the context of image retrieval*, in Proceedings of the 44th Annual Southeast Regional Conference, 2006, Melbourne, Florida, USA, March 10-12, 2006, 2006, p. 638–643.
- [87] L. MCMILLAN, *An image-based approach to three-dimensional computer graphics*, Technical report, Ph.D. Dissertation, UNC Computer Science, (1999).
- [88] L. MCMILLAN ET G. BISHOP, *Plenoptic modeling : An image-based rendering system*, Computer Graphics (SIGGRAPH'95), (1995), p 39—46.
- [89] B. MERCIER, *Reconstruction et analyse automatiques pour le rééclairage d'objets basés-image*, Thèse doctorat, 2006.
- [90] H. MOBAHI, S. R. RAO, A. Y. YANG, S. S. SASTRY ET Y. MA, *Segmentation of natural images by texture and boundary compression*, International Journal of Computer Vision, 95 (2011), p. 86–98.
- [91] K. MUELLER, *Core-experiments on mpeg-4 video—efficient coding*, ISO/IEC JTC/SC29/WG11, MPEG-96-N1250, (1996).
- [92] P. MÜLLER, G. ZENG, P. WONKA ET L. J. V. GOOL, *Image-based procedural modeling of facades*, ACM Trans. Graph., 26 (2007), p 85.
- [93] T. NAPOLÉON, *Indexation multi-vues et recherche d'objets 3D. (Multi-view indexing and 3D objects retrieval)*, Thèse doctorat, Télécom ParisTech, France, 2010.
- [94] A. NATSEV, R. RASTOGI ET K. SHIM, *WALRUS : A similarity retrieval algorithm for image databases*, IEEE Trans. Knowl. Data Eng., 16 (2004), p. 301–316.
- [95] V. NOZICK, *Méthodes de Rendu à Base de Vidéos et Applications à la Réalité Virtuelle*, Thèse doctorat, 2006.
- [96] R. OHBUCHI, T. OTAGIRI, M. IBATO ET T. TAKEI, *Shape-similarity search of three-dimensional models using parameterized statistics*, in 10th Pacific Conference on Computer Graphics and Applications (PG 2002), 9-11 October 2002, Beijing, China, 2002, p. 265–275.
- [97] OLIVEIRA ET M. , *Image-based modeling and rendering techniques : A survey.*, International Journal of Computer Vision, 98 (2012), p. 187–201.

- [98] R. OSADA, T. A. FUNKHOUSER, B. CHAZELLE ET D. P. DOBKIN, *Matching 3d models with shape distributions*, in 2001 International Conference on Shape Modeling and Applications (SMI 2001), 7-11 May 2001, Genoa, Italy, 2001, p. 154–166.
- [99] C. PALOMO ET M. GATTASS, *An efficient algorithm for depth image rendering*, in Proceedings of the 9th International Conference on Virtual Reality Continuum and its Applications in Industry, VRCAI 2010, Seoul, Republic of Korea, December 12-13, 2010, 2010, p. 271–276.
- [100] E. PAQUET ET M. RIOUX, *Nefertiti : A query by content software for three-dimensional models databases management*, in International Conference on Recent Advances in 3-D Digital Imaging and Modeling (3DIM '97), May 12-15, 1997, Ottawa, Ontario, Canada, 1997, p. 345–352.
- [101] E. PAQUET, M. RIOUX, A. MURCHING ET T. N. ET A. TABATABAI, *Description of shape information for 2-d and 3-d objects*, Signal Processing : Image Communication, (2000).
- [102] G. PASQUALOTTO, PIETROZANUTTIGH ET G. M. CORTELAZZO, *Combining color and shape descriptors for 3d model retrieval*, SignalProcessing : Image Communication, 28 (2013), p 608–623.
- [103] G. PASSALIS ET T. T. ET IOANNIS A. KAKADIARIS, *A novel depth buffer-based shape descriptor for three-dimensional object retrieval*, Visual Comput, (2007).
- [104] R. W. PICARD, T. KABIR ET F. LIU, *Real-time recognition with the entire brodatz texture database*, in Conference on Computer Vision and Pattern Recognition, CVPR 1993, 15-17 June, 1993, New York, NY, USA, 1993, p. 638–639.
- [105] J. PODOLAK, P. SHILANE, A. GOLOVINSKIY ET S. R. ET THOMAS FUNKHOUSER, *A planar-reflective symmetry transform for 3d shapes*, ACM Trans. Graph, (2006).
- [106] B. G. PRASAD, K. K. BISWAS ET S. K. GUPTA, *Region-based image retrieval using integrated color, shape, and location index*, vol. 94, 2004, p. 193–233.
- [107] J. PU, Y. LIU, G. XIN, H. ZHA, W. LIU ET Y. UEHARA, *3d model retrieval based on 2d slice similarity measurements*, (2004), p. 95–101.
- [108] P. RAMANATHAN ET B. GIROD, *Rate-distortion analysis for light field coding and streaming*, Signal Processing : Image Communication, 21 (2006), p 462–475.
- [109] A. RAO, R. SRIHARI ET Z. ZHANG, *Geometric histogram : A distribution of geometric configuration of color subsets*, in in Proc. SPIE Conf. Electronic Imaging 2000, vol. 3964-09, San Jose, CA, 2000, p. 91–101.

- [110] J. RICARD, *Indexation et recherche d'objets 3D à partir de requêtes 2D et 3D*. déc. 2005.
- [111] O. ROBERT, F. THOMAS, C. BERNARD ET D. DAVID, *Shape distributions*, ACM Trans. Graph., 21 (2002), p. 807–832.
- [112] Y. RUBNER, C. TOMASI ET L. J. GUIBAS, *A metric for distributions with applications to image databases*, in ICCV, 1998, p. 59–66.
- [113] S. SAKJI-NSIBI ET A. BENZAZZA-BENYAHIA, *Region-based image retrieval using a joint scalable bayesian segmentation and feature extraction*, (2016), p. 1272–1276.
- [114] G. SALTON ET C. BUCKLEY, *Improving retrieval performance by relevance feedback*, JASIS, 41 (1990), p. 288–297.
- [115] D. SAUPE ET D. V. VRANIC, *3d model retrieval with spherical harmonics and moments*, in Pattern Recognition, 23rd DAGM-Symposium, Munich, Germany, September 12-14, 2001, Proceedings, 2001, p. 392–397.
- [116] Y.-T. SHEN, D.-Y. CHEN, X.-P. TIAN ET M. OUHYOUNG, *On visual similarity based 3d model retrieval*, Computer Graphics Forum, 22 (2003), p. 223–232.
- [117] H. SHI-SHENG, S. ARIEL, S. CHAO-HUI, Z. HAO, S. ALLA, H. SHI-MIN ET C.-O. DANIEL, *Qualitative organization of collections of shapes via quartet analysis*, ACM Trans. Graph., 32 (2013), p. 71 :1–71 :10.
- [118] H. SHUM ET L. HE, *Rendering with concentric mosaics*, in Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1999, Los Angeles, CA, USA, August 8-13, 1999, 1999, p. 299–306.
- [119] H. SHUM ET S. B. KANG, *Review of image-based rendering techniques*, in Visual Communications and Image Processing 2000, Perth, Australia, June 20, 2000, 2000, p. 2–13.
- [120] H.-Y. SHUM, S.-C. CHAN ET S. B. KANG, *Image-Based Rendering*, 1997.
- [121] H.-Y. SHUM, S. B. KANG ET S.-C. CHAN, *Survey of image-based representations and compression techniques*, IEEE Transaction On Circuits And Systems For Video Technology, 13 (2003), p 1020—1037.
- [122] A. W. M. SMEULDERS, M. WORRING ET S. SANTINI, *Content based image retrieval at the end of the early years*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22 (2000), p 1349–1380.
- [123] M. SUBRAHMANYAM, R. P. MAHESHWARI ET R. BALASUBRAMANIAN, *Directional local extrema patterns : a new descriptor for content based image retrieval*, IJMIR, 1 (2012), p. 191–203.

- [124] M. SUBRAHMANYAM ET Q. M. J. WU, *Expert content-based image retrieval system using robust local patterns*, J. Visual Communication and Image Representation, 25 (2014), p. 1324–1334.
- [125] H. SUNDAR, D. SILVER, N. GAGVANI ET S. J. DICKINSON, *Skeleton based shape matching and retrieval*, in 2003 International Conference on Shape Modeling and Applications (SMI 2003), 12-16 May 2003, Seoul, Korea, 2003, p. 130–142, 290.
- [126] M. J. SWAIN ET D. H. BALLARD, *Color indexing*, International Journal of Computer Vision, 7 (1991), p. 11–32.
- [127] R. SZELISKI ET H. SHUM, *Creating full view panoramic image mosaics and environment maps*, in Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1997, Los Angeles, CA, USA, August 3-8, 1997, 1997, p. 251–258.
- [128] J. W. H. TANGELDER ET R. C. VELTKAMP, *A survey of content based 3d shape retrieval methods*, Multimedia Tools Appl., 39 (2008), p. 441–471.
- [129] L. THOMAS ET K. LINUS, *Fast and Robust Approximation of Smallest Enclosing Balls in Arbitrary Dimensions*, Computer Graphics Forum, (2013).
- [130] X. TONG ET R. M. GRAY, *Coding of multi-view images for immersive viewing*, in IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP 2000, 5-9 June, 2000, Hilton Hotel and Convention Center, Istanbul, Turkey, 2000, p. 1879–1882.
- [131] —, *Interactive view synthesis from compressed light fields*, in Proceedings of the 2001 International Conference on Image Processing, ICIP 2001, Thessaloniki, Greece, October 7-10, 2001, 2001, p. 85–88.
- [132] T. TUNG ET F. SCHMITT, *Augmented reeb graphs for content-based retrieval of 3d mesh models*, in 2004 International Conference on Shape Modeling and Applications (SMI 2004), 7-9 June 2004, Genova, Italy, 2004, p. 157–166.
- [133] —, *Indexation de modèles 3d par graphe de reeb multirésolution augmenté*, Annales des Télécommunications, 60 (2005), p. 1309–1336.
- [134] W. N. VENABLES ET B. D. RIPLEY, *Modern applied statistics*, Springer, (2002).
- [135] D. V. VRANIC, *3d Model Retrieval*, Thèse doctorat, 2004.
- [136] D. V. VRANIC ET D. S. ET J. RICHTER, *Tools for 3d-object retrieval : Karhunen-loeve transform and spherical harmonics*, IEEE 2001 Workshop Multimedia Signal Processing, (2001).
- [137] D. V. VRANIC ET D. SAUPE, *3d shape descriptor based on 3d fourier transform*, EURASIP Conference on Digital Signal Processing for Multimedia Communications and Services, (2001).

- [138] D. WANG, J. ZHANG, H. WONG ET Y. LI, *3d model retrieval based on multi-shell extended gaussian image*, in Advances in Visual Information Systems, 9th International Conference, VISUAL 2007, Shanghai, China, June 28-29, 2007 Revised Selected Papers, 2007, p. 426–437.
- [139] J. Z. WANG, J. LI ET G. WIEDERHOLD, *Simplicity : Semantics-sensitive integrated matching for picture libraries*, IEEE Trans. Pattern Anal. Mach. Intell., 23 (2001), p. 947–963.
- [140] H. WONG, B. MA, Y. SHA ET H. H. IP, *Hierarchical indexing for 3d head model retrieval based on kernel PCA*, in 9th International Conference on Information Visualisation, IV 2005, 6-8 July 2005, London, UK, 2005, p. 848–853.
- [141] L. Y-S ET R. K, *Robust principal axes determination for point-based shapes using least median of squares*, Computer-Aided Design, (2009).
- [142] D. YUVARAJ ET S. HARIHARAN, *Content-based image retrieval based on integrating region segmentation and colour histogram*, Int. Arab J. Inf. Technol., 13 (2016), p. 203–207.
- [143] F. ZERNIKE, *Diffraction theory of the cut procedure and its improved form, the phase contrast method*, Physica, 56 (1934), p 689–704.
- [144] C. ZHANG ET T. CHEN, *Efficient feature extraction for 2d/3d objects in mesh representation*, in Proceedings of the 2001 International Conference on Image Processing, ICIP 2001, Thessaloniki, Greece, October 7-10, 2001, 2001, p. 935–938.
- [145] D. ZHANG, M. M. ISLAM, G. LU ET I. J. SUMANA, *Rotation invariant curvelet features for region based image retrieval*, International Journal of Computer Vision, 98 (2012), p. 187–201.
- [146] R. ZHANG ET Z. ZHANG, *A clustering based approach to efficient image retrieval*, in 14th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2002), 4-6 November 2002, Washington, DC, USA, 2002, p. 339–346.