

UNIVERSITÉ DU QUÉBEC EN OUTAOUAIS

**MODÉLISATION 3D INSPIRÉE PAR LA PERCEPTION HUMAINE
POUR DES APPLICATIONS EN RÉALITÉ VIRTUELLE**

MÉMOIRE PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DU PROGRAMME DE MAÎTRISE EN SCIENCES ET TECHNOLOGIES DE

L'INFORMATION

PAR

MAUDE CHAGNON-FORGET

Janvier 2017

JURY D'ÉVALUATION

Président du jury : Nadia Baaziz
Membre du jury : Mohand Saïd Allili
Directeur de recherche : Ana-Maria Cretu
Codirecteur de recherche : Stéphane Bouchard

REMERCIEMENTS

Je tiens tout d'abord à remercier la directrice de ce mémoire, madame Ana-Maria Cretu pour sa confiance, ses encouragements, ses conseils, son support et ses idées. Madame Cretu a su faire preuve de beaucoup de générosité. Je la remercie également pour le temps qu'elle a toujours été prête à m'accorder, pour le partage de ses connaissances et pour tous les échanges enrichissants que nous avons eus.

J'aimerais également remercier monsieur Stéphane Bouchard, co-directeur de ce mémoire, pour son soutien, sa confiance et ses encouragements. J'aimerais aussi le remercier de m'avoir accueillie au laboratoire de cyberpsychologie où j'ai eu et ai toujours l'occasion d'approfondir mes connaissances dans le domaine de la réalité virtuelle. Je remercie d'ailleurs toute l'équipe du laboratoire.

Merci à monsieur Mohand Saïd Allili et madame Nadia Baaziz pour leur participation au jury de soutenance.

Merci à Ghazal Rouhafzay, candidate au doctorat en sciences et technologies de l'information, pour son aide.

Finalement, je tiens à remercier mes parents pour leur éternel support, leurs encouragements et l'aide qu'ils n'hésitent jamais à m'offrir.

Table des matières

JURY D'ÉVALUATION	ii
REMERCIEMENTS.....	iii
Table des figures.....	vii
Table des abréviations, symboles et sigles	xii
Résumé.....	14
Abstract.....	15
1. Introduction	16
2. État de l'art	18
2.1. Modélisation 3D.....	18
2.2. Modélisation d'objets 3D à multiples niveaux de détails	23
2.3. Les modèles computationnels de l'attention visuelle.....	24
2.3.1. L'attention ascendante	25
2.3.2. L'attention descendante	28
2.4. Détection des points d'intérêt.....	28
2.5. Conclusion de l'état de l'art	32
3. Objectifs et contributions attendues	33
4. Méthodologie.....	35
4.1. Détection de points d'intérêt selon plusieurs points de vue en utilisant l'attention visuelle	35
4.1.1. Prise d'images de points de vue multiples.....	36
4.1.2. Modèle computationnel de l'attention visuelle.....	37
4.1.3. Extraction des points d'intérêt	52
4.1.4. Détection des points d'intérêt dans la carte de saillance.....	54
4.1.5. Projection des points d'intérêt 2D sur le modèle 3D	55

4.2.	Simplification sélective d'objets 3D	57
4.3.	Évaluation de la qualité du maillage	59
4.3.1.	Évaluation via des mesures d'erreur	59
4.3.2.	Évaluation via une réalité de terrain générée par l'humain	60
4.4.	Incorporation des maillages sélectivement densifiés dans une méthode discrète à multiples niveaux de détails	61
4.5.	Incorporation des maillages sélectivement densifiés dans une méthode continue à multiples niveaux de détails	62
5.	Résultats expérimentaux.....	64
5.1.	Influence des points de vue utilisés.....	64
5.2.	Influence du voisinage	67
5.3.	Influence de la couleur de fond.....	69
5.4.	Influence de la lumière dans la scène graphique.....	70
5.5.	Influence du matériau de l'objet.....	71
5.6.	Influence de la courbure.....	72
5.7.	Influence de la symétrie	74
5.8.	Influence du contraste, de l'entropie et de différentes combinaisons de canaux	78
5.9.	Influence de la texture	79
5.10.	Comparaison de différents détecteurs de points d'intérêt	80
5.11.	Évaluation avec la réalité de terrain	87
5.12.	Compacité des modèles sélectivement simplifiés	94
5.13.	Incorporation dans une méthode discrète à multiples niveaux de détails.....	94
5.14.	Incorporation dans une méthode continue à multiples niveaux de détails	95
5.15.	Temps de calcul	95
5.16.	Généralisation de la méthode proposée pour d'autres objets 3D	96
6.	Conclusion.....	98

7. Références 100

Table des figures

Figure 1- Les trois composantes d'un modèle 3D.....	18
Figure 2- Normale de surface.	19
Figure 3- Normales de sommets.	20
Figure 4- Représentation des maillages sous forme de matrices.	21
Figure 5- (a) Ombrage de Gouraud; et (b) Ombrage plat (flat) sur une sphère.	21
Figure 6- Différentes résolutions d'un maillage polygonal.....	22
Figure 7- Niveaux de détails discrets avec préservation des régions d'intérêt pour le modèle de chien.....	23
Figure 8- Aperçu de l'influence d'une tâche dans le déploiement de l'attention visuelle. Adapté de [17].....	26
Figure 9- Attributs guidant le déploiement de l'attention : (a) couleur, (b) orientation, (c) dimension, (d) fermeture, (e) densité, (f) intensité, (g) courbure, (h) profondeur, et (i) direction de la lumière.....	27
Figure 10- Diagramme de flux de l'approche proposée.....	36
Figure 11- Exemple de différents points de vue utilisés pour la prise de photos.....	37
Figure 12- Aperçu du modèle computationnel de l'attention visuelle d'Itti.	38
Figure 13- (a) Le modèle 3D de la main et (b) la carte de saillance.....	41
Figure 14- (a) Carte de saillance sans les canaux de couleur et (b) régions d'intérêt sans les canaux de couleur; (c) carte de saillance avec canaux de couleur et (d) régions d'intérêt avec canaux de couleur.....	42
Figure 15- Diagramme de flux modifié pour l'ajout du canal de courbure.....	43
Figure 16-Points de vue pour le canal de la courbure.....	44
Figure 17- Une paire de points p_i et p_j évaluée pour la symétrie bilatérale.....	46
Figure 18 - Exemple de symétrie bilatérale : (a) image du modèle original, (b) 219 points clés trouvés, (c) axes de symétrie associés aux 62 paires réfléchives trouvées, l'intensité étant proportionnelle à la magnitude de symétrie, (d) axes de symétrie dans l'espace Hough, (e) 7 caractéristiques symétriques associées avec les axes de symétrie dominants, et (f) axes de symétrie dominants et les caractéristiques symétriques associées.	47

Figure 19-Exemple de symétrie rotationnelle. (a) image originale, (b) points clés détectés, (c) centres de rotation des paires de points symétriques, (d) centre de rotation symétrique dominant et caractéristiques symétriques associées.	48
Figure 20-Diagramme de flux modifié pour l'ajout du canal de symétrie.	49
Figure 21- Exemple d'une carte de saillance Sym pour un point de vue du modèle de la main.....	50
Figure 22- Exemple de carte de saillance du contraste pour le modèle de la main	51
Figure 23- (a) Carte de saillance et (b) régions d'intérêt à la suite de l'application d'un seuillage	52
Figure 24- Comparaison de différents niveaux de seuillage. (a) 0.1, (b) 0.25 et (c) 0.5 .	53
Figure 25- Effet de l'application d'un seuil de 0.3 sur le modèle de la main	53
Figure 26- (a) Modèle de la main initial; (b) carte de saillance de l'image correspondant à la main vue de ce point de vue; (c) régions d'intérêt identifiées.	54
Figure 27- Détection des points d'intérêt à partir des régions d'intérêt comme points centraux des régions sur la carte de saillance et projetés en 3D.	54
Figure 28- Exemple de l'intersection triangle/rayon.....	55
Figure 29- Contraction d'arêtes : l'arête en gras est contractée en un seul point. Les triangles en surbrillance sont éliminés durant la contraction.....	57
Figure 30- Points d'intérêt identifiés par 3 individus différents (a)-(c) sur le modèle de la fourmi et de l'Armadillo.	60
Figure 31- Points de vue pour le calcul de l'attention visuelle : (1) placement initial, rotation de : (2) 90° selon z, (3) 180° selon z, (4) 270° selon z, (5) 45° selon x, (6) 120° selon z et 45° selon x, (7) 240° selon z et 45° selon x, (8) -45° selon x, (9) 120° selon z et -45° selon x, (10) 240° selon z et -45° selon x, (11) 90° selon x (12) 180° selon x.....	64
Figure 32- Erreurs Metro selon le nombre de points de vue. Ici, « quatre tour » utilise les points de vue 1, 2, 3 et 4 alors que « quatre » utilise les points de vue 1, 3, 11 et 12.	65
Figure 33- (a) Le modèle vu de deux points de vue avec les points d'intérêt identifiés en rouge et les modèles à différentes résolutions obtenus avec l'approche proposée pour (b) 4000, (c) 1500 et (d) 700 faces respectivement dans le maillage simplifié.	66
Figure 34- Mesures d'erreur en fonction de la grandeur du voisinage.....	67
Figure 35- Comparaison des erreurs pour un voisinage de 3 ou un voisinage adaptatif. .	68

Figure 36- Simplification avec voisinage adaptatif pour la méthode VisAttCurvSymCon et 1500 faces pour (a) le modèle de la fille et un voisinage de 1; (b) le modèle de la main et un voisinage de 2; et (c) le modèle de la théière et un voisinage de 3.....	69
Figure 37- Différentes couleurs de fond pour le modèle de la main.	69
Figure 38- Influence de la couleur de fond.	70
Figure 39- Influence du modèle de lumière, soit (a) Gouraud ou (b) plat (Flat en Matlab).	70
Figure 40- Erreur Metro pour l'influence de la lumière, pour la méthode VisAtt et une réduction à 1500 faces.	71
Figure 41- Modèle de la fourmi avec les 3 formes de matériaux soit a) dull, b) shiny et c) metal.....	72
Figure 42- Influence du matériau sur les erreurs Metro.	72
Figure 43- Influence de l'utilisation de la courbure pour les méthodes : (a) VisAtt, (b) Curv, (c) VisAttCurvP, et (d) VisAttCurv.....	73
Figure 44- Influence de l'utilisation de la courbure dans la mesure d'erreurs.....	74
Figure 45- Influence du canal de courbure pour différentes résolutions de simplification.	74
Figure 46- Points détectés par la méthode de symétrie bilatérale. Aucun point n'est détecté dans la deuxième figure (à partir de la gauche).	75
Figure 47- Point d'intérêt détecté par la symétrie rotationnelle.	75
Figure 48- Influence de la symétrie : (a) SymBiLat, (b) SymRot, (c) 2Sym, (d) VisAttSymP et (e) VisAttSym. En haut : modèle initial avec points d'intérêt identifiés en rouge. Milieu : simplification à 3000 faces. En bas : simplification à 1500 faces.	76
Figure 49- Valeur des erreurs Metro pour l'influence de la symétrie.	77
Figure 50- Influence du nombre de faces dans le modèle simplifié pour l'ajout du canal de symétrie.....	77
Figure 51- Influence de différentes combinaisons de canaux.....	78
Figure 52- Influence de différentes combinaisons de canaux pour un nombre différent de faces.	79
Figure 53-(a) Maillage d'objet texturé et comparaison des taux d'erreur lorsque la texture est utilisée ou non; (b) moyenne des erreurs Metro; et (c) erreur perceptuelle.	80

Figure 54- Comparaison de différentes solutions avec la méthode proposée pour différentes combinaisons de canaux.	81
Figure 55- Comparaison de différentes méthodes avec la solution proposée pour un nombre de faces différent.	81
Figure 56- Erreurs perceptuelles basées sur la similarité pour un nombre de faces différent.	82
Figure 57- Distance dans le domaine laplacien pour un nombre différent de faces.	82
Figure 58- Distance dans le domaine laplacien normalisé.....	83
Figure 59- Nombre moyen de points d'intérêt retournés par différentes méthodes et par la solution proposée	83
Figure 60- Résultats d'une simplification pour : (a) un grand nombre de points d'intérêt (SDC), (b) un nombre moyen de points d'intérêt (VisAttCurvSymCon) et (c) un petit nombre de points d'intérêt (HKS)	84
Figure 61- Comparaison de différents détecteurs de points d'intérêt : (a) MS, (b) SP, (c) 3DH, (d) 3DS, (e) SDC, (f) HKS, (g) VisAtt, (h) VisAttCurv, (i) VisAttSym, (j) VisAttCon, (k) VisAttEnt, (l) VisAttAll, (m) VisAttCurvSym, (n) VisAttCurvP, (o) SymBiLat, (p) SymRot, (q) 2Sym, (r) VisAttSymP, (s) Curv et (t) VisAttCurvSymCon.....	85
Figure 62- Résultats obtenus en utilisant Cloud Compare pour différents détecteurs de points d'intérêt pour le modèle de la main simplifié à 1500 faces : (a) MS, (b) SP, (c) 3DH, (d) 3DS, (e) SDC, (f) HKS, (g) VisAtt, (h) VisAttCurv, (i) VisAttSym, (j) VisAttCon, (k) VisAttEnt, (l) VisAttAll, (m) VisAttCurvSym, (n) VisAttCurvP, (o) SymBiLat, (p) SymRot, (q) 2Sym, (r) VisAttSymP, (s) Curv et (t) VisAttCurvSymCon.....	86
Figure 63- Première colonne : tous les points d'intérêt marqués par les participants, deuxième colonne : les points d'intérêt pour $nsujets = 2$ et troisième colonne, les points d'intérêt pour $nsujets = 9$, pour : (a) Armadillo, (b) fourmi, (c) fille et (d) main	88
Figure 64- Moyenne du nombre de points d'intérêt retournés par chaque méthode.....	89
Figure 65- Comparaison des points d'intérêt pour le GT (rouge) et la méthode VisAtt (en bleu). Première colonne : tous les participants, deuxième colonne : $nsujets = 2$, troisième colonne : $nsujets = 9$, pour : (a) Armadillo, (b) fourmi, (c) fille et (d) main.....	90
Figure 66- Comparaison des points d'intérêt pour le GT (rouge) et la méthode VisAttCurvSymCon (en bleu). Première colonne : tous les participants, deuxième colonne :	

<i>nsujets</i> = 2 , troisième colonne : <i>nsujets</i> = 9, pour : (a) Armadillo, (b) fourmi, (c) fille et (d) main.	91
Figure 67- Comparaison des points d'intérêt pour le GT (rouge) et la méthode VisAttAll (en bleu). Première colonne : <i>nsujets</i> = 2 , deuxième colonne : <i>nsujets</i> = 9, pour : (a) Armadillo, (b) fourmi, (c) fille, et (d) main La troisième colonne est quoi?	92
Figure 68- Erreurs Metro pour un nombre de faces de 1500 et un voisinage de $n=3$	93
Figure 69 - Modèle de l'Armadillo réduit à 1500 faces en préservant les points d'intérêt pour la méthode (a) MS, (b) SDC, (c) VisAtt, et (d) VisAttCurvSymCon.....	94
Figure 70- Modèle discret de l'objet à différents niveaux de détails en utilisant la méthode d'identification de points d'intérêt basée sur l'attention visuelle.....	95
Figure 71- Modèle de l'objet et les erreurs en code de couleur Cloud Compare à différents niveaux de détails en utilisant l'identification des points d'intérêt basée sur l'attention visuelle humaine (méthode VisAttCurvSymCon)	96

Table des abréviations, symboles et sigles

Abréviations/symboles /sigle	Signification
2Sym	Symétrie bilatérale et rotationnelle
3DH	3D-Harris
3DS	3D-SIFT
Curv	Courbure
GT(n=2)	Réalité de terrain avec un consensus de 2 individus
GT(n=9)	Réalité de terrain avec un consensus de 9 individus
HKS	Signature thermique du noyau (Heat kernel signature)
MS	Saillance du maillage (Mesh Saliency)
SDC	Coins dépendants de l'échelle (SD-Corners)
SP	Points saillants (Salient points)
SymBiLat	Symétrie bilatérale
SymRot	Symétrie rotationnelle
VisAtt	Modèle de l'attention visuelle
VisAttAll	Modèle de l'attention visuelle en plus de tous les autres canaux (courbure, symétrie, contraste et entropie)
VisAttCon	Modèle de l'attention visuelle et de la saillance provenant du contraste
VisAttCurv	Modèle de l'attention visuelle et de la saillance provenant de la courbure
VisAttCurvSym	Modèle de l'attention visuelle et de la saillance provenant de la courbure et de la symétrie

VisAttCurvSymCon	Modèle de l'attention visuelle et de la saillance provenant de la courbure, de la symétrie et du contraste
VisAttEnt	Modèle de l'attention visuelle en plus de la saillance provenant de l'entropie
VisAttCurvP	Modèle de l'attention visuelle en plus des points extraits directement de la courbure
VisAttSym	Modèle de l'attention visuelle et de la saillance provenant de la symétrie
VisAttSymP	Modèle de l'attention visuelle en plus des points extraits directement de la symétrie

Résumé

Il a été préalablement démontré que les mécanismes de l'attention visuelle humaine peuvent contribuer à l'amélioration des systèmes visuels computationnels et à l'amélioration du rendu des objets 3D. Ce travail a comme objectif principal de contribuer à l'avancement des techniques de modélisation 3D pour qu'elles soient mieux adaptées aux capacités visuelles de l'humain. L'approche proposée exploite les caractéristiques de l'attention visuelle humaine et les incorpore dans un processus de modélisation 3D à multiples niveaux de détails pour assurer une meilleure qualité des modèles simplifiés, surtout à basse résolution. Dans ce contexte, ce travail propose un cadre de modélisation 3D employant les caractéristiques de l'attention visuelle humaine afin d'obtenir des modèles compacts mieux adaptés aux capacités visuelles humaines.

Bien que de nombreuses publications mentionnent l'importance de tenir compte des mécanismes de l'attention visuelle humaine dans la détection de régions ou de points d'intérêt, très peu s'inspirent en effet d'un modèle complet de l'attention visuelle pour créer leurs solutions. Un modèle computationnel amélioré de l'attention visuelle avec des canaux de saillance supplémentaires comme la courbure, la symétrie, le contraste et l'entropie est initialement employé pour détecter les points d'intérêt à la surface d'objets 3D. L'impact de l'ajout de ces canaux supplémentaires ainsi que l'influence de différents paramètres, tels que la lumière ou le matériau des modèles, sont évalués expérimentalement. Les régions identifiées comme saillantes par le modèle de l'attention visuelle sont préservées dans un modèle sélectivement simplifié obtenu en utilisant une version adaptée de l'algorithme QSlim. Les modèles résultants sont caractérisés par une densité de points plus élevée dans les régions saillantes, assurant par le fait même une meilleure qualité perceptuelle tout en offrant une représentation du modèle plus compacte et moins complexe. La qualité des modèles obtenus est comparée avec d'autres détecteurs de points d'intérêt intégrés dans le même algorithme de simplification. La solution proposée résulte en général en des modèles de meilleure qualité, principalement à de très basses résolutions. En tant qu'exemple d'application, les modèles sélectivement densifiés sont inclus dans un cadre de modélisation à multiples niveaux de détails dans lequel un réseau de neurones détermine la taille et la résolution appropriées d'un objet.

Abstract

It has been previously shown that the mechanisms of human visual attention can contribute to the improvement of computational visual systems and to the improvement of 3D object rendering. The main objective of this thesis is to contribute to the advancement of 3D modeling techniques in order to make them more adapted to human visual capabilities. The proposed approach exploits human visual attention characteristics and incorporates them in a multiple level-of-detail 3D object modeling scheme in order to ensure a better quality of simplified models at various resolutions.

Although many publications mention the importance of considering the mechanisms of human visual attention in the detection of regions or points of interest, very few are using inspiration from a complete model of visual attention to create their solutions. An enhanced computational visual attention model with additional saliency channels, such as curvature, symmetry, contrast and entropy, is initially employed to detect points of interest over the surface of a 3D object. The impact of the use of these supplementary channels is experimentally evaluated. The regions identified as salient by the visual attention model are preserved in a selectively-simplified model obtained using an adapted version of the QSlim algorithm. The resulting model is characterized by a higher density of points in the salient regions, therefore ensuring a higher perceived quality, while at the same time ensuring a less complex and more compact representation for the object. The quality of the resulting models is compared with the performance of other interest point detectors incorporated in a similar manner in the simplification algorithm. The proposed solution results overall in higher quality models, especially at lower resolutions. As an example of application, the selectively-densified models are incorporated in a continuous multiple level of detail modeling framework, in which an original neural-network solution selects the appropriate size and resolution of an object.

1. Introduction

De nos jours, les applications graphiques ont les capacités de créer des modélisations 3D très détaillées et réalistes. Par contre, cette augmentation de la complexité (et de la taille) des modèles pose un défi sur le plan du traitement et de la manipulation de ces modèles. Ces défis sont encore plus flagrants lorsqu'il s'agit d'applications interactives où les utilisateurs s'attendent à obtenir les résultats de leurs actions en temps réel. Pour offrir un équilibre entre ces deux exigences, une approche propose de contrôler le niveau de détail des modèles 3D. Plusieurs méthodes existent pour effectuer ce contrôle. Entre autres, il est possible d'ajuster le niveau de détail des modèles en trois dimensions de manière adaptative et automatique, et ce, en utilisant les points d'intérêt identifiés sur la surface des objets 3D.

Différents algorithmes existent déjà pour l'identification des points d'intérêt. Les points d'intérêt sont ceux qui se distinguent par leur position; ils sont présents et stables dans toutes les instances d'un objet ou d'une catégorie d'objets [1]. Ce projet s'intéresse à l'identification des points d'intérêt, en se basant sur les théories de l'attention visuelle humaine.

L'être humain démontre des performances significativement supérieures, pour l'extraction et l'interprétation de l'information visuelle, à tout modèle artificiel de vision actuellement connu. Par conséquent, l'utilisation des connaissances biologiques et psychologiques dérivées des mécanismes de vision humains peut contribuer à l'amélioration des systèmes de vision computationnels [2] ainsi qu'à l'amélioration des systèmes de modélisation d'objets en 3D [3].

Le rôle de l'attention représente un intérêt particulier. Les modèles computationnels d'attention visuelle ont permis d'améliorer significativement la vitesse de traitement d'une scène visuelle et la reconnaissance d'objets [4] en tenant compte uniquement des régions d'intérêt et en distribuant les ressources là où elles sont nécessaires. Il a été démontré que les systèmes d'attention sont particulièrement bien adaptés pour détecter les caractéristiques discriminantes et que la répétabilité des régions saillantes est plus élevée que la répétabilité des régions non saillantes obtenues par des descripteurs de caractéristiques classiques tels que les coins ou les points SIFT [5, 6]. De tels modèles d'attention représentent donc une avenue de recherche prometteuse, non seulement afin de

limiter le traitement à certaines régions d'une image, mais aussi pour identifier des caractéristiques discriminantes. Ils peuvent être utilisés ultérieurement comme base pour la classification et la reconnaissance, deux aspects importants de la compréhension de scènes visuelles complexes, ou pour la simplification sélective d'objets, comme c'est le cas dans ce travail.

La contribution la plus importante de ce travail est la conception et validation d'une nouvelle solution utilisant les caractéristiques dérivées de l'attention visuelle pour contrôler automatiquement le niveau de détail d'un objet dans le contexte des applications de réalité virtuelle. Les principaux points qui seront achevés incluent l'adaptation d'un modèle d'attention visuelle ascendante 2D pour la détection des points d'intérêt sur la surface d'objets 3D, une étude expérimentale sur l'impact de l'incorporation d'autres caractéristiques saillantes inspirées de la perception humaine, telles que la courbure, la symétrie, le contraste et l'entropie dans le modèle d'attention visuelle ainsi que l'incorporation des points d'intérêt identifiés dans la carte de saillance dans un algorithme de simplification pour assurer que les traits caractéristiques d'un objet restent visibles même à la plus basse résolution. De nouvelles méthodes pour le choix du voisinage qui affecte la simplification, pour l'incorporation des maillages simplifiés dans une méthode discrète à plusieurs niveaux de détails et pour l'incorporation des maillages simplifiés dans une méthode continue à plusieurs niveaux de détails seront aussi proposées. Finalement, ce mémoire présente une démonstration et une validation expérimentale de l'importance de considérer les capacités visuelles humaines dans la modélisation 3D, ayant pour conséquence une perception améliorée des objets modélisés.

Ce travail se présente de la façon suivante : le chapitre 2 présentera une revue de l'état de l'art, et donc des travaux et des connaissances dans le domaine. On discutera des objectifs généraux et spécifiques du projet ainsi que les contributions attendues dans le chapitre 3. La méthodologie sera présentée dans le chapitre 4 et les résultats expérimentaux, dans le chapitre 5. Finalement le chapitre 6 exposera la conclusion et les contributions.

2. État de l'art

Ce chapitre vise à présenter les principaux travaux de recherche en lien avec le sujet de ce mémoire. Précisément, nous donnerons une définition de la modélisation 3D à la section 2.1; nous présenterons les travaux dans le domaine de la modélisation d'objets 3D à multiples niveaux de détails dans la section 2.2 ainsi que les travaux sur l'attention visuelle humaine et ses modèles computationnels dans la section 2.3; finalement, il sera question des travaux sur la détection de points d'intérêt dans la section 2.4.

2.1. Modélisation 3D

Lorsqu'on parle d'un modèle 3D, il s'agit de la représentation d'un objet du monde réel réalisée par ordinateur, et la modélisation représente le processus de création des composantes de l'environnement virtuel [7]. La modélisation 3D est utilisée dans plusieurs domaines, tels que la réalité virtuelle, les jeux vidéo, l'impression 3D et les dessins animés.

Les modèles 3D sont construits de trois composantes, tel qu'illustré à la Figure 1 : (1) les sommets, qui n'ont ni longueur, ni largeur, ni hauteur, soit aucune dimension; (2) les arêtes, qui ont une seule dimension, soit la longueur; et (3) les polygones, qui ont deux dimensions, soit la longueur et la largeur.

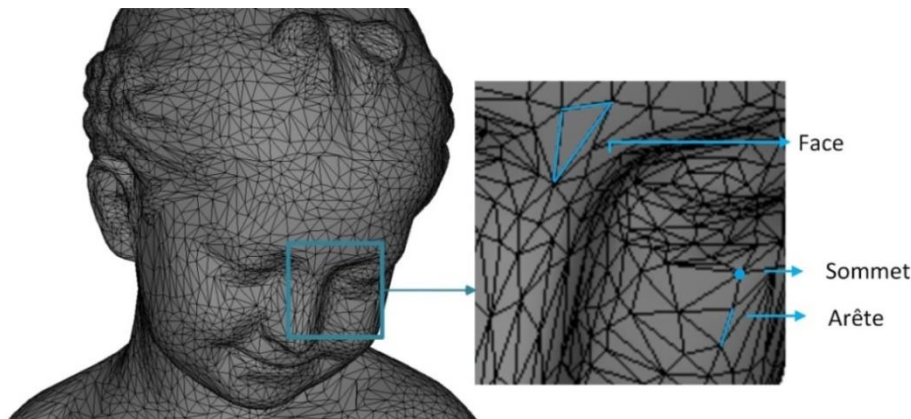


Figure 1- Les trois composantes d'un modèle 3D.

Finalement, les modèles 3D sont composés de ces trois dimensions (longueur, largeur, hauteur) [8]. Les sommets sont des points de l'espace à trois dimensions définis par leurs coordonnées (x, y, z) . Les arêtes sont l'association de deux sommets. Les triangles (ou

faces) sont les objets les plus simples qu'il soit possible de créer; ils sont représentés par l'association de trois sommets. Les polygones formant les modèles 3D peuvent être composés de n'importe quel nombre d'arêtes et de sommets, avec un minimum de trois (triangle). Les modèles utilisés dans ce travail sont des maillages de triangles, c'est-à-dire un groupe de polygones qui forment un modèle 3D. Les triangles sont idéaux pour la modélisation 3D puisqu'ils sont simples et facilement gérables par les ordinateurs [8]. Dans un maillage, toutes les arêtes sont partagées par d'autres faces, c'est-à-dire qu'il n'y a aucun trou pour pénétrer à l'intérieur du modèle [8]. Dans les applications en temps réel telles que les jeux, les représentations en maillage polygonaux dominent. Il faut préciser qu'il ne s'agit pas d'une représentation exacte de l'objet, mais plutôt d'une approximation où les sommets du maillage sont les points d'échantillonnage d'une surface lisse [7].

Les triangles composant le maillage sont définis par leur normale, qui représente la direction vers laquelle ils sont orientés. Les normales de surface jouent un rôle clé dans les représentations graphiques par ordinateur [7]. Soit un triangle $\langle p_1, p_2, p_3 \rangle$ tel qu'illustré à la Figure 2, où v_1 représente le vecteur reliant le sommet p_1 et le sommet p_2 , et v_2 , le vecteur reliant le sommet p_1 et le sommet p_3 ; la normale du triangle est alors calculée par le produit vectoriel basé sur la règle de la main droite.

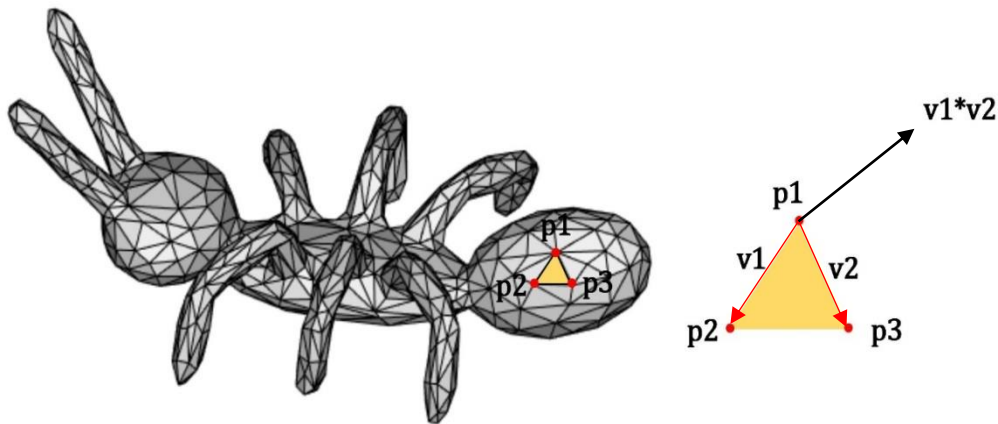


Figure 2- Normale de surface.

Puisque v_1 et v_2 sont représentés par (x_1, y_1, z_1) et (x_2, y_2, z_2) respectivement, alors $v_1 \times v_2$ est défini par $(y_1 z_2 - z_1 y_2, z_1 x_2 - x_1 z_2, x_1 y_2 - y_1 x_2)$ [7]. Il est important de noter

que les sommets p_1 , p_2 et p_3 sont ordonnés dans le sens contraire des aiguilles d’une montre. Dans le cas contraire, la normale pointerait vers la direction opposée. Dans la modélisation 3D, les normales de surfaces doivent pointer vers l’extérieur du modèle et c’est pourquoi les sommets sont considérés dans le sens contraire des aiguilles d’une montre.

Un autre aspect important réside dans les normales de sommets (Figure 3). Puisque les maillages polygonaux sont une approximation d’une surface lisse, les normales de sommets doivent estimer la normale de la surface lisse à leur position. Une normale de sommet est généralement calculée en considérant la moyenne des normales de tous les polygones partageant le sommet [7].

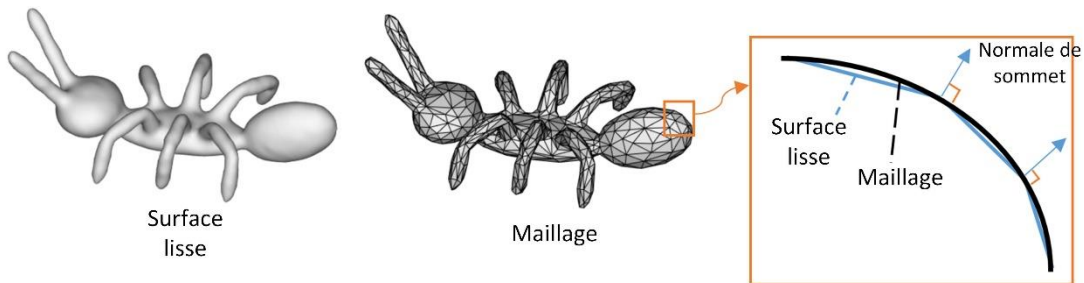


Figure 3- Normales de sommets.

Les maillages utilisés dans ce projet sont représentés dans Matlab sous forme de deux matrices, tel qu’illustré à la Figure 4. La première contient les coordonnées de tous les sommets composant le modèle et la deuxième, les indices des trois sommets composant chaque face du maillage. Cette représentation permet de sauver de l’espace puisque les coordonnées de chaque sommet ne sont emmagasinées qu’une seule fois. Dans cette représentation des triangles sous la forme de liste inversée, les sommets sont enregistrés dans le sens contraire des aiguilles d’une montre.

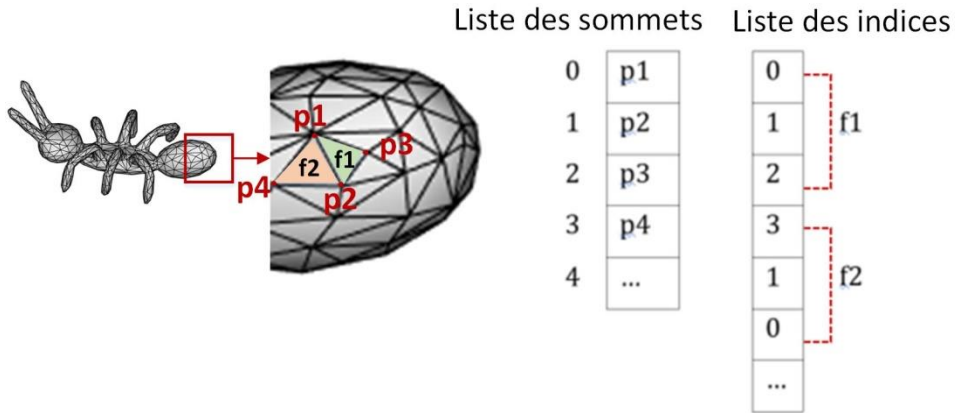


Figure 4- Représentation des maillages sous forme de matrices.

Il faut aussi mentionner que la portée de la modélisation n'est pas limitée à la construction des modèles 3D. En effet, la modélisation inclut aussi la création des textures qui sont ajoutées aux modèles 3D pour augmenter leur réalisme [7]. Certains autres aspects joueront aussi un rôle dans la qualité du maillage finale. Par exemple, le modèle de lumière utilisé influencera l'apparence de l'objet 3D. Deux modèles d'ombrage seront mentionnés plus loin dans ce travail soit *Gouraud* et plat (*Flat*). L'ombrage de Gouraud (Figure 5a) utilise les normales aux sommets de chaque polygone pour interpoler la couleur.

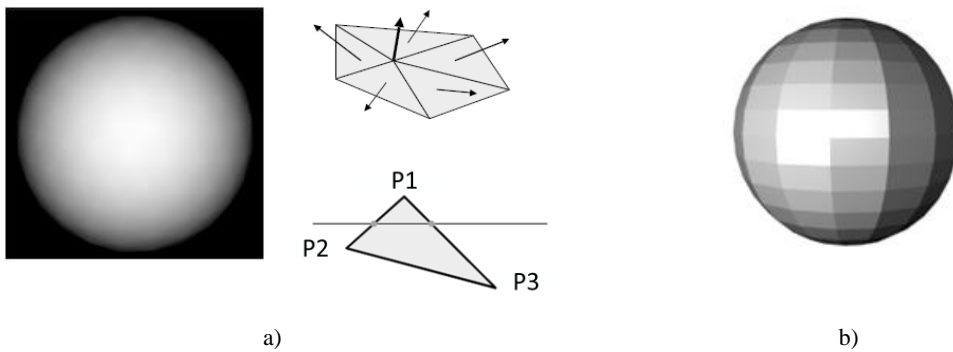


Figure 5- (a) Ombrage de Gouraud; et (b) Ombrage plat (flat) sur une sphère.

La normale à un sommet est calculée comme la moyenne des normales des polygones voisins de ce sommet. Les couleurs interpolées aux sommets sont ensuite interpolées le long des arêtes des polygones (Figure 5a). Pour compléter l'ombrage, des lignes sont

dessinées sur le modèle qui interpolent entre les intensités d'arêtes préalablement calculées. Un problème pouvant être observé avec ce type d'ombrage est que les couleurs sont uniformément interpolées le long des arêtes et des lignes dessinées, ce qui implique que les points d'intensité lumineuse plus élevée à l'intérieur d'un triangle ne seront pas pris en compte [7]. L'ombrage plat (*flat*) (Figure 5b) nécessite quant à lui l'information sur toutes les normales des polygones. La lumière affectée à chaque face est proportionnelle à l'angle entre la normale du polygone et la source lumineuse. La couleur est donc uniforme sur une même face. Cependant, des discontinuités d'ombrage sont créées aux intersections entre chacun des polygones.

Finalement, lorsqu'une surface est estimée par un maillage polygonal, différentes résolutions peuvent être considérées, tel que montré à la Figure 6.

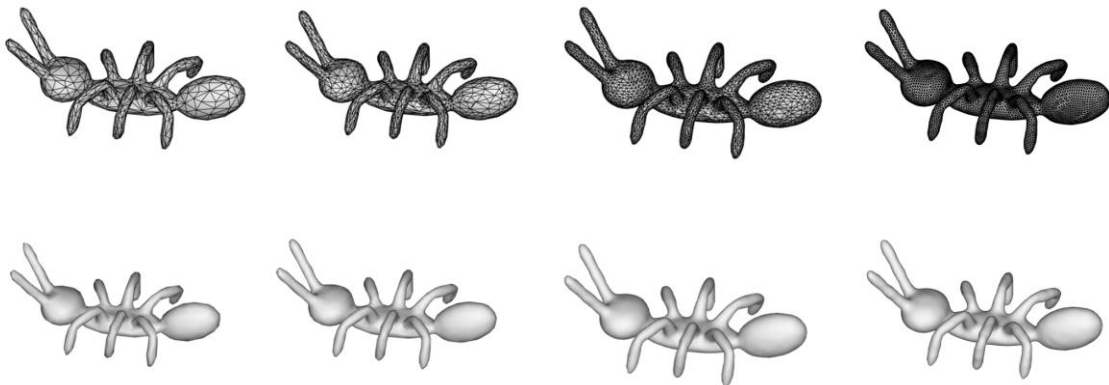


Figure 6- Différentes résolutions d'un maillage polygonal.

Il existe alors un compromis entre la précision et l'efficacité. Lorsque la résolution augmente, le maillage se rapproche de la surface originale mais le temps nécessaire pour traiter le maillage augmente. Le rendu d'un maillage de faible résolution se fait rapidement, mais la nature polygonale du modèle est facilement révélée. D'un autre côté, le rendu d'un maillage à haute résolution est lent, mais les résultats sont en général meilleurs. La gestion de différentes résolutions d'un même maillage polygonal est un sujet de recherche important. Le processus permettant de convertir un maillage de faible résolution en un maillage de haute résolution est appelé raffinement alors que le processus inverse est appelé

simplification. Le processus de simplification employé dans le cadre de ce projet sera expliqué en détails dans la section présentant la méthodologie.

2.2. Modélisation d'objets 3D à multiples niveaux de détails

Une des approches proposées dans la littérature pour gérer les contraintes graphiques des applications en temps réel est de contrôler le niveau de détail (*level of detail*, LOD) des objets. Selon Luebke et coll. [9], il existe trois catégories majeures de méthodes pour contrôler le niveau de détail d'un objet : les méthodes discrètes, les méthodes continues et les méthodes dépendantes du point de vue. La solution la plus souvent utilisée est la *méthode discrète*, comme l'illustre la Figure 7. Avec cette approche, plusieurs copies d'un même objet sont créées à des résolutions différentes, où les détails sont uniformément réduits, et ce, dans un processus hors-ligne et de prétraitement. Ces copies multiples sont employées dans les applications graphiques en fonction de la distance par rapport à l'utilisateur. De cette façon, les objets plus près ont un maillage plus dense associé à une plus haute résolution (voir le côté droit de la Figure 7) alors que les objets plus éloignés ont une résolution plus grossière puisqu'ils sont moins visibles (voir le côté gauche de la Figure 7). Cette procédure permet une vitesse de rendu plus élevée. La méthode est simple, mais la préparation des modèles est généralement longue.

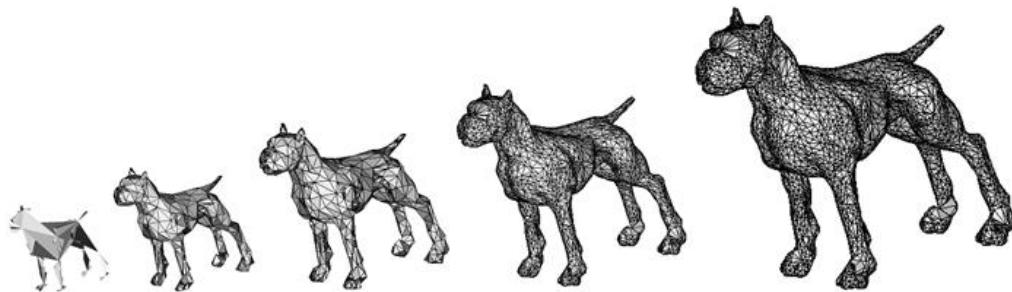


Figure 7- Niveaux de détails discrets avec préservation des régions d'intérêt pour le modèle de chien

Les *méthodes continues* utilisent quant à elles une structure de données qui encode un spectre continu de détails à partir de laquelle le niveau de détails désiré est extrait lors de l'exécution. Cette méthode d'encodage est avantageuse puisqu'elle offre un chargement interrompu ainsi que des propriétés de rendu progressif [9].

Les *méthodes dépendantes du point de vue*, étendent les méthodes continues en sélectionnant dynamiquement le niveau de détails le plus approprié à l'angle de vue actuel d'un objet. Dans ce cas, les parties de l'objet étant plus près de l'utilisateur offrent un niveau de détails plus élevé que les parties les plus éloignées. Cela résulte en une fidélité meilleure, mais aussi en une plus grande utilisation de la mémoire et un temps de traitement plus long.

Les trois méthodes présentent des résultats intéressants. Par contre, puisqu'elles simplifient la géométrie des objets uniformément, sont en général moins performantes à de très bas niveaux de détails. Certaines régions ou caractéristiques qui pourraient être plus importantes que d'autres peuvent être complètement éliminées durant le processus de simplification [9]. Afin de résoudre ce problème, certains chercheurs proposent l'intervention de l'utilisateur afin de guider la qualité désirée d'un modèle d'objet. L'utilisateur peut alors sélectionner les régions de plus basse qualité qui sont sujettes à des améliorations locales [10-13]. Seulement quelques solutions ont été proposées pour guider automatiquement la densité locale du maillage d'un modèle [14], comme nous l'envisageons dans ce projet.

2.3. Les modèles computationnels de l'attention visuelle

Le système visuel humain accomplit deux étapes de traitement visuel: une étape pré-attentive parallèle, où tout le champ visuel est traité en même temps, et une étape attentive lente et sérielle, où une région d'intérêt est sélectionnée par l'attention pour une analyse plus approfondie. C'est lors de cette deuxième étape que l'œil fera des fixations, c'est-à-dire le positionnement des objets sur la fovéa, ce qui maximise le focus mis sur l'objet. L'attention permet de segmenter la tâche d'analyse d'une scène visuelle en une série rapide de problèmes d'analyse visuelle localisés et moins exigeants sur le plan computationnel [15]. Le rôle de l'attention est de déterminer l'ordre dans lequel la scène est explorée [16], ou l'ordre de fixations, et de permettre à seulement une petite partie de la quantité énorme d'information sensorielle d'atteindre la mémoire à court terme [15].

Les études dans le domaine de la psychologie ont démontré que l'attention visuelle est divisée en deux processus : l'attention ascendante (*bottom-up*) et l'attention descendante

(*top-down*). L'attention ascendante est dirigée uniquement par la scène visuelle alors que l'attention descendante est déployée en fonction de facteurs cognitifs tels que les connaissances, les attentes et les objectifs actuels (voir *spécifications de la tâche* et *connaissances préalables* à la Figure 8). La plupart des implémentations computationnelles sont basées sur les caractéristiques ascendantes qui peuvent attirer l'attention durant des exercices de visionnement libre.

La Figure 8 illustre un aperçu de l'influence de la tâche dans le déploiement de l'attention. Soit une tâche telle que « trouver des humains dans la scène », les connaissances préalables à propos de la cible vont influencer l'extraction des caractéristiques de bas niveau en amorçant les caractéristiques désirées. Ces caractéristiques de bas niveau sont utilisées pour calculer la saillance ascendante des régions de la scène. Finalement, la carte de saillance ascendante est combinée avec la tâche et les connaissances préalables pour guider l'attention vers la cible.

2.3.1. L'attention ascendante

Afin d'identifier les caractéristiques attirant l'attention ascendante, une mesure qui a démontré sa pertinence est la saillance locale dans une image, c'est-à-dire le degré de visibilité entre la région de l'image et son entourage. En d'autres mots, la caractéristique doit être suffisamment discriminante par rapport à ce qui l'entoure afin d'occasionner le déploiement de l'attention, tel qu'illustré dans le bas de la Figure 8. Les deux personnes se distinguent suffisamment de leur entourage, entre autres par leur couleur, pour attirer l'attention visuelle. En dépit du fait que l'opinion est encore controversée en ce qui concerne les caractéristiques qui guident l'attention visuelle humaine, l'intensité, la couleur, l'orientation (voir *détecteurs de caractéristiques de bas niveau* à la Figure 8) et le mouvement sont des attributs indubitables de l'attention.

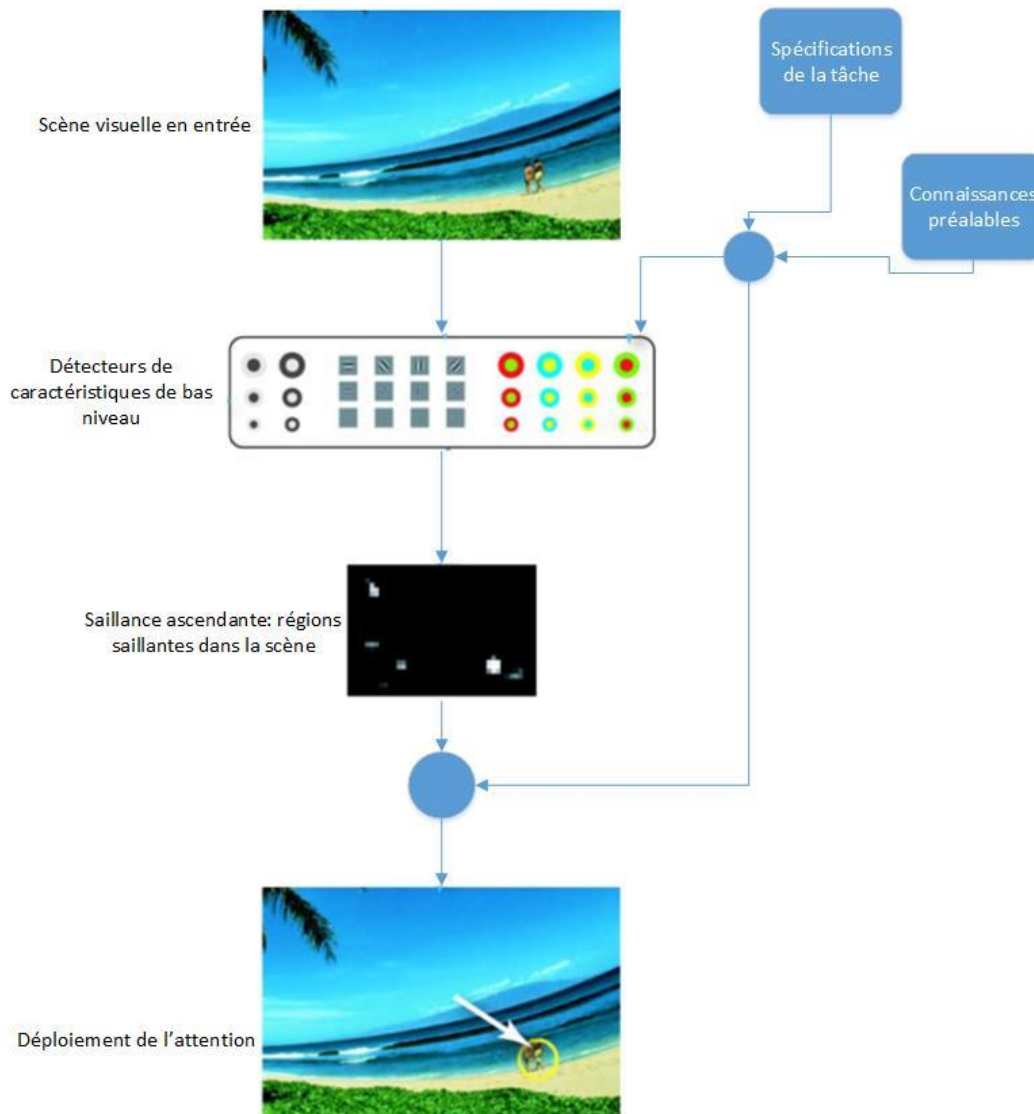


Figure 8- Aperçu de l'influence d'une tâche dans le déploiement de l'attention visuelle. Adapté de [17]

L'étude sur le déploiement de l'attention dans des tâches de recherche visuelle menée par Wolfe et Horowitz [18] a conduit à une description plus complète des attributs incluant : (1) les attributs indubitables : couleur (Figure 9a), mouvement, orientation (Figure 9b) et taille (Figure 9c); (2) les attributs probables : clignotement, polarité de la luminance, offset, profondeur (Figure 9h) et inclinaison stéréoscopique, forme, terminaison de ligne, fermeture (Figure 9d) et courbure (Figure 9g); et (3) les attributs possibles : direction de l'éclairage (Figure 9i), brillance, expansion et ratio d'aspect.

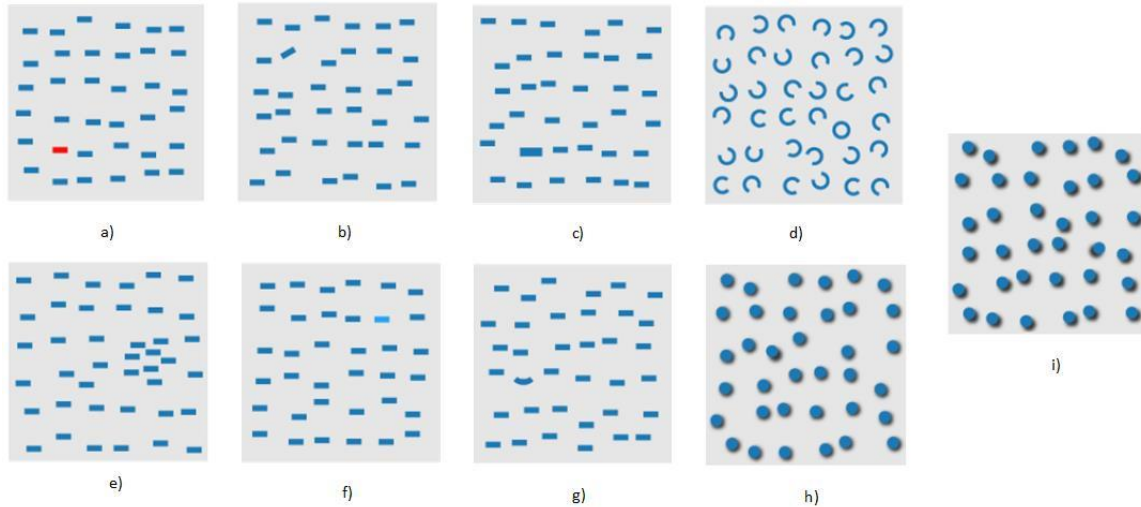


Figure 9- Attributs guidant le déploiement de l'attention : (a) couleur, (b) orientation, (c) dimension, (d) fermeture, (e) densité, (f) intensité, (g) courbure, (h) profondeur, et (i) direction de la lumière

De plus, Locher et Nodine [19] ont déterminé que lorsqu'une symétrie est présente, les fixations se concentrent le long de l'axe de symétrie, appuyant l'idée que la symétrie agit comme attribut dans le déploiement de l'attention visuelle.

L'idée derrière tous les systèmes computationnels ascendants proposés dans la littérature est de calculer plusieurs caractéristiques dérivées d'une image en couleur fournie comme élément d'entrée, chacune des caractéristiques correspondant à un attribut guidant le déploiement de l'attention, et de fusionner leur saillance dans une représentation appelée carte de saillance. Tel qu'illustré à la Figure 8 (saillance ascendante), cette carte de saillance prend la forme d'une image en niveau de gris dans laquelle l'intensité est proportionnelle à la saillance. Parmi les implémentations actuelles de modèles d'attention visuelle ascendants qui sont basés exclusivement sur la couleur, l'intensité et l'orientation, on trouve : le modèle d'Itti [15], une version plus récente du modèle proposée par l'auteur lui-même [17], [20], [21] ainsi que le modèle de Frintrop [22], pour n'en nommer que quelques-uns. La différence entre ces implémentations se trouve dans l'implémentation des canaux de couleur, d'intensité et d'orientation. Par exemple, la couleur peut être implémentée comme un code RG-BY (opposition rouge-vert/bleu-jaune), RGB (rouge, vert, bleu) ou DKL (Derrington-Krauskopf-Lennie, voir la section 4.1.2.1). Pour l'intensité, il est possible d'utiliser la différence de gaussiennes (DoG) ou les filtres de

Gabor, et pour l'orientation, les filtres de Gabor ou les pyramides orientables. D'autres attributs ayant aussi été incorporés dans le calcul de cartes de saillance incluent : le mouvement [23], le contraste [24-26], la taille [26] et l'excentricité [24, 26], la profondeur [27], la symétrie [24, 26, 28] et la texture sous forme de motifs binaires locaux [29], les ondes de Gabor [30] ou l'entropie [23], l'entropie n'étant pas incluse directement dans la carte de saillance.

2.3.2. L'attention descendante

D'autre part, l'attention descendante est orientée vers un objectif et tente de mettre en lumière les régions d'une image contenant des objets représentant un intérêt prédéfini. Tandis que la carte de saillance créée selon le système ascendant montre le niveau d'intérêt des régions dans une scène à un observateur objectif, lorsque l'observateur est à la recherche de quelque chose de spécifique, par exemple une voiture rouge dans un stationnement, les objets de couleur rouge ou en forme de voiture lui paraîtront plus intéressants même si ces objets ne sont pas les plus saillants de la scène. Ce type de comportement est modélisé en utilisant la saillance descendante, où les objets ou les caractéristiques d'intérêt se voient attribuer délibérément un poids plus élevé afin qu'ils deviennent plus saillants. Seulement quelques chercheurs ont exploré l'introduction de caractéristiques descendantes dans des systèmes computationnels d'attention : [17] ont introduit une fonction d'apprentissage d'un objet cible (et de distracteurs) comme moyen de contrôle de l'intensité des caractéristiques désirées dans la carte de saillance ascendante. L'approche de [31] pour introduire un réglage descendant est basée sur un réseau de la théorie de résonance adaptative modifiée, qui maintient la plasticité nécessaire pour apprendre de nouveaux modèles, tout en empêchant la modification des modèles déjà appris. Le mécanisme d'apprentissage de [22] combine une variété de caractéristiques ascendantes avec un poids prédéterminé dérivé des régions les plus saillantes et leur fond pour calculer l'attention descendante.

2.4. Détection des points d'intérêt

Étant donné que le projet s'intéresse plus spécialement au contexte de détection de points d'intérêt sur la surface des objets 3D, un sommaire est présenté sur les travaux existants dans la littérature à ce sujet.

Bien que de nombreuses publications mentionnent l'importance de considérer les mécanismes de l'attention visuelle humaine dans la détection de régions ou de points d'intérêt [14] [32] [33] [34] [35] [36], très peu s'inspirent en effet d'un modèle complet de l'attention visuelle pour créer leurs méthodes de détection des points d'intérêt.

Par exemple, dans [14], bien qu'on discute de l'attention visuelle humaine, seul l'attribut de la courbure est utilisé pour détecter les régions saillantes. En particulier, en se basant sur l'observation selon laquelle les changements dans la courbure sont associés à des différences dans la géométrie environnante, les auteurs calculent la saillance d'un maillage comme sa carte de courbure. La carte de courbure est une association entre chaque sommet et sa courbure moyenne. Leur méthode pour calculer la saillance utilise un mécanisme «*center-surround*» puisque celui-ci a l'avantage de pouvoir identifier les régions qui sont différentes de leur entourage. Leur mécanisme «*center-surround*» s'inspire des théories de [37] (tel que décrit à la section 0). Par contre, contrairement aux images où la couleur représente un des attributs les plus importants, les auteurs de [14] considèrent la géométrie du maillage comme l'attribut de saillance le plus important. Cette méthode de détection de la saillance sera, à partir d'ici, appelée saillance du maillage (*mesh saliency* ou MS).

Les auteurs de [35] s'intéressent quant à eux à l'inclusion des attributs descendants de l'attention visuelle pour créer la carte de saillance. Les attributs de bas niveau sont d'abord détectés en utilisant une méthode se basant sur l'intensité et les caractéristiques de couleur, alors que les caractéristiques de haut niveau sont détectées à l'aide d'une méthode utilisant l'histogramme de gradients orientés. Cependant, cette méthode s'applique seulement à des images (2D).

Dans [32], les auteurs utilisent une chaîne de Markov absorbante (*absorbing Markov chain*) pour décrire la transformation de l'attention visuelle en détection de la saillance sur un maillage.

Un courant différent de détecteurs de saillance s'inspire seulement partiellement ou pas du tout du système d'attention visuelle humain et met plutôt l'accent sur les propriétés de l'image telles que la fréquence [38, 39], les indices spatiotemporels [40] ou le contraste de région [41].

Les auteurs de [42] ont élaboré l'approche en complétant la saillance du maillage avec la théorie de Morse. Une autre amélioration de l'approche de [14] est proposée par [13]. Leur solution intègre des informations multi-échelle dans un cadre de champ aléatoire conditionnel pour imposer des contraintes de cohérence entre les points voisins.

Dans [43], le maillage d'un objet est remaillé à différents niveaux de décimation et un opérateur de différence de gaussiennes (DoG) est appliqué pour identifier les points d'intérêt à la surface d'un objet. Cette méthode est celle des points saillants (*salient points* ou SP).

Les auteurs de [44] s'intéressent au remaillage d'un objet en fonction d'un champ de saillance. Pour détecter les points d'intérêt, les auteurs prennent des photos du maillage selon différents points de vue et appliquent ensuite deux filtres (un filtre gaussien et un filtre de silhouette utilisant le squelette de l'objet) sur les images avant de projeter les points obtenus en 3D.

Les auteurs de [45] utilisent une série de 10 descripteurs locaux, soit la première et la deuxième courbure principale, la courbure gaussienne, la courbure moyenne, l'indice de forme, le log de la courbure, la distance au plan local, le volume local, l'histogramme de rotation de l'image et l'histogramme sphérique, afin de trouver les points d'intérêt à la surface de visages humains en 3D.

Dans [46], la saillance du maillage est calculée à l'aide d'une information mutuelle entre les polygones. La saillance d'un polygone est la dissemblance entre lui-même et ses voisins, mesurée comme la divergence de Jensen-Shannon.

Sipiran et Bustos [47] proposent une méthode (*3D-Harris* ou 3DH) pour calculer l'opérateur de Harris sur des maillages en 3D.

Dans [48], les auteurs utilisent une grille de voxels plutôt que le maillage. Ils proposent l'utilisation de l'approche *Scale-invariant feature transform* (SIFT) et appliquent en parallèle des filtres gaussiens. Il s'agit de la méthode *3D-SIFT* ou 3DS.

Dans [49], le maillage est d'abord déroulé sur un plan 2D à partir duquel une carte de distorsion est calculée. La carte des normales du plan 2D est ensuite filtrée à l'aide de

gaussiennes. Les points d'intérêt sont les coins, soit les points qui ont une courbure isotrope élevée et les points qui ont une courbure élevée dans au moins deux directions tangentielles. Cette méthode est appelée *SD-Corners* ou *SDC*.

La méthode *HKS* (*heat kernel signature*) [50] utilise la signature thermique du noyau pour identifier les points d'intérêt.

Les auteurs de [51] calculent la saillance du maillage en considérant à la fois le contraste local et la rareté globale, c'est-à-dire la saillance globale de chaque sommet basée sur son contraste par rapport à tous les autres sommets.

Dans [52], la saillance du maillage 3D est une combinaison de saillance locale multi-échelle et de la réponse spectrale globale, mais aucun point d'intérêt n'est identifié. L'information sur la saillance est utilisée pour guider la segmentation du maillage et l'intégration de balayage.

Leifman et coll. [53] détectent les régions d'intérêt sur une surface triangulée en cherchant les régions distinctes, et ce, en utilisant trois critères : la singularité locale d'un sommet, les extrémités de formes et l'association de patch. Une méthode est ensuite proposée afin de sélectionner les points de vue en se basant sur ces régions saillantes.

Dans [54], les caractéristiques saillantes basées sur la courbure sont obtenues en estimant localement une surface à chaque sommet (p. ex. un élément de surface analytique quadrique), en sélectionnant un point représentatif sur chaque élément de surface et en calculant analytiquement les propriétés différentielles (descripteurs) à ce point. Les caractéristiques saillantes sont obtenues par regroupement. Les ensembles de descripteurs qui ont une courbure élevée par rapport à leur voisinage et des valeurs de variance de courbure élevées sont regroupées afin d'obtenir les caractéristiques saillantes.

Les auteurs de [55] identifient les régions saillantes sur un maillage en diffusant premièrement, à l'aide d'un opérateur aléatoire de centre entourant, le champ d'index de forme (calculé en utilisant le maximum et le minimum de courbure à chaque sommet) et en l'améliorant en utilisant la théorie Retinex, un filtrage bilatéral basé sur la propagation photométrique. Les résultats sont segmentés et les points d'intérêt sont sélectionnés soit en

utilisant une approche d'échantillonnage de points [56] ou en utilisant l'entropie des valeurs de saillance [55] de chaque segment.

Yang et coll. [57] obtiennent la saillance des sommets en utilisant la distance entre un sommet et son voisinage, et l'incluent dans le calcul de la saillance de face. Les régions des objets sont ensuite classées selon leur saillance de face et celles avec les valeurs les plus élevées sont retenues comme régions d'intérêt.

Finalement, dans [1], quelques-unes des méthodes précédemment discutées pour la détection des points d'intérêt sont comparées en utilisant la vérité de terrain produite par l'humain.

2.5. Conclusion de l'état de l'art

Lorsqu'il s'agit de simplifier un maillage, l'importance de considérer la préservation de régions d'intérêt à la surface de l'objet a fait l'objet de nombreux travaux. Plusieurs se sont intéressés et s'intéressent toujours à trouver dans des images ou à la surface d'objets 3D les éléments, caractéristiques ou régions qui se distinguent. Différentes théories sont utilisées afin de déterminer ce qui sera saillant pour l'être humain, que ce soit la courbure, la couleur, l'orientation, etc. Cependant, plusieurs des méthodes actuellement disponibles dans la littérature présentent de faibles performances lors d'une simplification à très bas niveau. Les objets apparaissent donc distordus ou encore les éléments caractéristiques des objets sont perdus.

Afin d'améliorer la détection des régions d'intérêt pour la perception humaine, certains se tournent vers la théorie de l'attention visuelle humaine. Par contre, les modèles classiques de l'attention visuelle humaine n'exploitent en général que quelques caractéristiques du système visuel humain. De plus, peu de travaux se sont intéressés à l'utilisation d'un modèle d'attention visuelle humaine pour la détection de points d'intérêt pour des modèles en 3D.

3. Objectifs et contributions attendues

À la lumière des travaux sur ce sujet dans la littérature, l'objectif de ce travail est de proposer une nouvelle solution utilisant les caractéristiques dérivées de l'attention visuelle pour contrôler automatiquement le niveau de détail d'un objet dans le contexte des applications de réalité virtuelle. La méthode proposée assure que les traits caractéristiques d'un objet restent visibles même à la plus basse résolution. Pour ce faire, les régions et les points d'intérêt sont identifiés à la surface des modèles d'objets en 3D (sur un maillage) à l'aide d'un modèle computationnel adapté de l'attention visuelle humaine ascendante et sont inclus dans un cadre de modélisation d'objets à multiples niveaux de détails améliorant la perception par l'utilisateur.

Puisque, dans le contexte de ce travail, le but est de proposer un cadre de modélisation général ne dépendant pas des connaissances spécifiques à propos des objets, les influences descendantes de l'attention visuelle, qui sont guidées par les objectifs et les connaissances de l'utilisateur, ne seront pas considérées. Le modèle d'attention visuelle adapté contient un nombre restreint de canaux de caractéristiques adaptés aux exigences de l'application. Un algorithme de simplification est ensuite adapté pour permettre une simplification sélective qui conserve les points d'intérêt et leur voisinage à une plus haute-fidélité afin d'améliorer la perception et la performance lors de la simplification à très basse résolution. Finalement les modèles construits aux différents niveaux de détails sont incorporés dans une méthode de modélisation à plusieurs niveaux de détails.

Afin d'atteindre ces objectifs, plusieurs étapes intermédiaires sont réalisées. Dans une étape initiale, un modèle computationnel basé sur la couleur, l'intensité et l'orientation est testé pour le but prévu du projet de mémoire. La solution préliminaire est ensuite améliorée entre autres par : l'ajout de nouveaux canaux d'attention visuelle (tels que la courbure, la symétrie, le contraste et l'entropie) et le choix d'un voisinage adaptatif pour la simplification sélective. Puis, l'algorithme de simplification proposé est incorporé dans une méthode discrète ainsi que dans une méthode continue de modélisation à plusieurs niveaux de détail. Plusieurs tests et validations sont effectués afin de comparer les résultats obtenus

selon les différents paramètres et avec la littérature. Les différentes implémentations ainsi que les tests réalisés dans le cadre de ce projet seront faits sur la plateforme Matlab.

La contribution la plus importante visée dans ce mémoire est la conception et la validation d'une nouvelle solution utilisant les caractéristiques dérivées de l'attention visuelle pour contrôler automatiquement le niveau de détails d'un objet dans le contexte des applications de réalité virtuelle, incluant :

- l'adaptation d'un modèle d'attention visuelle ascendante 2D pour la détection des points d'intérêt sur la surface d'objets 3D;
- une étude expérimentale sur l'impact de l'incorporation d'autres caractéristiques saillantes inspirées de la perception humaine, telles que la courbure, la symétrie, le contraste et l'entropie dans le modèle de base d'attention visuelle;
- l'incorporation des points d'intérêt identifiés dans la carte de saillance dans un algorithme de simplification pour assurer que les traits caractéristiques d'un objet restent visibles même à la plus basse résolution;
- une nouvelle méthode pour le choix du voisinage qui affecte la simplification;
- une nouvelle approche pour l'incorporation des maillages simplifiés dans une méthode discrète à plusieurs niveaux de détails;
- une nouvelle approche pour l'incorporation des maillages simplifiés dans une méthode continue à plusieurs niveaux de détails;
- une démonstration et une validation expérimentale de l'importance de tenir compte des capacités visuelles humaines dans la modélisation 3D, ayant pour conséquence une perception améliorée des objets modélisés.

4. Méthodologie

Ce chapitre vise à présenter la méthodologie utilisée pour la conception de la méthode proposée dans ce mémoire. Précisément, les différentes étapes de la détection de points d'intérêt selon plusieurs points de vue seront discutées à la section 4.1. La section 4.2 présentera la méthode utilisée pour la simplification sélective d'objet alors que la section 4.3 présentera les outils et les mesures utilisées pour l'évaluation de la qualité des maillages simplifiés. Finalement, les sections 4.4 et 4.5 décriront l'incorporation des maillages sélectivement densifiés dans une méthode discrète et une méthode continue à multiples niveaux de détails.

L'approche proposée est la suivante : un modèle computationnel d'attention visuelle est initialement employé pour déterminer les régions d'intérêt. Les régions identifiées sont ensuite incorporées dans une version adaptée de l'algorithme de simplification QSlim [58] afin de créer des maillages de densités sélectives qui encodent la forme des objets 3D en préservant leurs traits caractéristiques. Les différentes simplifications des objets ainsi obtenues seront ensuite incorporées dans deux méthodes de modélisation à multiples niveaux de détails, l'une discrète, l'autre continue. Le diagramme de flux de l'approche proposée est illustré à la Figure 10.

En raison du coût de calcul associé au calcul du modèle d'attention visuelle amélioré, sur lequel des détails sont donnés dans la section des résultats expérimentaux, la détection des points d'intérêt s'effectue dans un processus hors ligne. Une fois les points d'intérêt identifiés, les modèles simplifiés avec préservation des points d'intérêt sont construits en ligne.

4.1. Détection de points d'intérêt selon plusieurs points de vue en utilisant l'attention visuelle

La section suivante présentera la méthodologie proposée afin d'identifier les points d'intérêt ainsi qu'une description des canaux ajoutés au modèle computationnel de l'attention visuelle initiale.

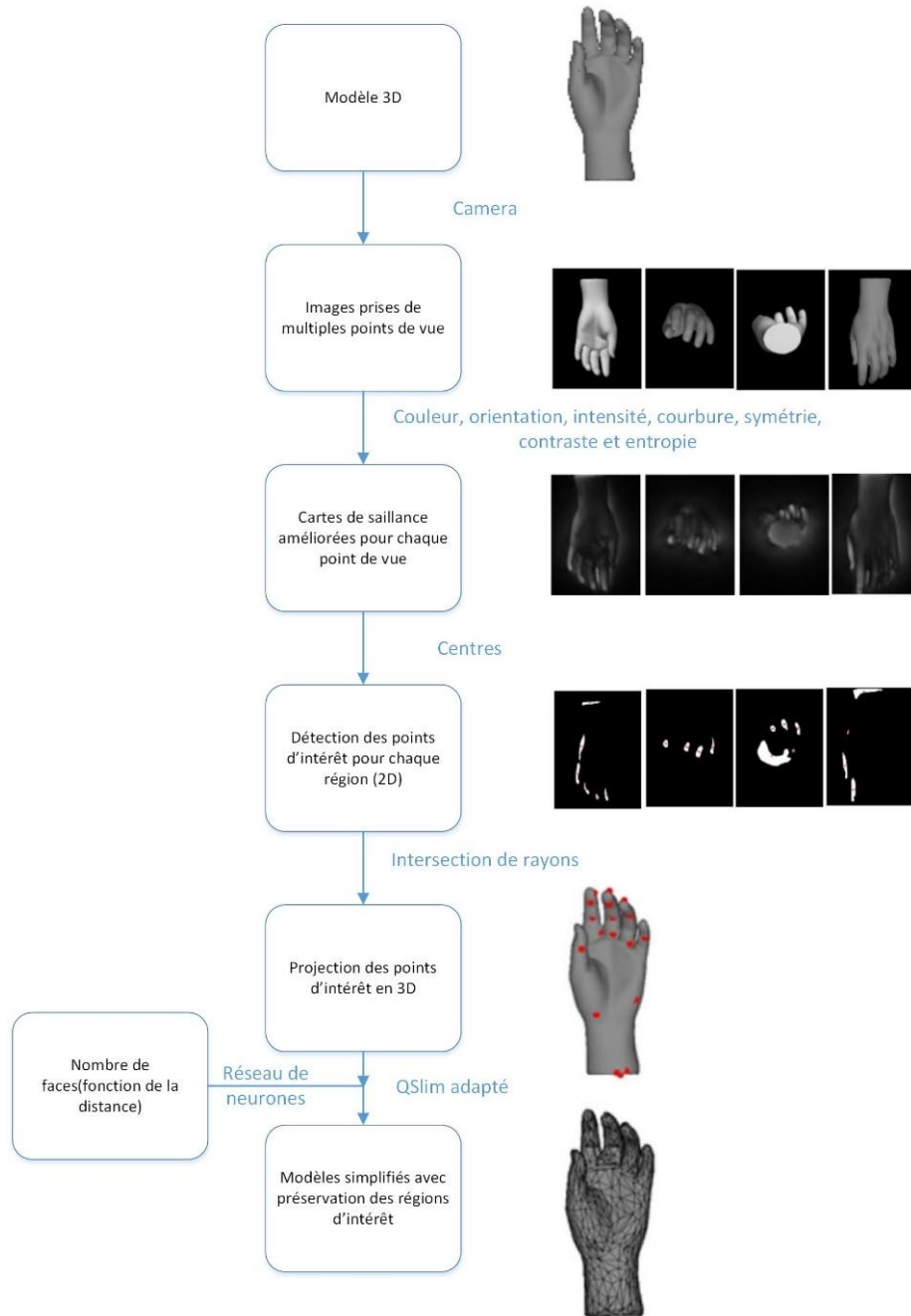


Figure 10- Diagramme de flux de l'approche proposée

4.1.1. Prise d'images de points de vue multiples

À l'instar des travaux réalisés dans [44], la première étape consiste à prendre plusieurs images d'un objet selon différents points de vue pour assurer une description relativement complète des régions d'intérêt sur la surface entière de l'objet. Cette étape est nécessaire

puisque le modèle computationnel d'attention visuelle est un modèle pour les images (et donc en 2D). Ainsi, on doit obtenir des images des modèles 3D afin d'appliquer le modèle d'attention à des objets en 3D. La Figure 11 montre un exemple des points de vue utilisés. La caméra de Matlab est utilisée pour prendre des photos de chaque objet selon plusieurs points de vue. Puisque seulement le maillage de l'objet est disponible, les objets sont rendus avec un matériau lisse, d'une couleur gris neutre. La seule source de lumière utilisée dans la scène est la *headlight*. Pour éviter que l'attention soit attirée par le contraste entre l'objet et ce qui l'entoure, un fond simple et noir est utilisé. Le nombre de points de vue ainsi que l'orientation de l'objet affectent le nombre et la position des points d'intérêt résultants, tel qu'il sera démontré dans la section des résultats (section 7).

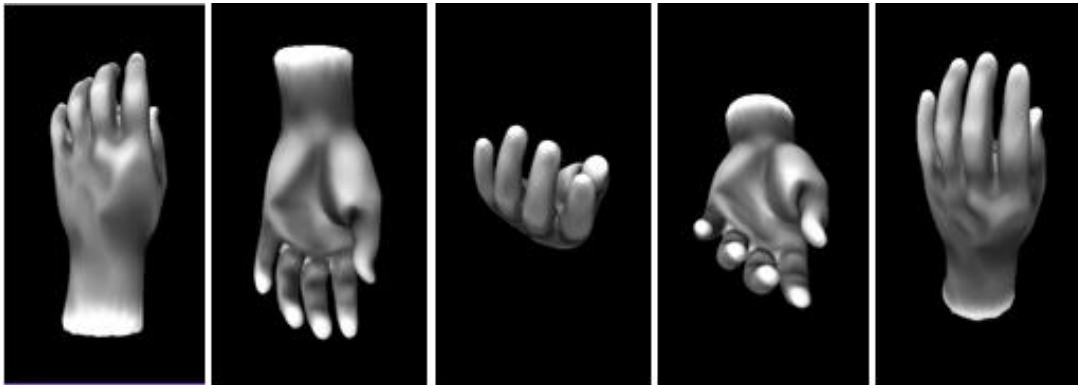


Figure 11- Exemple de différents points de vue utilisés pour la prise de photos

4.1.2. Modèle computationnel de l'attention visuelle

Une fois les images de l'objet prises, un modèle computationnel de l'attention visuelle, dont la structure générale est visible à la Figure 12, est appliqué sur chaque image des différents points de vue pour produire la carte de saillance correspondante.

Le modèle computationnel de l'attention visuelle considéré est celui d'Itti et coll. [37]. Le modèle prend en entrée une image statique en couleur. Neuf échelles spatiales sont créées en utilisant des pyramides dyadiques gaussiennes qui, progressivement, filtrent passe-bas et sous-échantillonnent l'image d'entrée, créant des réductions d'image verticales et horizontales selon des facteurs allant de 1 : 1 (échelle 0) à 1 : 256 (échelle huit) dans huit octaves. Chaque caractéristique est calculée comme un ensemble d'opérations

linéaires «*center-surround*» similaires aux champs réceptifs visuels. En effet, les neurones visuels typiques sont plus sensibles dans une petite région de l'espace visuel, soit le centre («*center*»), alors que les stimuli présents dans une région antagoniste plus large et plus faible concentrique avec le centre, soit le «*surround*», inhibent les réponses neuronales.

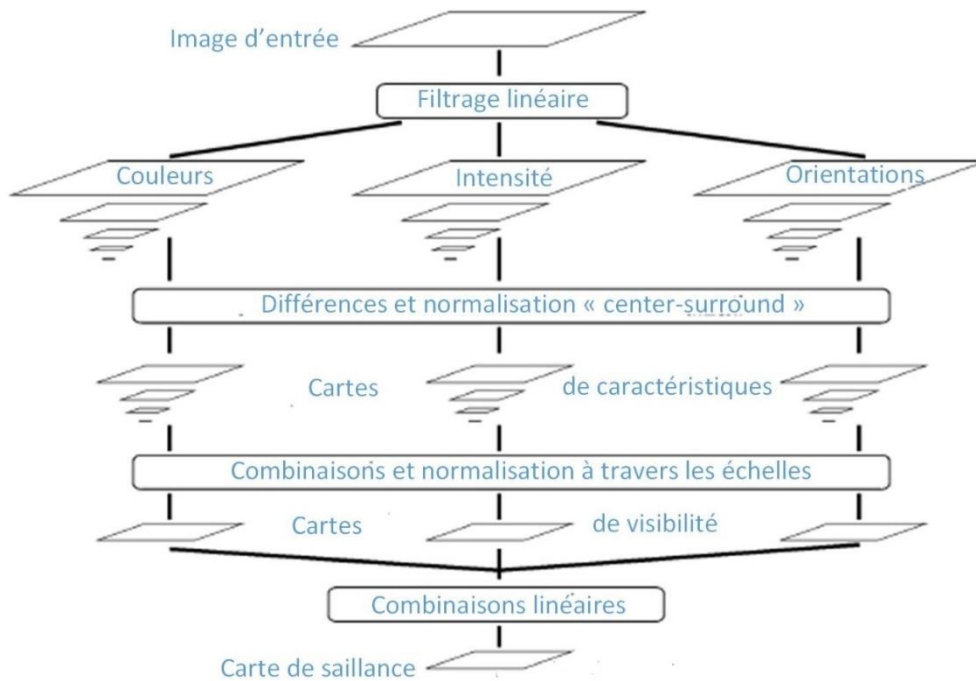


Figure 12- Aperçu du modèle computationnel de l'attention visuelle d'Itti.

Dans le modèle, le mécanisme «*center-surround*» est implémenté comme la différence entre des échelles fines et grossières. Le centre est un pixel à une échelle $c \in \{2, 3, 4\}$ et le «*surround*» est le pixel correspondant à une échelle $s = c + \delta$, où $\delta \in \{3, 4\}$. La différence entre échelles entre deux cartes (noté \ominus) est obtenue par une interpolation à l'échelle la plus fine et une soustraction point par point. En utilisant plusieurs échelles pour c et pour s , une réelle extraction des caractéristiques multi-échelles est produite puisque différents ratios sont inclus entre le centre et les régions environnantes («*surround*»).

L'extraction des caractéristiques visuelles est faite de la façon suivante : soit r , g et b étant les canaux rouge, bleu et vert respectivement de l'image d'entrée, une image d'intensité I est obtenue par : $I = (r + g + b)/3$. L'image I est ensuite utilisée pour créer une pyramide

gaussienne $I(s)$, où $s \in [0 \dots 8]$ représente l'échelle. Les canaux r , g et b sont normalisés par I afin de découpler la teinte de l'intensité. Par contre, comme les variations de teinte ne sont pas perceptibles à de très basses luminosités, la normalisation est appliquée seulement sur les endroits où I est plus grand que le 1/10 du maximum de l'image entière. Quatre canaux de couleur calibrés sont créés :

$$R = r - \frac{(g + b)}{2} \quad (1)$$

$$G = g - \frac{(r + b)}{2} \quad (2)$$

$$B = b - \frac{(r + g)}{2} \quad (3)$$

$$Y = \frac{r + g}{2} - \frac{|r - g|}{2} - b \quad (4)$$

soit les canaux rouge (R), vert (G), bleu (B) et jaune (Y), respectivement. Quatre pyramides gaussiennes $R(\sigma)$, $G(\sigma)$, $B(\sigma)$ et $Y(\sigma)$ sont créées à partir de ces canaux de couleur. La différence « *center-surround* », \ominus , produit les cartes de caractéristiques. Le premier ensemble de cartes de caractéristiques se rapporte à l'intensité du contraste et est calculé dans un ensemble de 6 cartes $J(c, s)$ où $c \in \{2, 3, 4\}$ et $s = c + \delta, \delta \in \{3, 4\}$:

$$J(c, s) = |I(c) \ominus I(s)| \quad (5)$$

Le deuxième ensemble de cartes est construit pour les canaux de couleur. Des cartes $RG(c, s)$ sont créées pour considérer les oppositions rouge/vert et vert/rouge et des cartes $BY(c, s)$ pour considérer les oppositions bleu/jaune et jaune/bleu :

$$RG(c, s) = |(R(c) - G(c)) \ominus (G(s) - R(s))| \quad (6)$$

$$BY(c, s) = |(B(c) - Y(c)) \ominus (Y(s) - B(s))| \quad (7)$$

Un troisième ensemble de cartes est créé pour l'orientation. L'information sur l'orientation locale est obtenue de I en utilisant des pyramides de Gabor orientées $O(\sigma, \theta)$ où $\sigma \in [0 \dots 8]$ représente l'échelle et $\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ est l'orientation préférée. Les cartes de

caractéristiques d'orientation, $O(c, s, \theta)$ encodent le contraste d'orientation locale entre les échelles de centre et de « *surround* »:

$$O(c, s, \theta) = |O(c, \theta) \ominus O(s, \theta)| \quad (8)$$

Au total, 42 cartes de caractéristiques sont calculées, soit 6 pour l'intensité, 12 pour la couleur et 24 pour l'orientation.

Afin de créer la carte de saillance, les auteurs de [37] proposent un opérateur de normalisation $\mathcal{N}(\cdot)$ qui promeut globalement les cartes où un petit nombre de forts pics de saillance sont présents, et supprime globalement des cartes contenant plusieurs pics similaires. La création de $\mathcal{N}(\cdot)$ se fait en 3 étapes, soit : 1) normaliser les valeurs dans les cartes dans un intervalle de $[0, M]$, où M est le maximum global de la carte, 2) déterminer la position du maximum global de la carte et calculer la moyenne \bar{m} de tous les autres maximums locaux et 3) multiplier globalement la carte par $(M - \bar{m})^2$. Seulement les maximums locaux d'activité sont considérés, de sorte que $\mathcal{N}(\cdot)$ compare les réponses associées à des endroits d'activation significatifs et ignore les régions homogènes.

Les cartes de caractéristiques sont ensuite combinées en trois cartes de visibilité : $\bar{\mathcal{J}}$, pour l'orientation, $\overline{C_{RGB}}$ pour la couleur et \bar{O} pour l'orientation et ce à l'échelle de la carte de saillance ($\sigma = 4$). Les cartes de visibilité sont obtenues par une addition entre échelles, \oplus , qui consiste en la réduction de chaque carte à l'échelle 4 et une addition point par point.

$$\bar{\mathcal{J}} = \oplus_{c=2}^4 \oplus_{s=c+3}^{c+4} \mathcal{N}(\mathcal{J}(c, s)) \quad (9)$$

$$\overline{C_{RGB}} = \oplus_{c=2}^4 \oplus_{s=c+3}^{c+4} [\mathcal{N}(RG(c, s)) + \mathcal{N}(BY(c, s))] \quad (10)$$

Pour l'orientation, quatre cartes intermédiaires sont créées en combinant les 6 cartes de caractéristiques pour un certain θ et sont ensuite combinées :

$$\bar{O} = \sum_{\theta \in \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}} \mathcal{N} \oplus_{c=2}^4 \oplus_{s=c+3}^{c+4} \mathcal{N}(O(c, s, \theta)) \quad (11)$$

Finalement, la carte de saillance est créée de la façon suivante :

$$S = \frac{1}{3} (N(\overline{J}) + N(\overline{C_{RGB}}) + N(\overline{O})) \quad (12)$$

Un exemple d'une carte de saillance créée à partir de ce modèle est montré à la Figure 13. Tel que déjà mentionné, la carte de saillance est une carte en niveau de gris où les parties les plus saillantes correspondent aux endroits les plus blancs. Ce sont donc ces parties qui sont considérées comme intéressantes du point de vue de l'attention visuelle sur l'image.

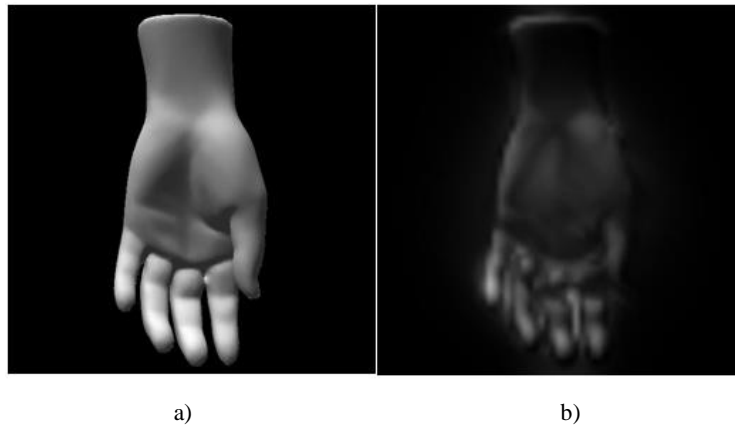


Figure 13- (a) Le modèle 3D de la main et (b) la carte de saillance

4.1.2.1. Ajout du canal de couleurs DKL

En plus du code de couleur RGB, le code de couleur Derrington-Krauskopf-Lennie (DKL) est utilisé comme canal dans le calcul de la saillance. L'espace colorimétrique DKL [59] fait référence au modèle d'opposition de couleurs dans les premiers traitements visuels [60]. La vision de la couleur débute par l'extraction de différents signaux émis par les cônes et enregistrés par trois mécanismes post-récepteurs, soit un mécanisme de luminance et deux mécanismes d'opposés chromatiques. Les axes du code de couleur DKL correspondent au contraste de couleur de trois mécanismes théoriques soit la luminance, l'opposition rouge/vert et l'opposition jaune/bleu. Cet espace met l'emphase sur la réponse au fond et aux déviations par rapport au fond ce qui représente le contraste des stimuli [60].

En ajoutant le code couleur DKL en plus des caractéristiques existantes en [59], la carte de saillance (VisAtt) est créée de la façon suivante :

$$S = \frac{1}{4} (N(\bar{J}) + N(\bar{C}_{RGB}) + N(\bar{C}_{DKL}) + N(\bar{O})) \quad (13)$$

Bien que les modèles 3D utilisés dans le cadre de ce travail soient en niveaux de gris, les canaux de couleur sont tout de même utilisés. L'objectif de ce projet est en effet de créer une méthode qui pourra être utilisée sur tout modèle 3D, indépendamment de ses propriétés de couleur ou de texture. Pour la majorité des expérimentations, seulement des objets en niveaux de gris sont utilisés, mais la méthode peut être utilisée sur des objets colorés ou texturés sans ajustement. Par exemple, la section 5.9 montre les résultats de l'utilisation de méthode sur un objet texturé.

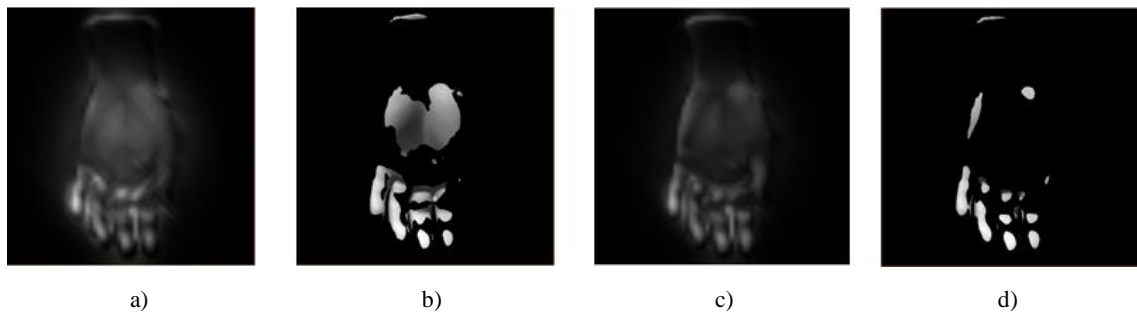


Figure 14- (a) Carte de saillance sans les canaux de couleur et (b) régions d'intérêt sans les canaux de couleur; (c) carte de saillance avec canaux de couleur et (d) régions d'intérêt avec canaux de couleur

De plus, malgré l'utilisation du niveau de gris, l'utilisation des canaux de couleur (RGB et DKL) permet d'identifier plus précisément les zones d'intérêt, tel qu'illustré à la Figure 14. Cette figure compare les régions d'intérêt détectées sans l'utilisation des canaux de couleur dans la Figure 14b avec le cas où les canaux de couleur sont employés, dans la Figure 14d. On peut observer que les régions sont plus précises à la Figure 14d.

4.1.2.2. Ajout du canal de courbure

La Figure 15 montre le diagramme de flux lorsque le canal de courbure est ajouté au modèle de l'attention visuelle.

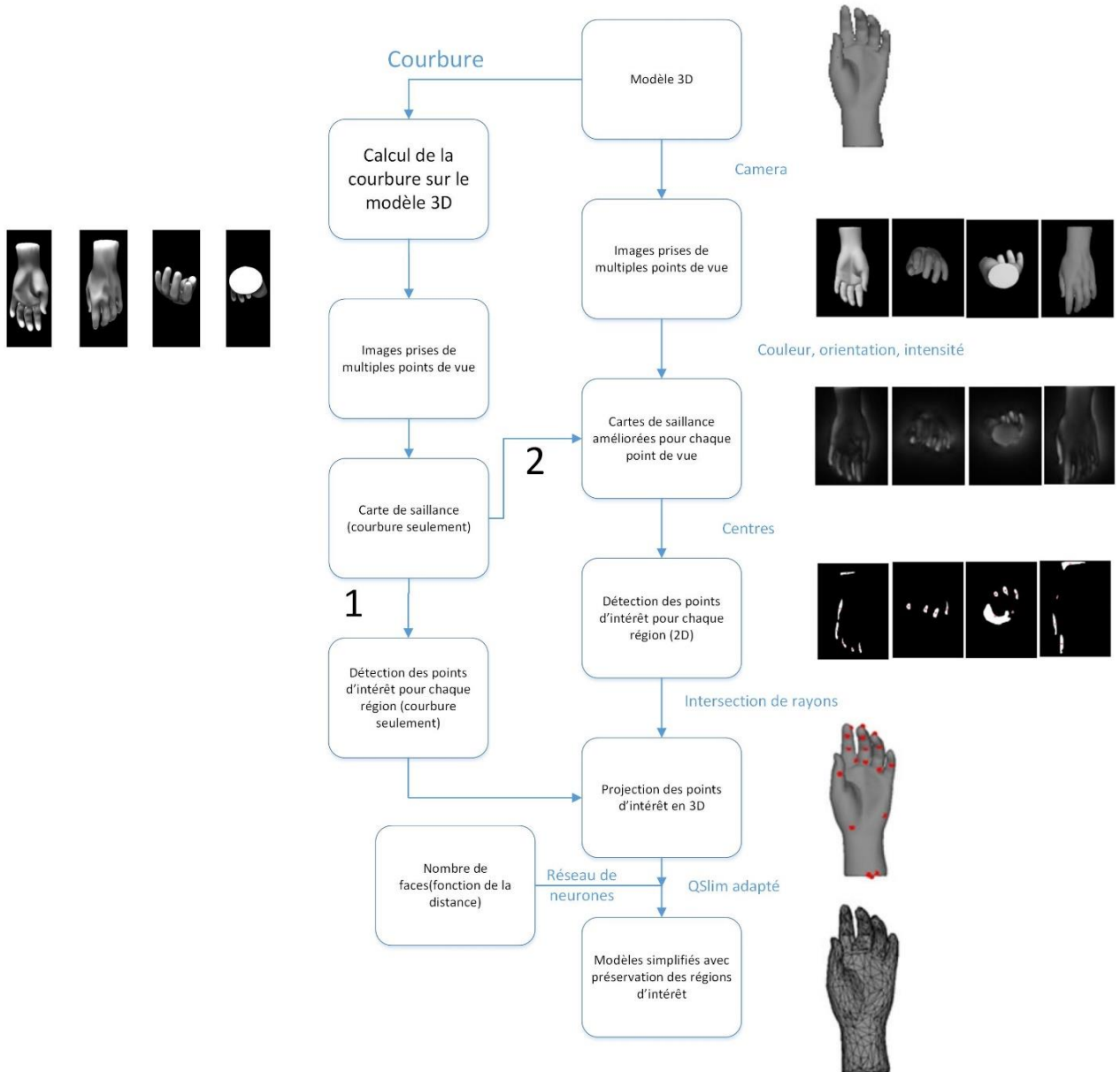


Figure 15- Diagramme de flux modifié pour l'ajout du canal de courbure

La Figure 16 montre les différents points de vue utilisés pour le calcul du canal de courbure. Les zones colorées en blanc sont celles ayant la courbure maximale.

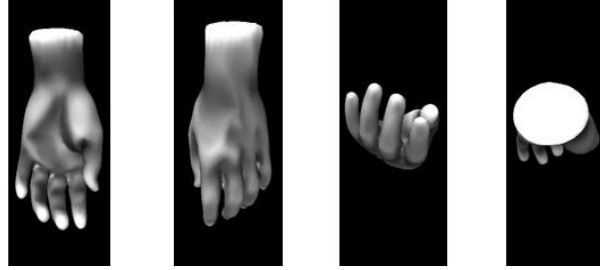


Figure 16-Points de vue pour le canal de la courbure

Afin de calculer la courbure à la surface d'un objet (en 3D), la fonction *compute_curvature* du Toolbox Graph [61] sur Matlab est utilisée. L'algorithme détaillant cette fonction se trouve dans [61]. La carte de saillance de la courbure (*Curv*) est ainsi obtenue de façon indépendante.

La carte de saillance de la courbure a été ajoutée au modèle de l'attention visuelle de deux façons différentes afin de comparer les résultats. Premièrement, les points d'intérêt sont extraits de la carte de saillance de la courbure afin de visualiser leur effet indépendamment du modèle de l'attention visuelle (*VisAttCurvP*) (identifié comme « 1 » à la Figure 15). Deuxièmement, la carte de saillance de la courbure est combinée avec les cartes créées par les canaux de couleur (RGB et DKL), d'intensité et d'orientation du modèle de l'attention visuelle (*VisAttCurv*) afin de créer la carte de saillance de l'objet qui est calculée par :

$$S = \frac{1}{5} (N(\mathcal{J}) + N(\overline{C}_{RGB}) + N(\overline{C}_{DKL}) + N(\overline{O}) + N(\overline{Curv})) \quad (14)$$

où *Curv* représente le canal de la courbure (identifié comme « 2 » à la Figure 15).

Les résultats obtenus avec l'influence de la courbure sont présentés dans la section 0.

4.1.2.3. Ajout du canal de symétrie

Les travaux de recherche dans la littérature [19] ont déterminé que lorsqu'une symétrie est présente, les fixations se concentrent le long de l'axe de symétrie. Cela appuie l'idée selon laquelle la symétrie agit comme attribut dans le déploiement de l'attention visuelle. Pour cette raison, des tests ont été effectués pour évaluer l'impact de la symétrie sur les points d'intérêt.

Afin d'évaluer l'influence de la symétrie dans la détection de points d'intérêt, le modèle proposé par les auteurs de [62] est utilisé. Ce modèle permet de calculer deux types de symétrie soit la symétrie bilatérale (miroir) et la symétrie rotationnelle.

Premièrement, un ensemble de points p_i sont identifiés en utilisant la méthode SIFT, qui distingue des points distinctifs avec une bonne répétabilité. Le vecteur point $p_i = (x_i, y_i, \phi_i, s_i)$ associé à chaque point décrit sa position (x_i, y_i) , son orientation (ϕ_i) et, optionnellement, son échelle (s_i) . Ensuite, un descripteur k_i est généré pour chaque point, encodant l'apparence locale de la caractéristique après la normalisation de son orientation (et de l'échelle s'il y a lieu). SIFT produit k_i comme étant un vecteur de 128 éléments.

Pour la *symétrie bilatérale (SymBiLat)*, un ensemble m_i de descripteurs caractéristiques de reflet est généré, où m_i décrit une version reflétée d'un élément de surface de l'image associée à la caractéristique k_i . Les descripteurs m_i sont produits en réordonnant les éléments du vecteur descripteur pour qu'ils représentent l'élément de surface original de l'image retourné selon l'axe ayant l'orientation dominante. Une association est ensuite faite entre les caractéristiques k_i et les caractéristiques miroirs m_i pour créer un ensemble (p_i, p_j) de caractéristiques potentiellement symétriques. La symétrie de chaque paire est quantifiée en fonction de la position, de l'orientation et de l'échelle relatives de p_i et p_j . Un poids de symétrie angulaire $\Phi_{ij} \in [-1, 1]$ est calculé à l'aide de la formule suivante [62]:

$$\Phi_{ij} = 1 - \cos(\phi_i + \phi_j - 2\theta_{ij}) \quad (15)$$

où les angles sont définis tel que montré dans la Figure 17.

Un poids d'échelle $\Phi_{ij} \in [-1, 1]$ quantifiant la similarité relative de l'échelle des deux vecteurs est calculé avec la formule :

$$S_{ij} = \exp\left(\frac{-|s_i - s_j|}{\sigma_s(s_i + s_j)}\right)^2 \quad (16)$$

où σ_s contrôle la quantité de variation d'échelle acceptée. Comme dans le cas des expérimentations réalisées par les auteurs de [62], $\sigma_s = 1$ est utilisé.

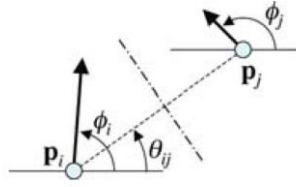


Figure 17- Une paire de points p_i et p_j évaluée pour la symétrie bilatérale

Finalement, une fonction gaussienne de poids de distance $D_{ij} \in [0,1]$ peut être utilisée pour récompenser les paires de caractéristiques plus près de l'axe de symétrie.

$$D_{ij} = \exp\left(\frac{-d^2}{2\sigma_d^2}\right) \quad (17)$$

où d est la distance séparant les deux caractéristiques. Dans ce cas-ci, une valeur $D_{ij} = 1$ est utilisée (aucune contrainte de diamètre). La magnitude de symétrie, quantifiant la quantité de symétrie que démontre une paire de vecteurs de points est définie comme suit :

$$M_{ij} = \begin{cases} \Phi_{ij} S_{ij} D_{ij} & \text{if } \Phi_{ij} > 0 \\ 0 & \text{sinon} \end{cases} \quad (18)$$

Les axes de symétrie potentiels qui se trouvent entre les deux points caractéristiques d'une paire peuvent être représentés par :

$$r_{ij} = x_c \cos \theta_{ij} + y_c \sin \theta_{ij} \quad (19)$$

où (x_c, y_c) sont les coordonnées image du point milieu de la ligne joignant p_i et p_j et où θ_{ij} est l'angle entre cette ligne et l'axe des x . Pour trouver les axes de symétrie dominants, la transformation linéaire de Hough est utilisée. Chaque paire (p_i, p_j) a un vote (r_{ij}, θ_{ij}) de poids M_{ij} dans l'espace de Hough. L'espace de Hough est brouillé avec un filtre gaussien et les maximums sont extraits et utilisés pour déterminer l'axe de symétrie dominant. Les points se trouvant dans le voisinage de ces maximums dans l'espace de Hough indiquent les paires symétriques associées avec cet axe de symétrie. Un exemple des résultats obtenus pour la symétrie bilatérale pour le modèle 3D d'Armadillo est illustré à la Figure 18.

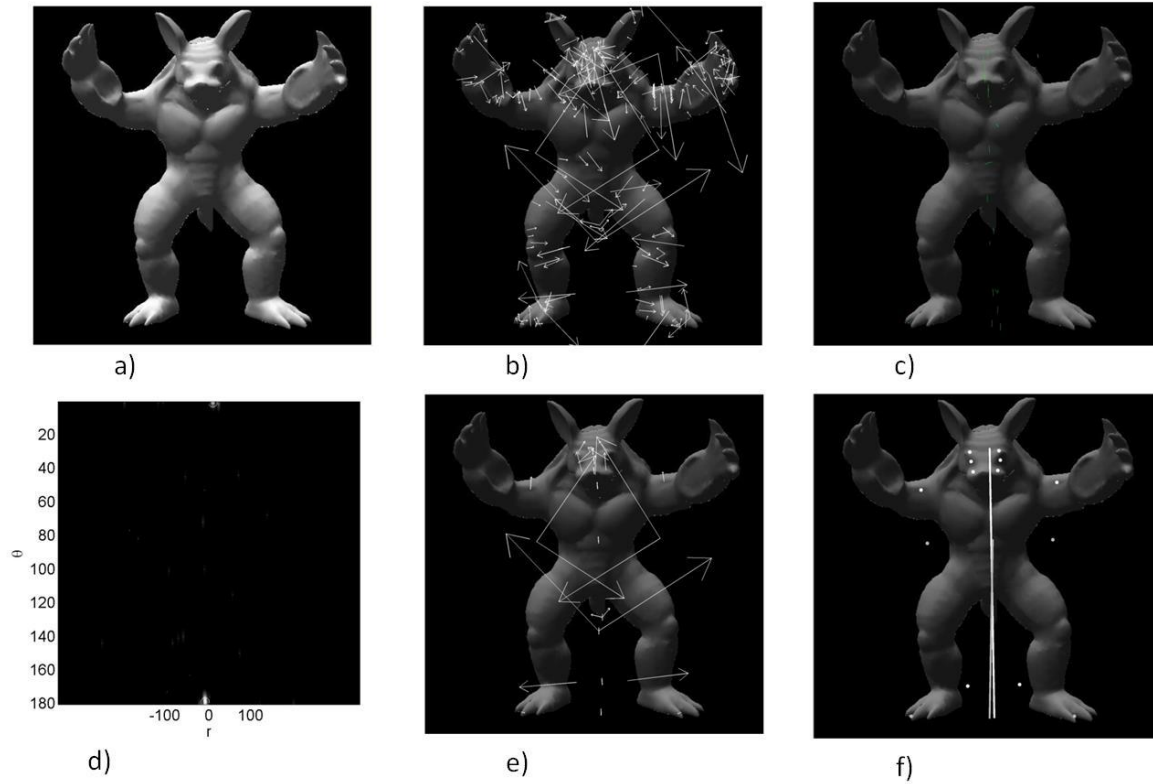


Figure 18 - Exemple de symétrie bilatérale : (a) image du modèle original, (b) 219 points clés trouvés, (c) axes de symétrie associés aux 62 paires réfléchives trouvées, l'intensité étant proportionnelle à la magnitude de symétrie, (d) axes de symétrie dans l'espace Hough, (e) 7 caractéristiques symétriques associées avec les axes de symétrie dominants, et (f) axes de symétrie dominants et les caractéristiques symétriques associées.

En ce qui concerne la symétrie rotationnelle (SymRot), il suffit d'associer les caractéristiques k_i les unes avec les autres [62]. Soit une paire de vecteurs point p_i et p_j non parallèles, il existe un point c_{ij} à une distance r de p_i et de p_j selon lequel p_i peut être tourné afin de devenir aligné avec p_j . Le centre de rotation c_{ij} est donné par :

$$c_{ij} = \begin{pmatrix} x_i \\ y_i \end{pmatrix} + \begin{pmatrix} r \cos(\beta + \gamma) \\ r \sin(\beta + \gamma) \end{pmatrix} \quad (20)$$

où x_i, y_i sont les coordonnées du point p_i et γ est l'angle formé par la ligne joignant p_i et p_j et l'axe des x .

Avec Pythagore, on obtient :

$$r^2 = \left(\frac{d}{2}\right)^2 + \left(\frac{d}{2} * \tan\beta\right)^2 \rightarrow \beta = \frac{\phi_i - \phi_j + \pi}{2} \quad (21)$$

où d est la distance entre p_i et p_j et ϕ_i l'orientation de p_i . On définit aussi:

$$\begin{aligned} \phi_i &= \gamma + \beta + \varphi \\ \phi_j &= \gamma + \pi - \beta + \varphi \end{aligned} \rightarrow \beta = \frac{\phi_i - \phi_j + \pi}{2} \quad (22)$$

Une fois que le centre de rotation est déterminé pour chaque paire de points, la magnitude de symétrie rotationnelle est calculée par :

$$R_{ij} = S_{ij}D_{ij} \quad (23)$$

Les centres de symétrie rotationnels sont déterminés en accumulant les centres c_{ij} dans une image de vote où chaque vote a un poids équivalent à sa magnitude. Un filtre gaussien est appliqué sur l'image et les maximums sont identifiés comme les centres de rotation dominants. Tous les centres de rotation près d'un maximum sont associés à ce maximum.

Dans leurs travaux, les auteurs de [62] font le pairing en construisant une matrice de similarité quantifiant la similarité entre les points caractéristiques. Ils utilisent la distance euclidienne entre les descripteurs SIFT. Un exemple des résultats obtenus pour la symétrie rotationnelle pour un modèle 3D d'une main est illustré à la Figure 19.

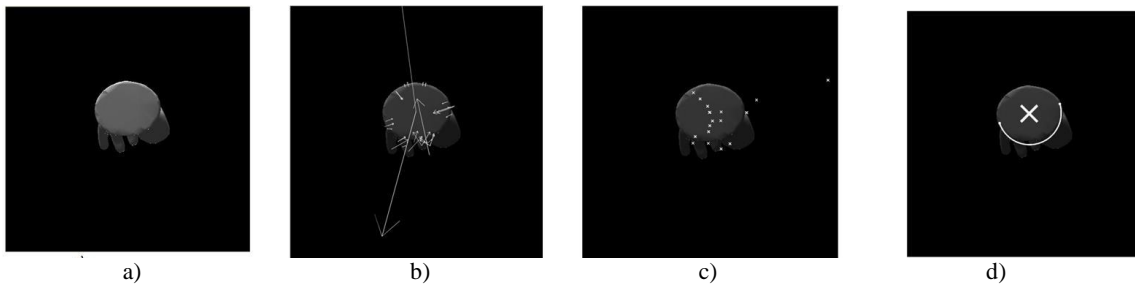


Figure 19-Exemple de symétrie rotationnelle. (a) image originale, (b) points clés détectés, (c) centres de rotation des paires de points symétriques, (d) centre de rotation symétrique dominant et caractéristiques symétriques associées.

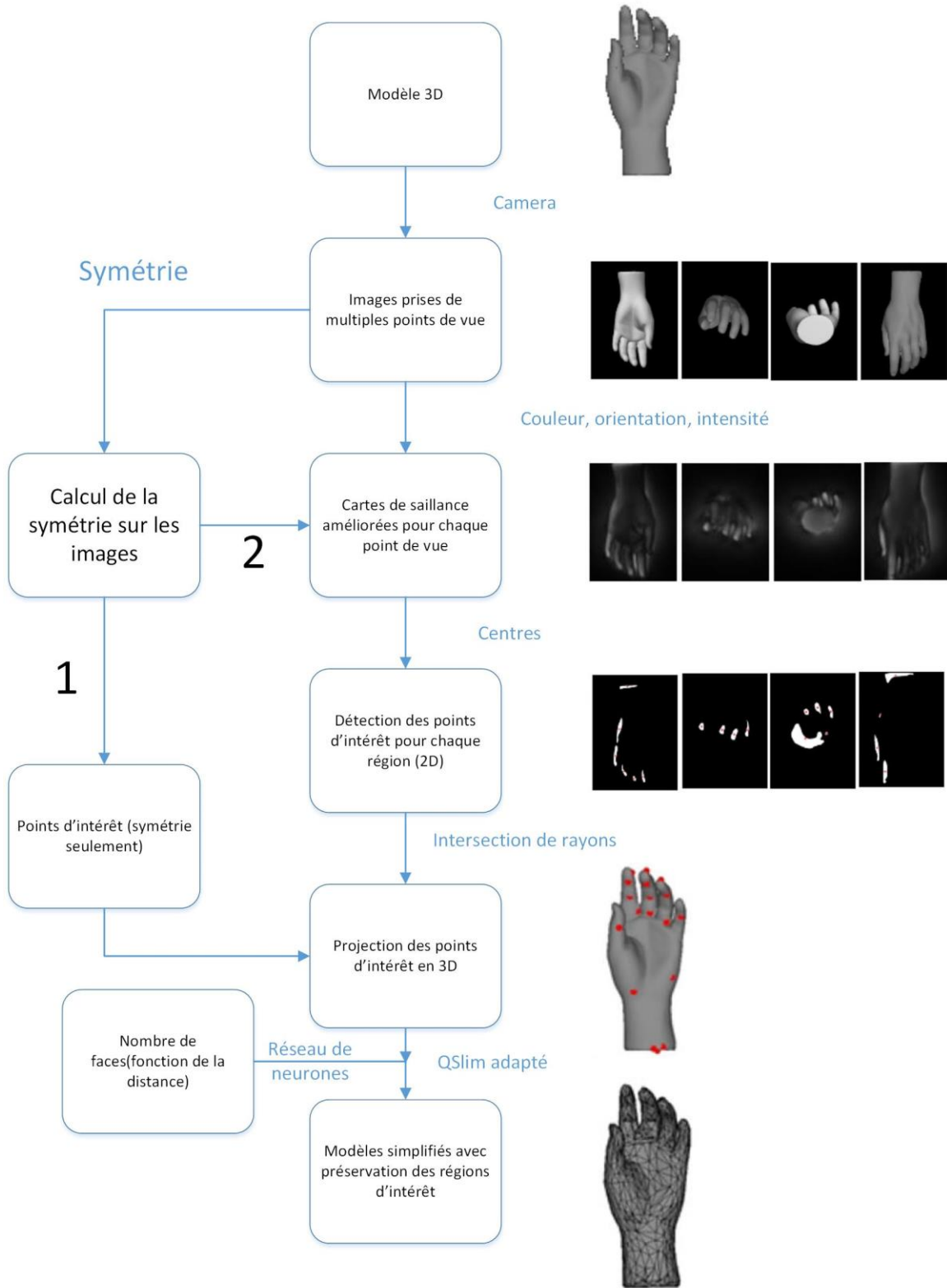


Figure 20-Diagramme de flux modifié pour l'ajout du canal de symétrie.

Afin d'inclure la symétrie dans le modèle de l'attention visuelle, deux méthodes sont proposées. Premièrement, les points d'intérêt détectés sont ajoutés directement aux points d'intérêt détectés par la méthode de l'attention visuelle seulement (VisAttSymP) (identifié comme « 1 » à la Figure 20). Dans un deuxième temps, la symétrie est incorporée sous forme de carte de saillance (identifié comme « 2 » à la Figure 20), c'est-à-dire que des cartes de saillance Sym (Figure 21) sont créées pour les différents points de vue où les points d'intérêt et leurs voisins immédiats apparaissent en blanc et le fond en noir. À l'aide d'un opérateur aléatoire de centre entourant appliqué sur les cartes Sym , la carte de symétrie (VisAttSym) devient :

$$\overline{Sym} = N(Sym) \quad (24)$$

Ceci permet d'inclure la symétrie de la façon suivante dans le modèle (VisAttCurvSym) :

$$S = \frac{1}{6} (N(\overline{J}) + N(\overline{C}_{RGB}) + N(\overline{C}_{DKL}) + N(\overline{O}) + N(\overline{Curv}) + N(\overline{Sym})) \quad (25)$$

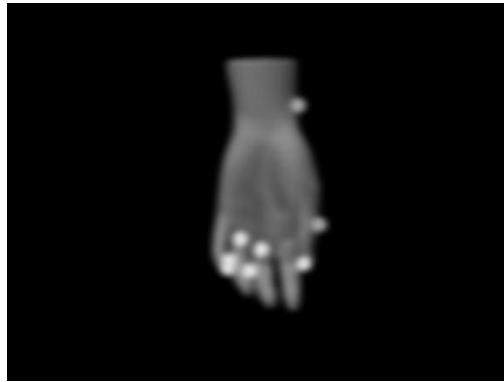


Figure 21- Exemple d'une carte de saillance Sym pour un point de vue du modèle de la main

Les résultats obtenus à partir de ce modèle de symétrie pour la détection des points d'intérêt sont présentés dans la section 5.7.

4.1.2.4. Ajout du canal de contraste

Lorsqu'un individu regarde une image, il aura tendance à être attiré par les régions de fort contraste alors que les régions peu contrastées seront davantage ignorées. Zhang et coll. [63] utilisent la luminance, la texture et le contraste de couleur comme les trois

composantes de leur modèle d'attention visuelle, alors que dans [41] une méthode de contraste basée sur l'histogramme est proposée pour améliorer la détection des régions saillantes. Dans le cadre de ce travail, la carte de saillance en niveaux de gris du contraste, Con , est calculée en utilisant la variance de luminance dans un voisinage de 80×80 pixels [64] et la carte du contraste (Figure 22), dénoté $VisAttCon$, est construite comme :

$$\overline{Con} = N(Con) \quad (26)$$

Ce qui crée l'équation suivante pour l'inclusion de tous les canaux ($VisAttCurvSymCon$):

$$S = \frac{1}{7} (N(\overline{J}) + N(\overline{C}_{RGB}) + N(\overline{C}_{DKL}) + N(\overline{O}) + N(\overline{Curv}) + N(\overline{Sym}) + N(\overline{Con})) \quad (27)$$



Figure 22- Exemple de carte de saillance du contraste pour le modèle de la main

4.1.2.5. Ajout du canal d'entropie

Kadir et Brady [65] proposent l'idée d'utiliser l'entropie comme une mesure de la complexité de signal local ou de l'imprévisibilité dans une image. On s'attend à ce que l'utilisation de l'entropie dans le calcul de l'attention visuelle mènera à de meilleurs résultats puisque des petites régions qui sont saillantes en raison de l'éclairage ou de l'unicité de la couleur ne sont pas nécessairement saillantes en général [66]. Dans le cadre de ce travail, l'image subit d'abord un prétraitement à l'aide d'un filtre médian, puis l'entropie est encodée comme une valeur d'entropie locale dans un voisinage de 9×9 autour du pixel correspondant dans l'image filtrée [67] :

$$Ent = - \sum_{i=0}^{L-1} p(J_i) \log_2 p(J_i) \quad (28)$$

où $p(J_i)$ est l'histogramme des niveaux d'intensité dans la région i et L le nombre possible de niveaux d'intensité (c.-à-d. $L=256$ pour l'expérimentation). La carte de perceptibilité pour l'entropie (VisAttEnt) est calculée comme :

$$\overline{Ent} = N(Ent) \quad (29)$$

La carte de saillance finale (VisAttAll) est calculée comme une moyenne des cartes de perceptibilités calculée indépendamment :

$$S = \frac{1}{8} \left(N(\overline{J}) + N(\overline{C}_{RGB}) + N(\overline{C}_{DKL}) + N(\overline{O}) + N(\overline{Curv}) + N(\overline{Sym}) \right. \\ \left. + N(\overline{Con}) + N(\overline{Ent}) \right) \quad (30)$$

4.1.3. Extraction des points d'intérêt

Une fois la carte de saillance calculée, afin de déterminer les régions d'intérêt à la surface de l'objet, il est nécessaire d'appliquer un seuillage sur cette carte. Le seuil permettra de conserver uniquement les régions où la saillance lui est supérieure, tel que montré dans la Figure 23.

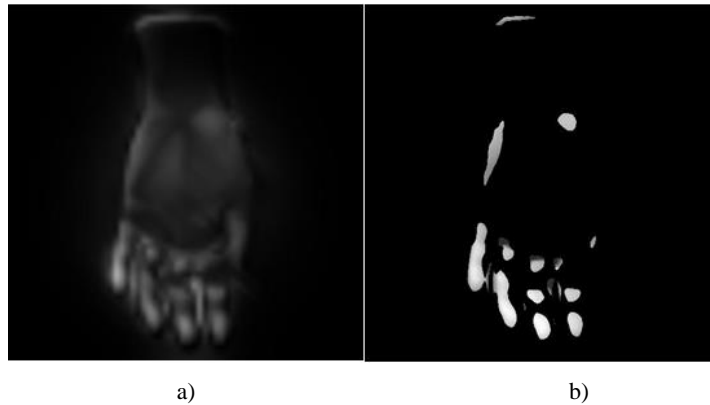


Figure 23- (a) Carte de saillance et (b) régions d'intérêt à la suite de l'application d'un seuillage

Il faut noter que le seuillage doit être déterminé en fonction des valeurs retournées sur la carte de saillance. Un exemple de l'impact des différents seuils est présenté dans la Figure 24.

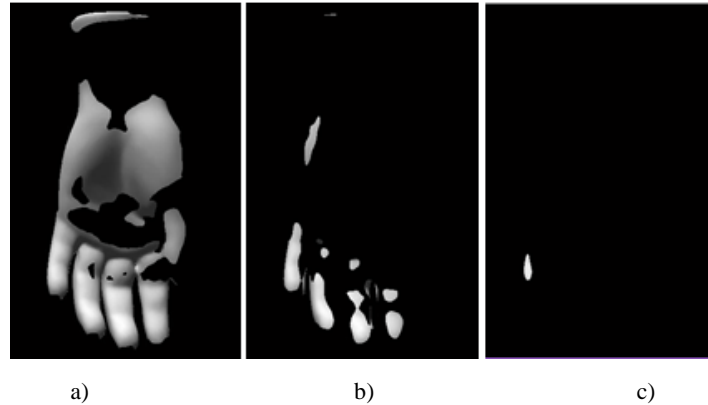


Figure 24- Comparaison de différents niveaux de seuillage. (a) 0.1, (b) 0.25 et (c) 0.5

On peut constater qu'une valeur trop basse de ce seuil engendre des régions trop larges (Figure 24a) et une valeur trop haute n'identifie pas suffisamment de régions (Figure 24c). Dans ce travail, le seuil appliqué sur les cartes de saillance en niveaux de gris permet de retenir 30 % (seuil de 0.3) des valeurs de saillance les plus élevées, comme dans la plupart des travaux sur le sujet dans la littérature, et par conséquent d'identifier les régions les plus intéressantes en termes d'attention visuelle, tel qu'illustré à la Figure 25.

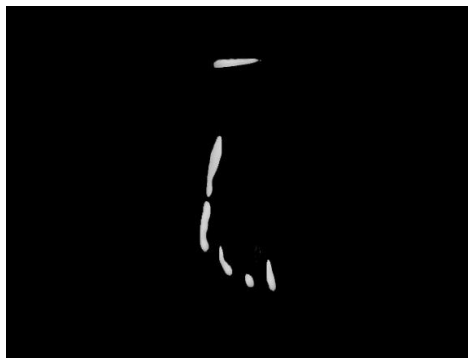


Figure 25- Effet de l'application d'un seuil de 0.3 sur le modèle de la main

Le modèle computationnel de l'attention visuelle, tel que montré dans la Figure 26, prend donc en entrée le modèle d'un objet à partir duquel une carte de saillance est créée de laquelle sont extraites des régions d'intérêt.

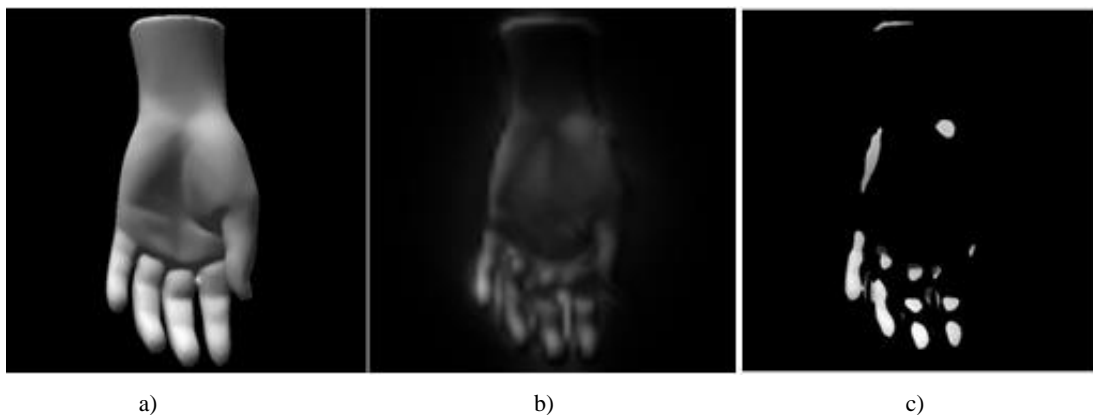


Figure 26- (a) Modèle de la main initial; (b) carte de saillance de l'image correspondant à la main vue de ce point de vue; (c) régions d'intérêt identifiées.

De plus, le modèle permet l'utilisation de points d'intérêt trouvés à partir de l'étude de la courbure ou de la symétrie, tel que décrit aux sections 4.1.2.2 et 4.1.2.3, respectivement.

4.1.4. Détection des points d'intérêt dans la carte de saillance

Puisque l'on cherche à retrouver des points d'intérêt plutôt que des régions, il est nécessaire de retrouver des points à l'intérieur de ces régions. La méthode proposée est d'identifier les points d'intérêt comme étant les centres des régions d'intérêt. Pour ce faire, la fonction Matlab `regionprops` est utilisée. Cette fonction permet entre autres de trouver le point qui identifie le centre d'une région à l'aide du paramètre `centroids`. Les coordonnées de ces points centraux sont retournées sous forme de matrices et il est possible de les visualiser sur l'image, tel que le montre la Figure 27.

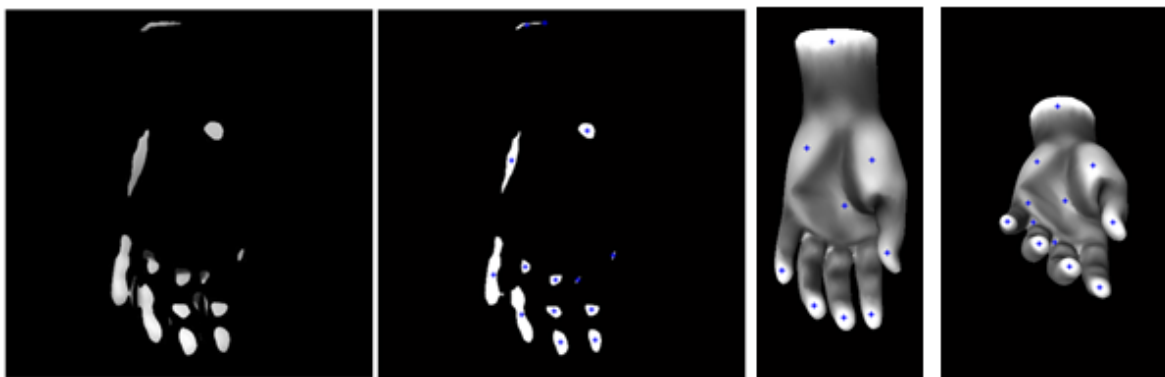


Figure 27- Détection des points d'intérêt à partir des régions d'intérêt comme points centraux des régions sur la carte de saillance et projetés en 3D.

4.1.5. Projection des points d'intérêt 2D sur le modèle 3D

La dernière étape du processus consiste à récupérer les coordonnées 3D des points trouvés, et ce, pour chacun des points de vue. Les quatre points de vue sont ceux où la caméra se situe sur l'axe des z positif, sur l'axe des z négatif, sur l'axe des y positif et sur l'axe des y négatif pointant chaque fois vers l'origine. Par conséquent, pour chaque point de vue, le troisième axe est caché.

Le plus long côté de l'objet suivant l'axe des z en pixels, calculé à partir de l'image, et divisé par la dimension réelle en unités 3D, résulte en le nombre de pixels par unité 3D, c'est-à-dire que pour chaque point, deux coordonnées peuvent être obtenues. Afin d'obtenir la troisième coordonnée, le modèle d'intersection triangle/rayon introduit par Moller et Trumbore [68] est utilisé. Cet algorithme fournit une solution rapide pour trouver toutes les intersections d'un rayon passant de chaque point perpendiculaire à la surface et donc, la plus proche intersection avec l'objet est considérée comme la troisième coordonnée du point d'intérêt du modèle de l'attention visuelle (Figure 28).

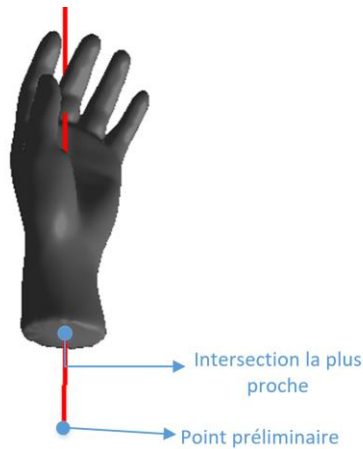


Figure 28- Exemple de l'intersection triangle/rayon

Soit un rayon $R(t)$ avec comme origine O et comme direction normalisée D défini comme :

$$R(t) = O + tD \quad (31)$$

où t représente la distance entre le point et la surface, et un triangle défini par trois sommets P_1, P_2 et P_3 . Un point $T(u, v)$ sur un triangle est défini par :

$$T(u, v) = (1 - u - v)P_1 + uP_2 + vP_3 \quad (32)$$

et en réarrangeant les termes, on obtient :

$$[-D, P_2 - P_1, P_3 - P_1] \begin{bmatrix} t \\ u \\ v \end{bmatrix} = O - P_1 \quad (33)$$

Cela implique que les coordonnées barycentriques (u, v) et la distance, t , de l'origine du rayon jusqu'au point d'intersection, peuvent être trouvées en résolvant le système d'équations ci-dessus.

Seulement les points trouvés à cette étape et leur voisinage immédiat seront préservés lors du processus de simplification. Ces points sont ensuite conservés dans une matrice sous la forme de coordonnées x, y, z . Cette matrice pourra ensuite être utilisée afin de projeter les points sur le maillage initial ou pour appliquer une simplification du maillage en tenant compte des points d'intérêt.

Le processus pour la détection des points d'intérêt utilisant le modèle d'attention visuelle humaine décrit dans les sections 4.1.1 à 4.1.5 peut être résumé par le pseudo-algorithme suivant :

Algorithme 1. Détection des points d'intérêt utilisant l'attention visuelle

// détection des points d'intérêt pour un objet **O** représenté par un maillage **M**

Recueillir un nombre prédéfini **n** de « photos » de l'objet de différents points de vue

pour chaque photo **faire**

 Appliquer le modèle d'attention visuelle pour construire la carte de saillance **SM_i**, $i=1, \dots, n$

 Récupérer les régions d'intérêt **R_j** en appliquant un seuillage sur la carte de saillance

pour chaque région d'intérêt **faire**

 Déterminer le point d'intérêt comme le centroïde de la région

fin

 Récupérer les coordonnées 3D des points d'intérêt

 Projeter les points d'intérêt sur le modèle initial de l'objet et les ajouter sur la liste des points, **P**

fin

4.2. Simplification sélective d'objets 3D

Une fois les points d'intérêt identifiés, l'algorithme de simplification proposé par [58], QSlim, est utilisé et adapté en une solution automatisée pour inclure les points d'intérêt détectés et leur voisinage afin de créer un maillage sélectivement simplifié.

L'algorithme QSlim peut brièvement être décrit comme suit : soit un maillage triangulaire $M(P, F)$ où P est un ensemble de sommets p_i et F , un ensemble de faces ou plus précisément un ensemble de triplets $\{p_j, p_k, p_l\}$ décrivant 3 sommets créant une face dans le maillage (dans le sens contraire des aiguilles d'une montre), QSlim simplifie un maillage en effectuant des contractions d'arêtes basées sur une métrique d'erreur quadrique calculée à chaque sommet du maillage, tel qu'illustré à la Figure 29.

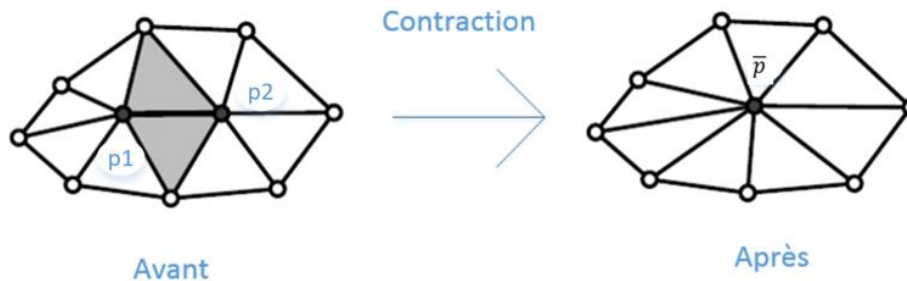


Figure 29- Contraction d'arêtes : l'arête en gras est contractée en un seul point. Les triangles en surbrillance sont éliminés durant la contraction.

Cette mesure d'erreur est une matrice 4×4 qui représente la somme des distances du sommet aux faces des triangles avoisinants. Si sa valeur est grande, le sommet correspondant pourrait être considéré comme un trait caractéristique à la surface de l'objet et donc être conservé plus longtemps dans le maillage. Sinon, le sommet sera retiré plus tôt. La même mesure est utilisée pour calculer le coût d'une contraction et la position optimale pour le sommet résultant de la suppression d'arêtes. De cette façon, chaque arête est associée à un coût de contraction et toutes les arêtes sont stockées dans une liste ordonnée selon leur coût. L'idée de l'algorithme est donc de retirer à chaque étape l'arête avec le coût le plus bas, de mettre à jour le voisinage résultant de la contraction et de mettre à jour le coût des arêtes connectées au nouveau sommet. Le paramètre le plus important est la résolution finale du maillage, soit le nombre de faces à conserver après la simplification. La plupart des solutions disponibles dans la littérature accordent un poids

plus élevé aux régions d'intérêt après la création d'un maillage simplifié [11, 13, 14] et ajustent leur coût dans le but de retarder leur simplification.

Dans ce travail, il est plutôt interdit au processus de simplification d'affecter les régions d'intérêt. En particulier, les faces du maillage contenant les points d'intérêt et leurs voisins immédiats (n -plus proches voisins) sont retirées de la liste de faces à être simplifiées par l'algorithme QSlim. Puisque le nombre de faces est déterminé au début de l'algorithme, les faces seront distribuées à la surface de l'objet d'une manière sélective, avec une densité plus élevée dans les régions d'intérêt.

Le pseudo-code pour l'algorithme de simplification peut être résumé comme suit :

Algorithme 2. Simplification sélective d'objets 3D

//simplification pour un maillage \mathbf{M} pour obtenir un maillage de densité sélective \mathbf{M}_s

Calculer \mathbf{PN} = les n -plus proches voisins du point \mathbf{P} dans \mathbf{M} , $n=1,2,\dots$

Récupérer les arêtes dans le maillage qui ne sont pas des points d'intérêt ou leurs voisins :

$$\mathbf{E}_D = \{e(u,v) \mid e(u,v) \in \mathbf{M}, e(u,v) \notin \mathbf{PN}\}$$

Initialiser $\mathbf{M}_s = \mathbf{E}_D$

// appliquer QSlim

Calculer l'erreur quadrique à chaque sommet de \mathbf{M}_s

Déterminer les coûts de contraction de chaque arête $e(u,v)$ dans \mathbf{M}_s

Créer une liste ordonnée des arêtes en fonction de leur coût

tant que la résolution désirée (nombre de faces) n'est pas atteinte **faire**

Retirer l'arête $e(u,v)$ de plus faible coût

Utiliser l'erreur quadrique pour choisir la cible de contraction optimale

Contracter u et v et recalculer le coût pour les arêtes adjacentes \mathbf{M}_s

fin

P , l'ensemble de points d'intérêt, peut être obtenu en utilisant n'importe lequel des détecteurs de points d'intérêt, par exemple, n'importe lequel des détecteurs utilisés à des fins de comparaison dans la section 6 ci-dessous. Puisque la méthode proposée de détection des points d'intérêt roule en Matlab, l'adaptation Matlab de QSlim est faite à partir de [69], et la même plateforme est utilisée pour implémenter la solution proposée.

Il est important de mentionner que tout autre algorithme de simplification disponible dans la littérature pourrait être utilisé plutôt que QSlim sans restreindre l'applicabilité de la

solution proposée. La principale raison de choisir QSlim est le fait que l'algorithme est reconnu pour assurer un des meilleurs équilibres entre la fidélité, la vitesse et la robustesse parmi les algorithmes de simplification disponibles [3].

4.3. Évaluation de la qualité du maillage

4.3.1. Évaluation via des mesures d'erreur

L'outil Metro [70] est utilisé pour calculer une mesure quantitative de qualité du maillage simplifié résultant. Cet outil permet de comparer deux maillages triangulaires qui représentent la surface d'un même objet à différentes résolutions. La mesure est basée sur le calcul de la distance Hausdorff entre un point et une surface. Le premier maillage (le maillage initial à pleine résolution de l'objet) est échantillonné et, pour chaque élément de surface, la distance au maillage simplifié est calculée. Metro retourne les distances (erreurs) maximales et moyennes ainsi que la variance. Ces mesures peuvent être considérées comme la mesure d'erreur entre les deux maillages [71] [72].

Afin d'évaluer la qualité des modèles sur un plan perceptuel, trois autres mesures d'erreur sont utilisées. La première se base sur la mesure de similarité structurelle (SSIM) et s'appuie sur l'observation selon laquelle la perception visuelle humaine est largement adaptée pour extraire l'information structurelle dans une scène [71]. L'inverse de cette mesure est employé comme mesure d'erreur puisqu'une similarité plus basse entre le maillage initial et le maillage simplifié indique une plus grande erreur. La deuxième catégorie d'erreurs est constituée par des métriques d'évaluation de qualité d'image basée sur des pyramides laplaciennes [72], soit deux mesures de la qualité d'image basées sur des transformations visuelles, c'est-à-dire la soustraction de la luminance locale et le contrôle de gain de contraste. Les auteurs suggèrent que de représenter les images dans une décomposition multi-échelles non linéaire peut mieux tenir compte des jugements sur la qualité de la perception humaine. Les deux formes reportées sont les distances prédites dans le domaine laplacien et dans le domaine laplacien normalisé. Puisque ces erreurs sont conçues pour être utilisées sur des images, des images sont prises des modèles simplifiés selon les mêmes points de vue que pour le modèle de l'attention visuelle et sont comparées avec les images des modèles initiaux (non simplifiés). Les mesures d'erreur pour chaque

objet sont présentées comme une moyenne pour tous les points de vue et les résultats finaux sont calculés comme une moyenne pour tous les objets dans la base de données.

Finalement, l'outil Cloud Compare [73] est utilisé pour visualiser les régions les plus atteintes par l'erreur.

4.3.2. Évaluation via une réalité de terrain générée par l'humain

Dans [1], les auteurs proposent une stratégie d'évaluation des détecteurs des points d'intérêt basée sur une réalité de terrain généré par l'humain. Ils offrent une mesure quantitative d'évaluation en fonction des points d'intérêt détectés par des sujets humains. Le jugement des individus en ce qui concerne les points d'intérêt est subjectif par nature. La Figure 30 illustre les points d'intérêt identifiés par trois individus sur le modèle de la fourmi ainsi que celui de l'Armadillo.

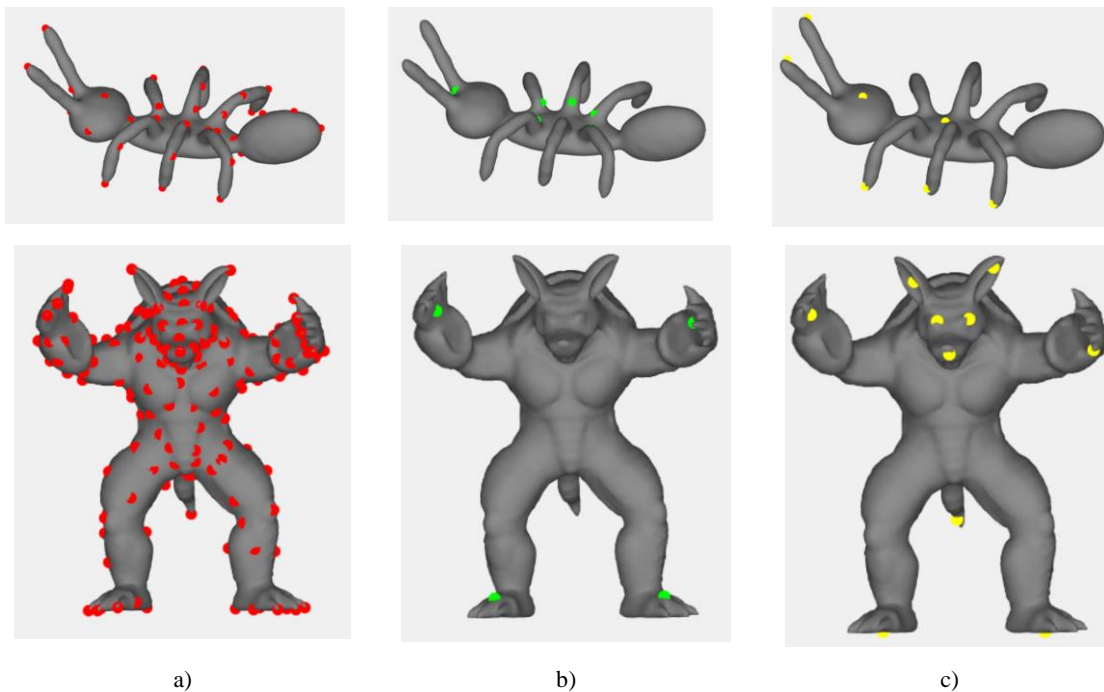


Figure 30- Points d'intérêt identifiés par 3 individus différents (a)-(c) sur le modèle de la fourmi et de l'Armadillo.

Les auteurs ont donc cherché un consensus entre les utilisateurs afin de créer un ensemble final de points d'intérêt constituant l'ensemble de réalité de terrain. Ils ont utilisé deux critères afin de construire cet ensemble soit : le rayon d'une région d'intérêt et le nombre d'utilisateurs n_{sujets} qui ont identifié un point dans cette région. Le rayon d'une région

d'intérêt est identifié comme d_M , où d_M représente le diamètre du modèle c'est-à-dire la plus grande distance euclidienne entre toutes les paires de sommets du modèle M et φ est une constante choisie lors de la construction de la réalité de terrain (p. ex. 0.03 ou 0.05). Tous les points d'intérêt marqués par différents individus qui sont à une distance plus petite que $2\varphi d_M$ les uns des autres sont regroupés. Si le nombre de points d'intérêt dans le groupe est plus petit que n_{sujets} , alors le groupe est rejeté. Sinon, un représentant du groupe est sélectionné, le point ayant la plus petite distance géodésique avec les autres points, et identifié comme un point faisant partie de la réalité de terrain. Le nombre de sujets ayant identifié ce point est aussi conservé comme étant la proéminence.

La méthode proposée dans ce travail est donc évaluée à l'aide de la réalité de terrain générée par les données compilées dans [1]. Les résultats obtenus sont présentés à la section 0 et sont comparés avec les résultats obtenus pour les autres détecteurs de points d'intérêt. Il est important à mentionner dès le départ que ce processus d'évaluation est en effet aussi biaisé que le processus individuel d'identification des points d'intérêt. Bien qu'il propose une approche intéressante pour trouver le consensus, ceci n'est pas appuyé par des études sur la perception ou la vision humaine. Il ne prend pas en considération les caractéristiques qui attirent l'attention humaine. De ce fait, il est fort possible que cette évaluation ne reflète pas les caractéristiques perceptives humaines. Par exemple, si plusieurs sujets identifient des points sur une même ligne (Figure 30) comme étant d'intérêt, au lieu de choisir un point parmi ceux-ci, il sera plus utile à considérer la ligne entière comme étant d'intérêt.

4.4. Incorporation des maillages sélectivement densifiés dans une méthode discrète à multiples niveaux de détails

Comme il est difficile de juger du nombre de faces nécessaire pour un modèle situé à une certaine distance par rapport à un utilisateur, la solution proposée est d'utiliser un réseau de neurones. L'intérêt d'utiliser des réseaux de neurones à cet effet découle de leur capacité de fournir des estimations pour des données qui ne font pas partie de l'ensemble d'entraînement, ce qui signifie dans le contexte actuel que le nombre de faces peut être prédit pour des objets dont les caractéristiques sont similaires, mais pas identiques à ceux utilisées pendant la phase d'entraînement. Afin d'incorporer les maillages sélectivement

densifiés dans une méthode discrète à multiples niveaux de détails, on utilise la solution proposée dans [74]. Le réseau de neurones est entraîné pour apprendre l'association entre le nombre de faces à utiliser pour les copies multiples de l'objet à différentes résolutions et le maillage initial. En entrée, le réseau de neurones prend le maillage initial (résolution maximale) et retourne une série de nombres de faces croissant. À l'exécution, le maillage d'un objet inconnu est entré dans le réseau de neurones et celui-ci retourne le nombre de faces à utiliser pour chaque copie, de la plus grossière à la plus détaillée. Finalement, l'algorithme de simplification sélective est appliqué sur le modèle pour chaque copie en retenant le nombre de faces proposées par le réseau de neurones.

4.5. Incorporation des maillages sélectivement densifiés dans une méthode continue à multiples niveaux de détails

Une série d'architectures en aval (*feedforward*) à deux couches est utilisée, un réseau étant associé à chaque version du modèle d'attention visuelle constitué de diverses combinaisons de canaux. Le rôle de chacun est d'apprendre le nombre de faces qui doivent être utilisées dans un modèle simplifié basé sur les caractéristiques de l'objet. Les caractéristiques des objets sont basées sur l'erreur tolérée, la distance entre l'objet et l'utilisateur, et la complexité de l'objet, c'est-à-dire la taille du maillage et le nombre de points d'intérêt identifiés. L'utilisation de ces caractéristiques est justifiée de la façon suivante : premièrement, l'utilisateur a le choix de contrôler le degré de précision du modèle (l'erreur tolérée); une valeur de 0.05 est sélectionnée par défaut pour l'erreur tolérée. La distance entre le modèle et l'objet joue un rôle important dans la modélisation à multiples niveaux de détails. Comme mentionné dans l'introduction, plus l'objet est loin de l'utilisateur dans l'environnement 3D, moins la représentation de l'objet devrait être détaillée afin de permettre une vitesse de rendu plus rapide. La complexité de l'objet a aussi un impact important sur la qualité des résultats. Un modèle plus complexe nécessite plus de faces pour préserver une bonne représentation de l'objet. Afin de considérer tous ces facteurs, chaque réseau de neurones proposé, d'une taille de 30 neurones dans la couche cachée, a quatre entrées, soit l'erreur tolérée, la distance entre l'utilisateur et l'objet, la taille initiale du maillage et le nombre de points d'intérêt et une sortie, le nombre de faces.

Chaque réseau (correspondant à une combinaison de canaux de caractéristiques) doit être entraîné pour apprendre l'association entre le nombre de faces en sortie et les variables d'entrées. Afin d'entraîner les réseaux, il est nécessaire de fournir des valeurs pour chaque variable d'entrée et de sortie. Les séries de mesures d'erreurs (p. ex. Metro max dans l'expérimentation, mais n'importe quel autre type d'erreur peut être utilisé) sont calculées comme décrit dans la section 4.3 pour une échelle de résolutions allant de 1500 faces au nombre total de faces dans le modèle initial, pour une différente combinaison de canaux d'attention visuelle. Pour chaque combinaison, la taille du modèle initial et le nombre de points d'intérêt sont aussi enregistrés. Les valeurs de distances sont déterminées dans VRML en déplaçant graduellement l'objet de plus en plus loin de l'utilisateur et en marquant les distances quand des caractéristiques importantes semblent disparaître. Un changement dans la résolution est attendu à ces distances. Une fois toutes ces données obtenues, chaque réseau est entraîné en utilisant la rétropropagation de descente de gradient avec une valeur constante de momentum de 0.95 et un taux d'apprentissage adaptatif. Une erreur de somme quadratique nulle est visée sur 1000 périodes d'apprentissage. Une fois le réseau entraîné, il peut fournir une estimation du nombre de faces pour chaque combinaison de valeurs d'entrée. Le nombre final de faces est calculé comme une moyenne des résultats obtenus par chaque réseau dans la série.

5. Résultats expérimentaux

Plusieurs tests ont été effectués afin de valider les diverses composantes de la solution proposée, et pour étudier l'impact des divers paramètres du modèle sur la performance. Les résultats de ces tests sont présentés dans ce chapitre. La base de données [75] a été choisie pour l'expérimentation. Ce choix est justifié par le fait qu'une série de méthodes de détection des points saillants sont incluses (en termes de coordonnées x , y , z pour chaque point d'intérêt et chaque méthode), en permettant une comparaison directe avec la solution proposée.

5.1. Influence des points de vue utilisés

Un des paramètres à déterminer dans l'approche proposée ayant un impact sur sa performance est le nombre de points de vue de l'objet utilisés pour trouver les points d'intérêt. Durant les premières expérimentations, le nombre de points de vue a été varié de 4 à 12, chacun étant le résultat d'une rotation spécifique selon l'axe des x et des z respectivement, tel que montré dans la Figure 31.

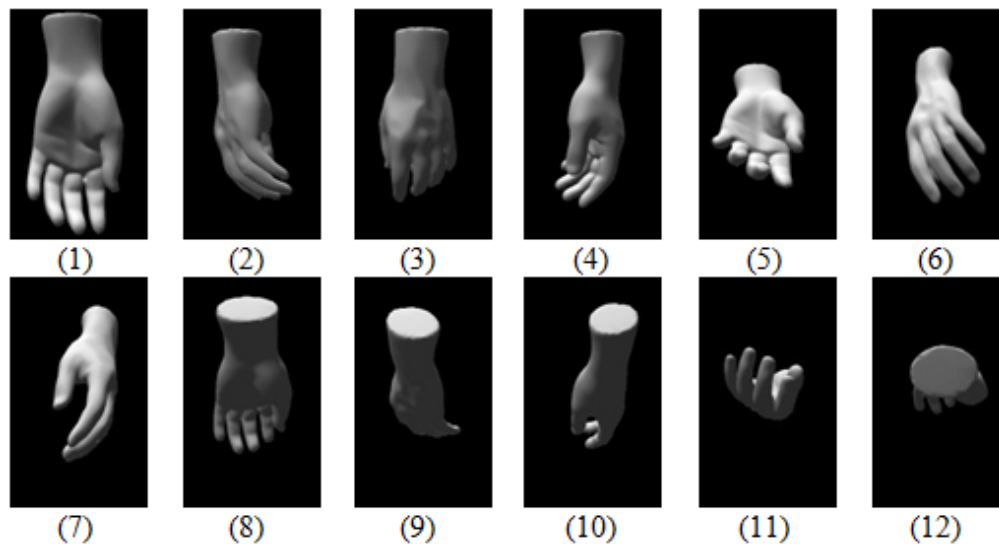


Figure 31- Points de vue pour le calcul de l'attention visuelle : (1) placement initial, rotation de : (2) 90° selon z , (3) 180° selon z , (4) 270° selon z , (5) 45° selon x , (6) 120° selon z et 45° selon x , (7) 240° selon z et 45° selon x , (8) -45° selon x , (9) 120° selon z et -45° selon x , (10) 240° selon z et -45° selon x , (11) 90° selon x (12) 180° selon x .

L'impact du nombre de points de vue ainsi que de l'utilisation des leurs différentes combinaisons a été étudié afin d'identifier la meilleure combinaison de points de vue possible. Les mesures d'erreur décrites précédemment, incluant l'erreur maximale, l'erreur moyenne et la variance, sont calculées comme une moyenne pour les objets dans

l'ensemble de données. La meilleure combinaison est celle ayant les mesures d'erreur les plus basses. Les résultats sont montrés dans la Figure 32.

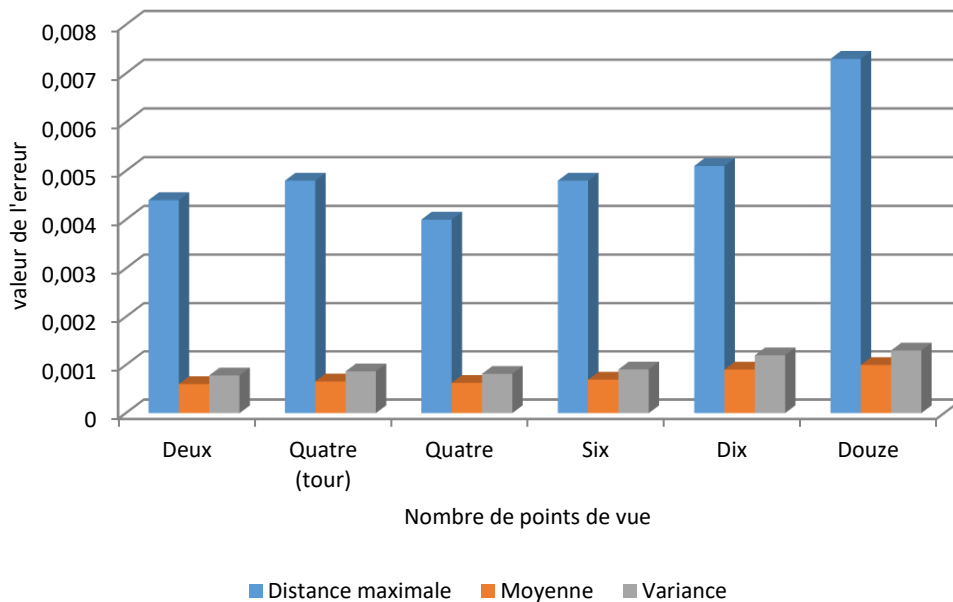


Figure 32- Erreurs Metro selon le nombre de points de vue. Ici, « quatre tour » utilise les points de vue 1, 2, 3 et 4 alors que « quatre » utilise les points de vue 1, 3, 11 et 12.

La figure ci-dessus montre qu'en général, plus le nombre de points de vue est élevé, plus les mesures d'erreur sont élevées. Cela s'explique par le fait que plus le nombre de points de vue augmente, plus le nombre de points d'intérêt identifiés augmente aussi. Puisque l'algorithme de simplification préserve les points d'intérêt et leurs voisins, seul un nombre limité de faces sont affectées par la simplification, résultant dans l'augmentation des mesures d'erreur. Pour les expérimentations préliminaires sur la base de données, la combinaison des points de vue 1, 3, 11 et 12 montrés à la Figure 31 (identifiée comme « quatre ») a été retenue puisqu'elle mène globalement aux mesures d'erreur plus faibles.

La Figure 33 montre un échantillon des résultats obtenus en utilisant la simplification avec la préservation des points d'intérêt pour 3 objets extraits de la base de données, pour un nombre de faces allant de 4000 à 700, pour le calcul de l'attention visuelle sur les quatre points de vue précédemment identifiés et pour un voisinage de préservation de $n=3$ autour d'un point d'intérêt. Il est possible de remarquer que même aux plus basses résolutions, les principaux traits caractéristiques des objets sont préservés dans le modèle.

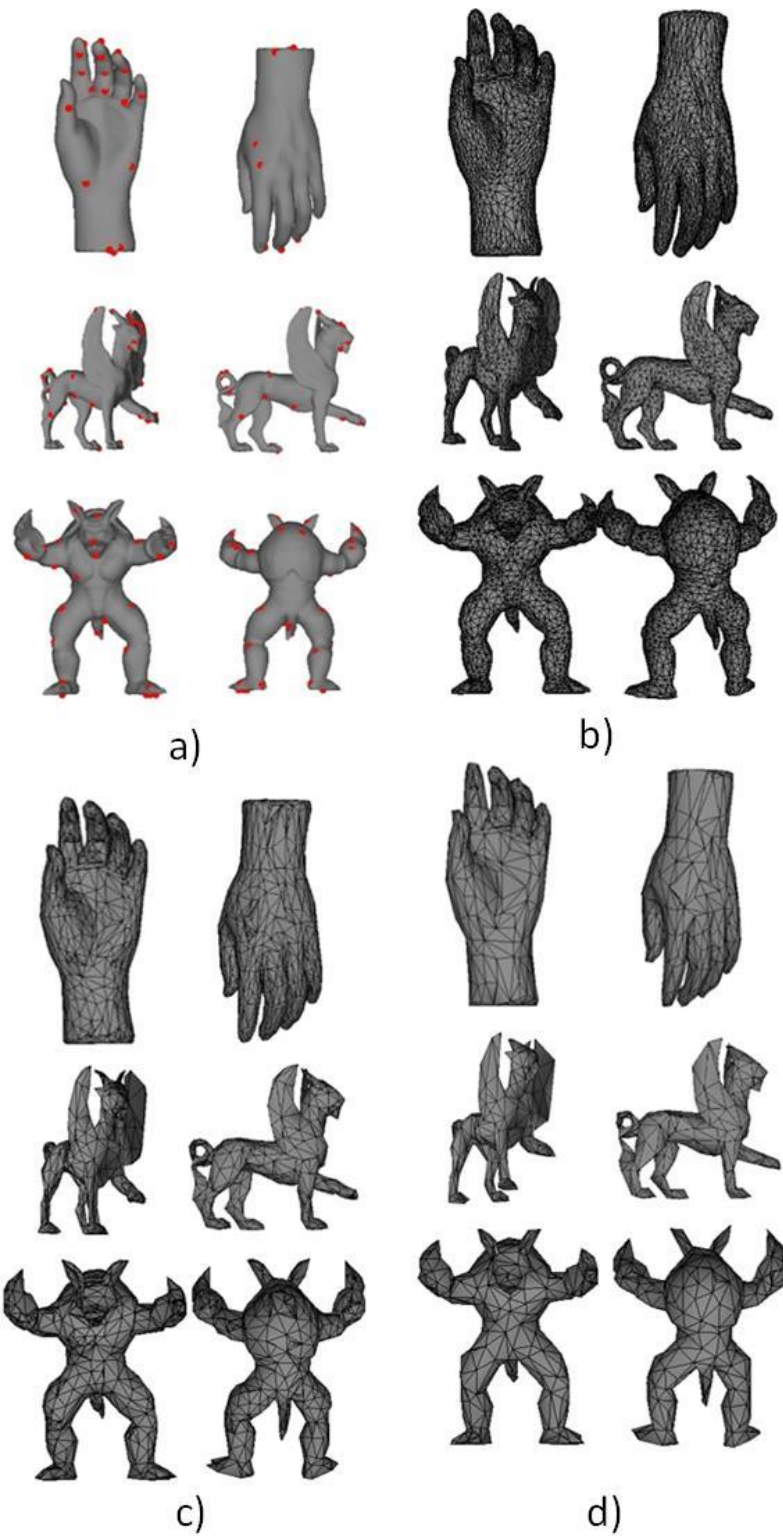


Figure 33- (a) Le modèle vu de deux points de vue avec les points d'intérêt identifiés en rouge et les modèles à différentes résolutions obtenus avec l'approche proposée pour (b) 4000, (c) 1500 et (d) 700 faces respectivement dans le maillage simplifié.

5.2. Influence du voisinage

À la lumière des résultats obtenus par [74], un voisinage de $n=3$ (3-plus-proches-voisins) est choisi pour les expérimentations. En effet, tel qu'illustré dans la Figure 34 pour un nombre de faces de 1500, les auteurs ont démontré que l'erreur tend à augmenter lorsque le voisinage augmente. En général, avec des maillages plus grossiers, un voisinage plus petit, tel que 1 ou 2, est préférable alors que pour des maillages plus détaillés, un voisinage de 3 ou 4 peut être utilisé. Lors de leurs expérimentations avec différents maillages comportant jusqu'à 35 000 faces, généralement, un voisinage de 3 donne de bons résultats pour les maillages de plus de 5 500 faces, un voisinage de 2 donne de bons résultats pour des maillages entre 3 500 et 5 500 faces et un voisinage de 1 pour ceux de moins de 3500 faces.

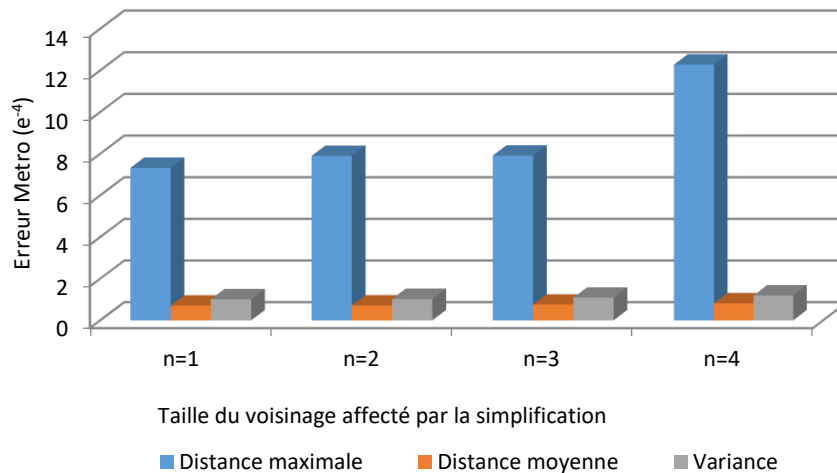


Figure 34- Mesures d'erreur en fonction de la grandeur du voisinage.

Bien que cette façon de procéder pour déterminer le voisinage soit fonctionnelle, une méthode de sélection adaptative du voisinage a été testée. Pour ce faire, l'aire des régions d'intérêt identifiées à partir de la carte de saillance est utilisée. L'idée exploitée est que plus un modèle produit des régions de grande surface, plus le voisinage sélectionné devrait être grand. L'aire des régions est donc calculée pour chacun des 4 points de vue et une moyenne est faite pour chaque objet. Cette moyenne est ensuite utilisée pour déterminer la

taille du voisinage à utiliser. Les tailles sélectionnées vont de 1 à 5. La Figure 35 ci-dessous montre les erreurs obtenues comme une moyenne sur l'ensemble de la base de données pour une réduction à 1500 faces. Un voisinage de 3, en bleu, et le voisinage adaptatif, en orange, sont comparés. La méthode de voisinage adaptatif est testée sur les méthodes décrites à la section 4.1.2, à savoir : VisAtt (attention visuelle seulement), VisAttCurv (attention visuelle et courbure), VisAttSym (attention visuelle et symétrie), VisAttCon (attention visuelle et contraste), VisAttEnt (attention visuelle et entropie), VisAttAll (attention visuelle, courbure, symétrie, contraste et entropie), VisAttCurvSym (attention visuelle, courbure et symétrie) et VisAttCurvSymCon (attention visuelle, courbure, symétrie et contraste). Il est possible de remarquer que le voisinage adaptatif mène généralement à des erreurs plus faibles, à l'exception de la méthode VisAtt et de la méthode VisAttSym.

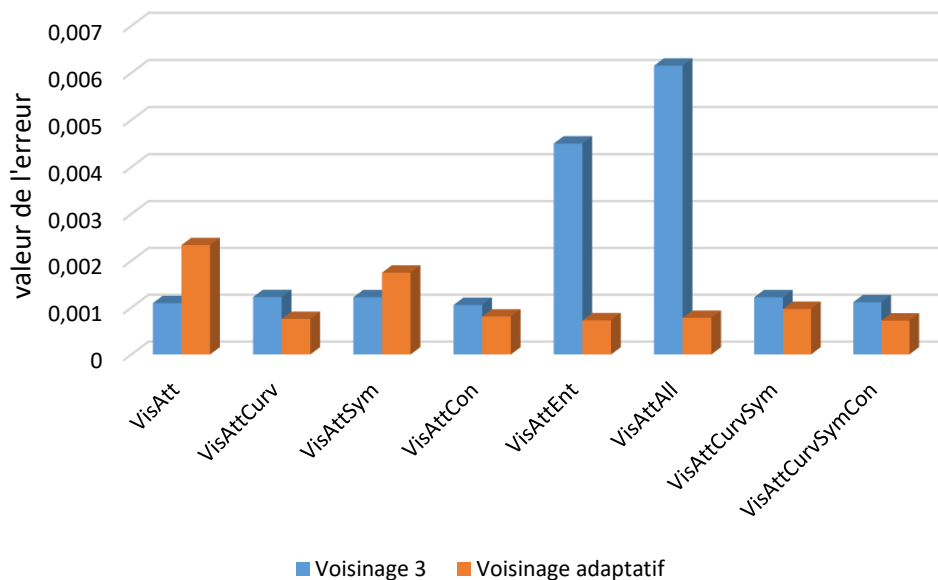


Figure 35- Comparaison des erreurs pour un voisinage de 3 ou un voisinage adaptatif.

La Figure 36 montre les résultats visuels d'une simplification avec un voisinage adaptatif pour quatre modèles de la base de données. La simplification est effectuée avec la préservation des points d'intérêt retournés par la méthode VisAttCurvSymCon pour un nombre de 1500 faces.

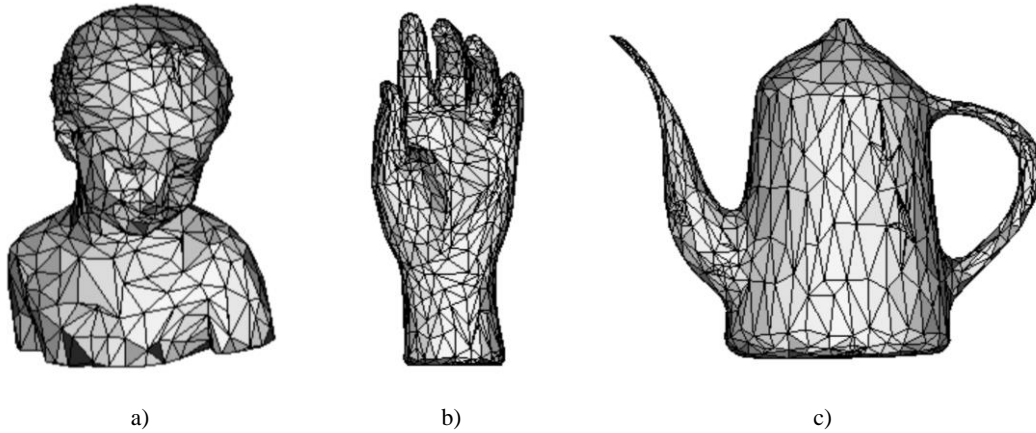


Figure 36- Simplification avec voisinage adaptatif pour la méthode VisAttCurvSymCon et 1500 faces pour (a) le modèle de la fille et un voisinage de 1; (b) le modèle de la main et un voisinage de 2; et (c) le modèle de la théière et un voisinage de 3.

5.3. Influence de la couleur de fond

Comme les modèles sont présentés sur un fond de couleur uni, l'étude de l'influence de cette couleur sur les résultats a été faite. Trois couleurs ont été testées, soit le noir (VisAtt-Noir), le blanc (VisAtt-Blanc) et le gris (VisAtt-Gris) (Figure 38). Les résultats obtenus en termes des mesures d'erreur sont illustrés à la Figure 38. On peut donc observer qu'un fond noir est plus approprié pour l'identification des points d'intérêt (les erreurs sont les plus basses) et c'est cette configuration qui sera utilisée pour la suite des expérimentations.

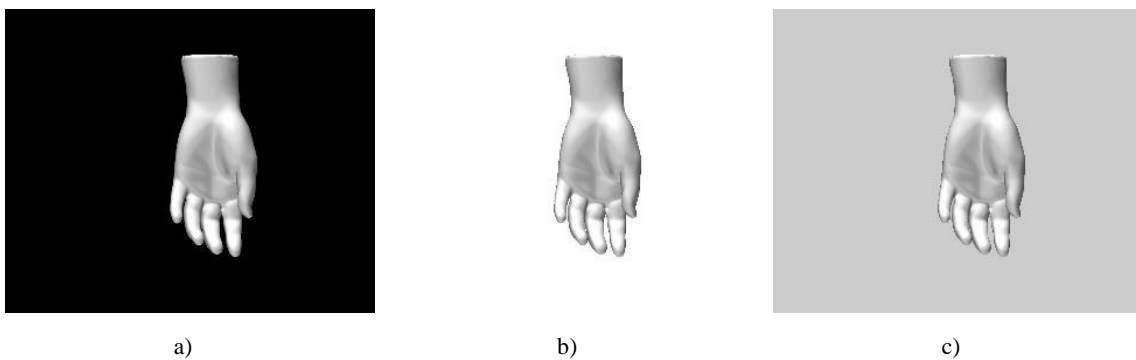


Figure 37-Différentes couleurs de fond pour le modèle de la main.

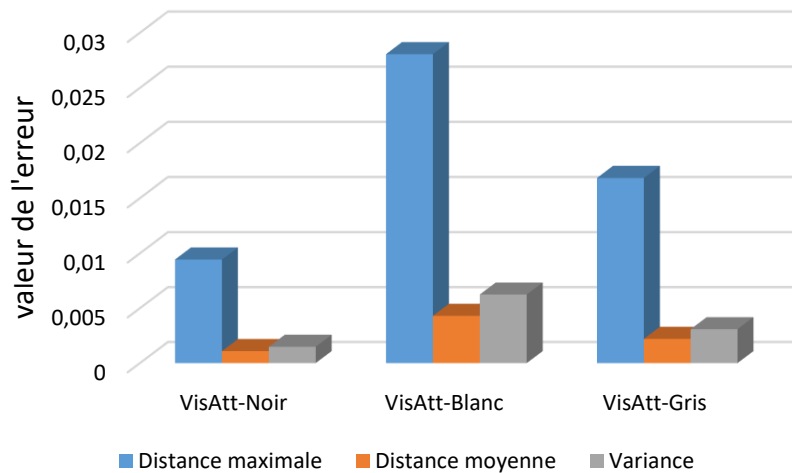


Figure 38- Influence de la couleur de fond.

5.4. Influence de la lumière dans la scène graphique

Puisque la lumière et les ombres peuvent affecter le déploiement de l'attention visuelle, des tests ont été effectués pour évaluer leur impact sur les objets dans notre base de données. Deux modèles de lumière ont été évalués afin de visualiser l'effet qu'ils provoquent sur la qualité des maillages simplifiés. Les modèles de lumière sont ceux offerts dans Matlab, soit *Gouraud* et plat (*Flat*) qui ont été décrits à la section 2.1. La Figure 39 montre le modèle de l'avion sous les deux modèles de lumière. Il est possible de remarquer que la lumière de type plat (Figure 39b) crée un modèle où chaque face peut être visible en raison des différences d'ombrage aux limites de celles-ci.

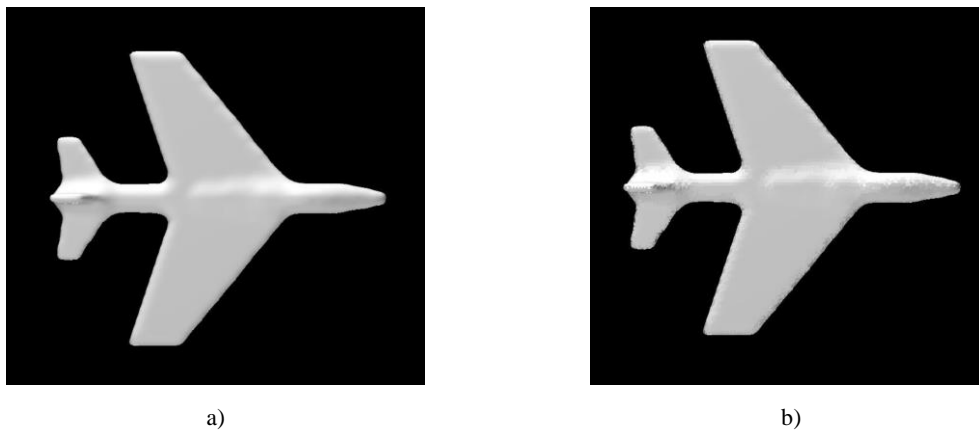


Figure 39- Influence du modèle de lumière, soit (a) *Gouraud* ou (b) *plat* (*Flat* en Matlab).

La Figure 40 montre les erreurs Metro obtenues pour les deux modèles de lumière, soit *Gouraud* en bleu et *Flat* en orange. Les erreurs sont calculées comme la moyenne sur l'ensemble de la base de données, pour la méthode de base du modèle computationnel de l'attention visuelle (VisAtt) et pour une réduction à 1500 faces. Il est possible de remarquer que les erreurs sont très similaires et par conséquent, le modèle de lumière *Gouraud* est employé pour les expérimentations suivantes.

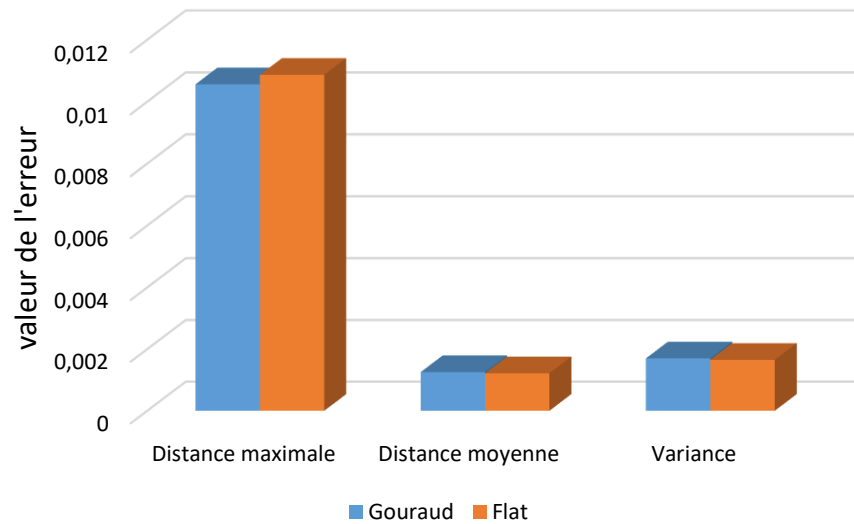
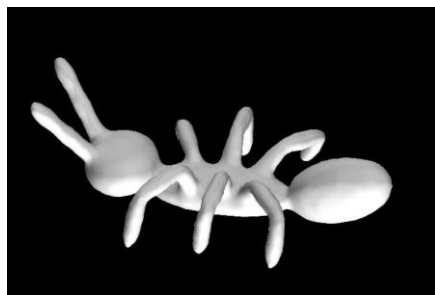


Figure 40- Erreur Metro pour l'influence de la lumière, pour la méthode VisAtt et une réduction à 1500 faces.

5.5. Influence du matériau de l'objet

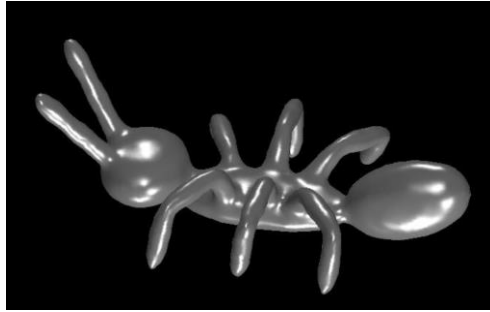
L'influence du matériau utilisé pour modéliser l'objet a aussi été étudiée. Encore une fois, les matériaux sélectionnés sont ceux disponibles dans Matlab, soit *Dull*, *Shiny* ou *Metal*. La Figure 41 montre le modèle de la fourmi créé avec les trois matériaux.



a)



b)



c)

Figure 41- Modèle de la fourmi avec les 3 formes de matériaux soit a) dull, b) shiny et c) metal.

La Figure 42 montre les erreurs Metro selon le matériau employé. Les erreurs sont calculées comme une moyenne sur l'ensemble de la base de données, pour la méthode VisAtt et une réduction à 1500 faces. Il est facilement observable que le matériau menant aux erreurs les plus basses est le matériau *Dull*. C'est donc celui-ci qui est employé pour la suite des expérimentations.

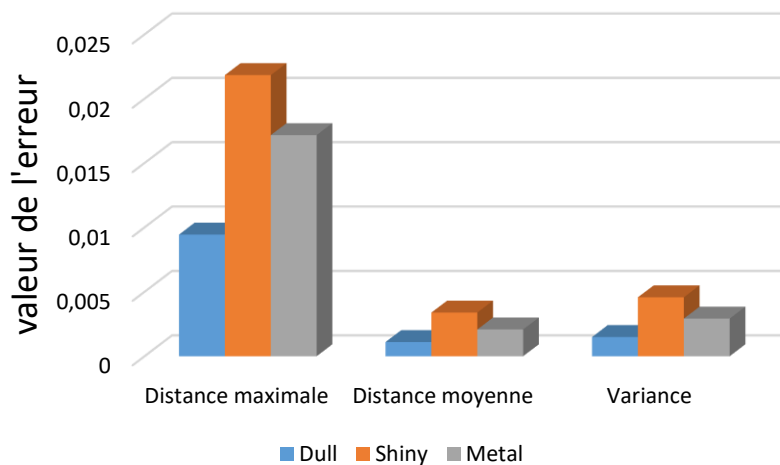


Figure 42- Influence du matériau sur les erreurs Metro.

5.6. Influence de la courbure

Pour étudier l'influence de la courbure, la Figure 43 compare les modèles simplifiés lorsque seulement le modèle de l'attention visuelle classique est utilisé (VisAtt) (couleur, intensité et orientation), lorsque seulement la courbure est utilisée (Curv) et lorsque la courbure est utilisée conjointement au modèle de l'attention visuelle, et ce, des deux façons expliquées à la section 4.1.2.2 (VisAttCurvP et VisAttCurv).

Encore une fois, les quatre points de vue identifiés précédemment (1, 3, 11 et 12) et un voisinage de $n=3$ sont utilisés. Un nombre de faces de 3000 (deuxième ligne de la Figure 43) et de 1500 (troisième ligne de la Figure 43) est utilisé pour le calcul des modèles simplifiés. Les mesures d'erreur obtenues dans ces quatre cas sont résumées dans la Figure 44. On peut observer que l'attention visuelle et la courbure, pris séparément, sont près en termes d'erreurs, alors que la combinaison des deux se solde par des erreurs plus élevées. Cependant, lorsque la courbure est incluse dans le modèle de l'attention visuelle sous la forme de l'équation 14, les erreurs obtenues sont légèrement plus basses. Les résultats obtenus en utilisant le modèle de l'attention visuelle, avec et sans la courbure, sont davantage comparés avec d'autres détecteurs de points d'intérêt à la section 5.10. Comme on peut s'y attendre et comme ce qui est confirmé dans la Figure 45, qui compare les erreurs pour une réduction à 1500 faces (en orange) et pour une réduction à 3000 faces (en beige), les erreurs diminuent avec une augmentation du nombre de faces dans le modèle simplifié. La différence dans les erreurs est aussi plus remarquable à de très basses résolutions.

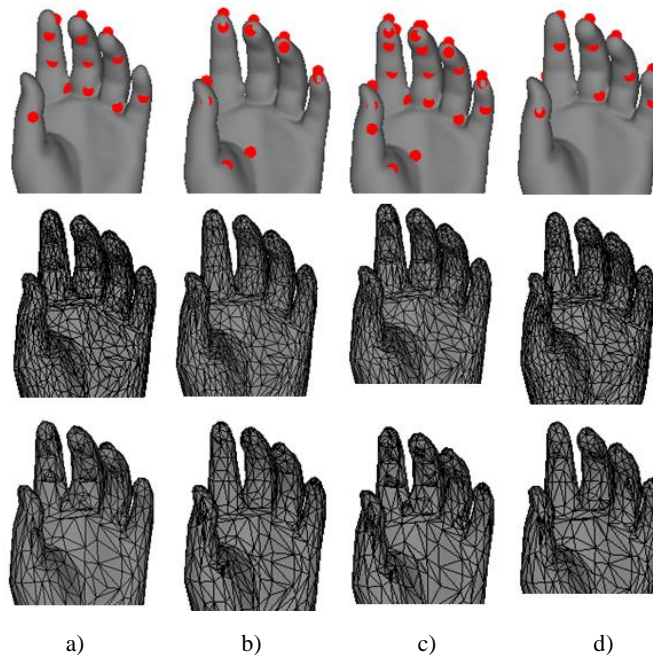


Figure 43- Influence de l'utilisation de la courbure pour les méthodes : (a) VisAtt, (b) Curv, (c) VisAttCurvP, et (d) VisAttCurv.

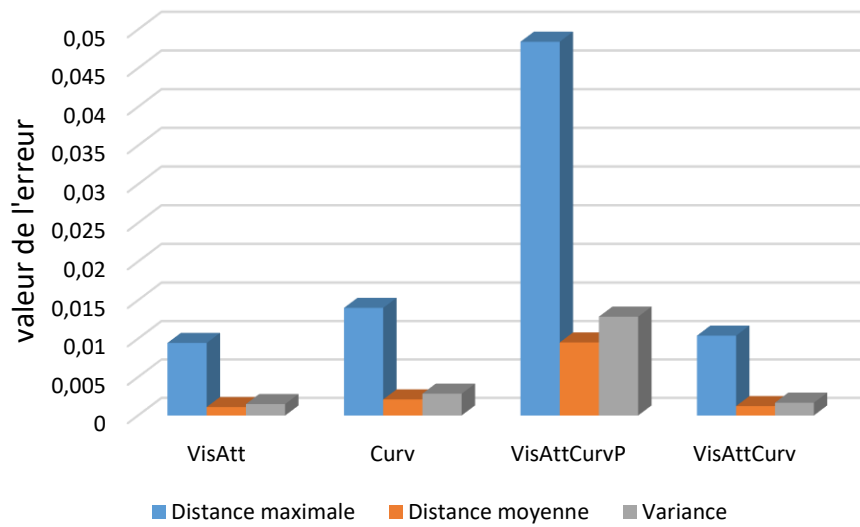


Figure 44- Influence de l'utilisation de la courbure dans la mesure d'erreurs.

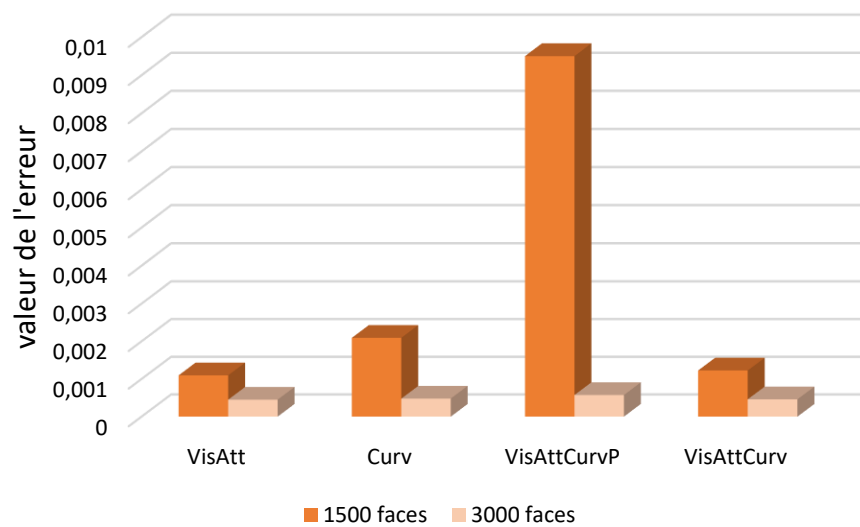


Figure 45- Influence du canal de courbure pour différentes résolutions de simplification.

5.7. Influence de la symétrie

L'influence de la symétrie a été évaluée pour les quatre points de vue 1, 3, 11 et 12 montrés à la Figure 31 (identifiée comme « quatre ») et identifiés à la section 5.1. En ce qui concerne la symétrie bilatérale, les points d'intérêt détectés sur le modèle de la main sont visibles en rouge à la Figure 46.



Figure 46- Points détectés par la méthode de symétrie bilatérale. Aucun point n'est détecté dans la deuxième figure (à partir de la gauche).

Pour la symétrie rotationnelle, le modèle ne détecte des points symétriques que pour un des points de vue, soit celui visible à la Figure 47.

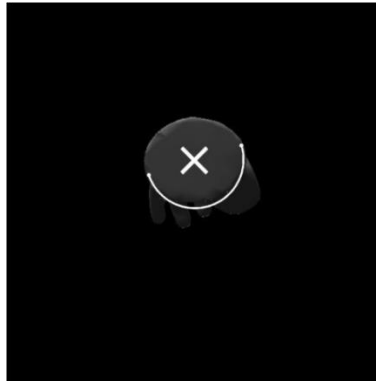


Figure 47- Point d'intérêt détecté par la symétrie rotationnelle.

La Figure 48 montre les résultats obtenus pour une simplification à 3000 (milieu) et à 1500 faces (bas), respectivement, pour la symétrie bilatérale (SymBiLat), la symétrie rotationnelle (SymRot), les deux symétries ensemble (2Sym) et la symétrie avec l'attention visuelle, et ce, des deux façons expliquées à la section 4.1.2.3 (VisAttSymP et VisAttSym).

En ce qui concerne les résultats obtenus au niveau des erreurs Metro, ils sont présentés à la Figure 49, ci-dessous. Ici, l'erreur maximale est plus grande lorsqu'on jumelle les points d'intérêt détectés par la symétrie et par l'attention visuelle. Il est possible que cette différence s'explique par l'augmentation du nombre de points d'intérêt.

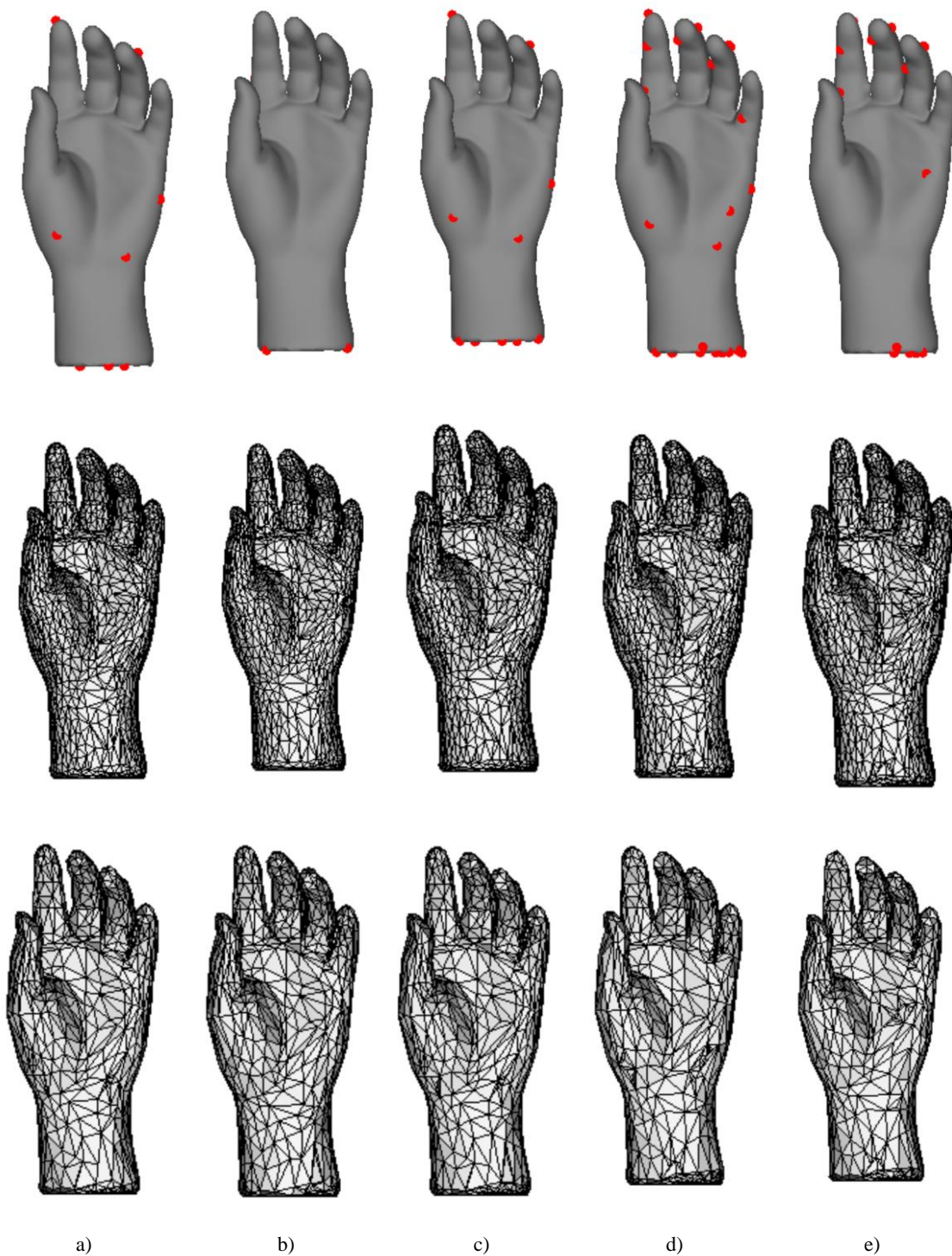


Figure 48- Influence de la symétrie : (a) SymBiLat, (b) SymRot, (c) 2Sym, (d) VisAttSymP et (e) VisAttSym. En haut : modèle initial avec points d'intérêt identifiés en rouge. Milieu : simplification à 3000 faces. En bas : simplification à 1500 faces.

Par contre, l'erreur moyenne est presque la même, que l'on considère seulement un type ou deux de symétrie, ou l'attention visuelle et la symétrie. Bien que la symétrie rotationnelle ne retourne que deux points d'intérêt, l'erreur maximale est près de celle de la symétrie bilatérale, mais avec une variance beaucoup plus importante.

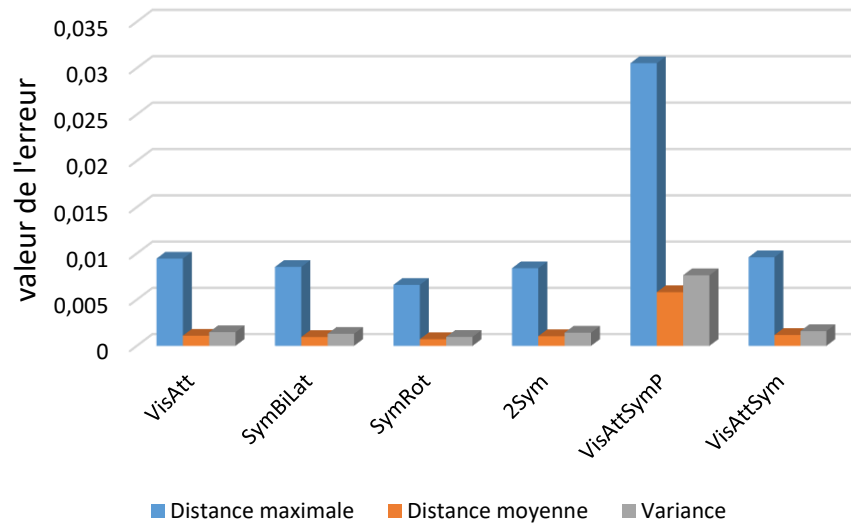


Figure 49- Valeur des erreurs Metro pour l'influence de la symétrie.

Encore une fois, les erreurs diminuent avec un nombre de faces plus élevé dans le modèle simplifié, tel qu'observable à la Figure 50.

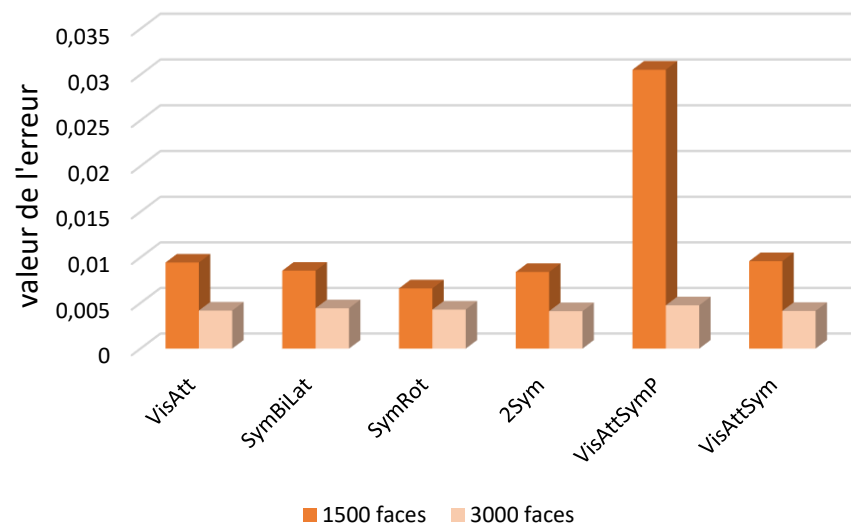


Figure 50- Influence du nombre de faces dans le modèle simplifié pour l'ajout du canal de symétrie.

5.8. Influence du contraste, de l'entropie et de différentes combinaisons de canaux

La Figure 51 et la Figure 52 présentent les mesures d'erreur quand différentes combinaisons de canaux sont utilisées pour le calcul de la carte de saillance. On peut observer que les différentes combinaisons, à l'exception du cas dans lequel tous les canaux soient considérés, obtiennent à peu près les mêmes erreurs que le modèle classique (une différence de 0.0002 est notée) ce qui implique que l'ajout de canaux apporte de l'information qui n'est pas déjà disponible dans le modèle classique. Par contre, quand tous les canaux sont utilisés, l'erreur est légèrement plus élevée. Cela est principalement dû à l'ajout du canal d'entropie (VisAttEnt) et provient probablement du fait que les objets testés n'ont pas de texture, alors que l'entropie peut être directement considérée comme une mesure de changement dans la texture. L'information sur l'entropie est plus pertinente pour les objets texturés, tel que démontré à la section 5.9. Finalement, le modèle utilisant la courbure, la symétrie et le contraste (VisAttCurvSymCon) mène aux meilleures qualités.

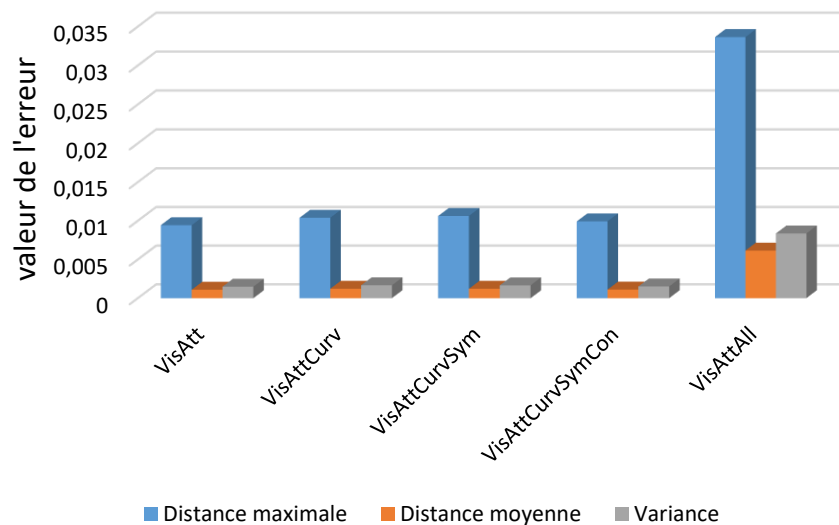


Figure 51- Influence de différentes combinaisons de canaux.

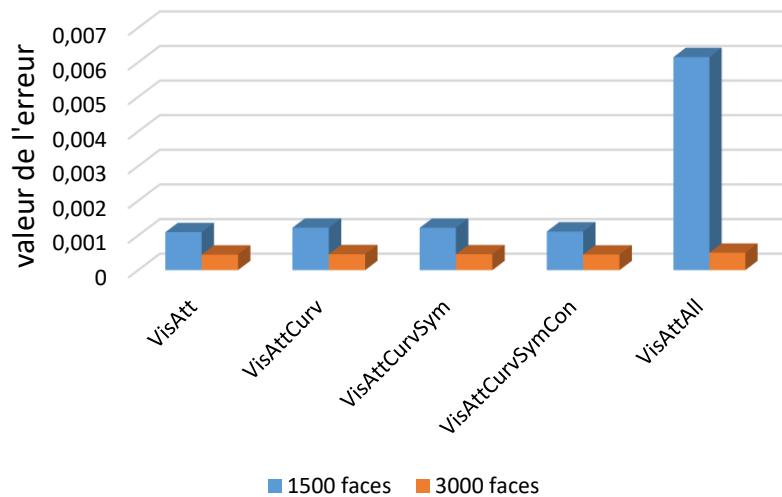


Figure 52- Influence de différentes combinaisons de canaux pour un nombre différent de faces.

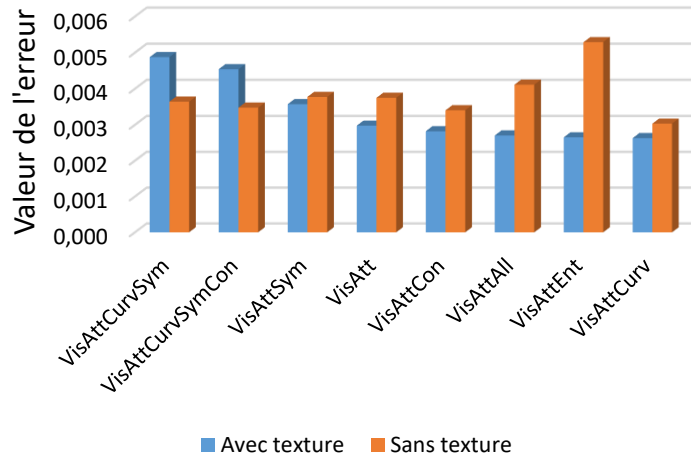
5.9. Influence de la texture

Les expérimentations réalisées dans le cadre de ce travail sont faites sur des modèles recouverts d'un matériau gris uniforme puisqu'aucune information de couleur n'est disponible pour les maillages dans la base de données utilisée. Cette base de données est utilisée afin de pouvoir comparer la méthode proposée avec des méthodes similaires pour la détection de points d'intérêt. Cependant, la méthode proposée fonctionne sans ajustement sur des maillages colorés et/ou texturés. La Figure 53 montre un exemple d'un maillage texturé et présente les résultats obtenus en comparant les taux d'erreurs pour l'objet avec ou sans la texture.

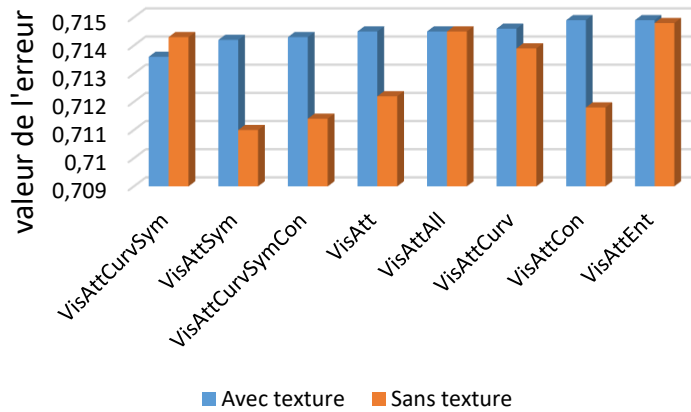
Comme il est possible de l'observer dans la Figure 53, en général, il n'y a pas de différence significative en termes d'erreur (maximum 0.01) entre l'utilisation de la texture ou non. L'erreur perceptuelle est calculée dans ce cas comme l'erreur moyenne normalisée basée sur SSIM et la distance dans le domaine laplacien normalisé. Une légère différence est attendue et est observable dans le cas des canaux de symétrie et de l'entropie. Dans le cas de la symétrie, la texture a un impact sur le calcul de ce canal puisque la texture n'est pas symétrique et cela mène à des erreurs légèrement plus élevées. Dans le cas de l'entropie, les erreurs résultantes sont plus basses lorsque la texture est utilisée, démontrant que l'utilisation de ce canal est bénéfique sur des objets texturés.



a)



b)



c)

Figure 53-(a) Maillage d'objet texturé et comparaison des taux d'erreur lorsque la texture est utilisée ou non; (b) moyenne des erreurs Metro; et (c) erreur perceptuelle.

5.10. Comparaison de différents détecteurs de points d'intérêt

La solution proposée, dans sa version classique et dans les versions avec canaux ajoutés, a aussi été comparée avec différents détecteurs de points d'intérêt provenant de la littérature

soit : la saillance du maillage (MS) [14], les points saillants (SP) [43], 3D-SIFT (3DS) [48], 3D Harris (3DH) [76], les coins dépendant de l'échelle (SDC) [49] et la signature thermique du noyau (HKS) [50] qui ont été intégrés dans l'algorithme de simplification décrit à la section 4.2. Les résultats sont montrés dans les Figures 58 à 62.

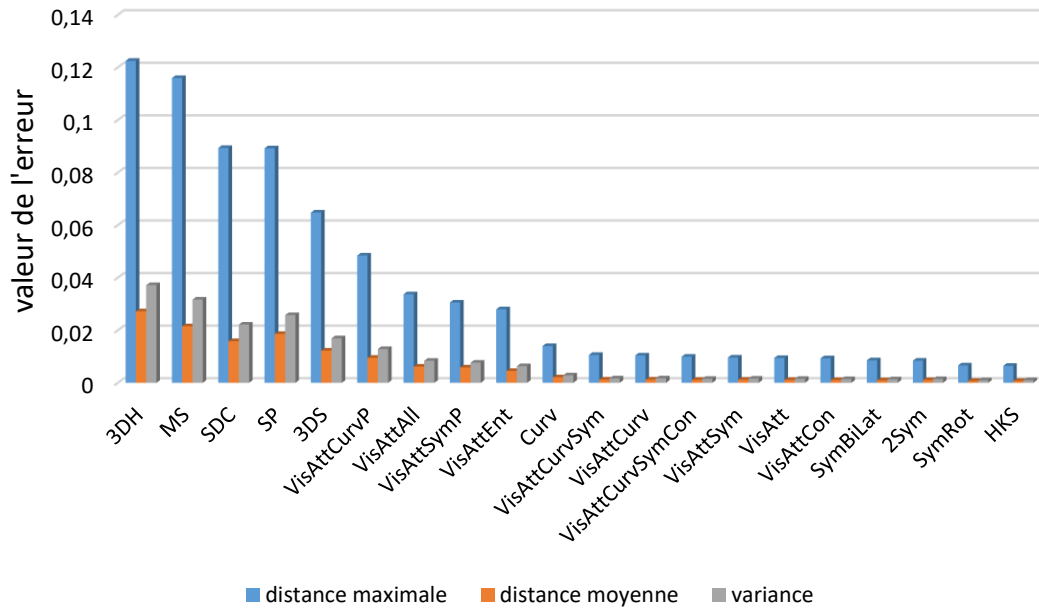


Figure 54- Comparaison de différentes solutions avec la méthode proposée pour différentes combinaisons de canaux.

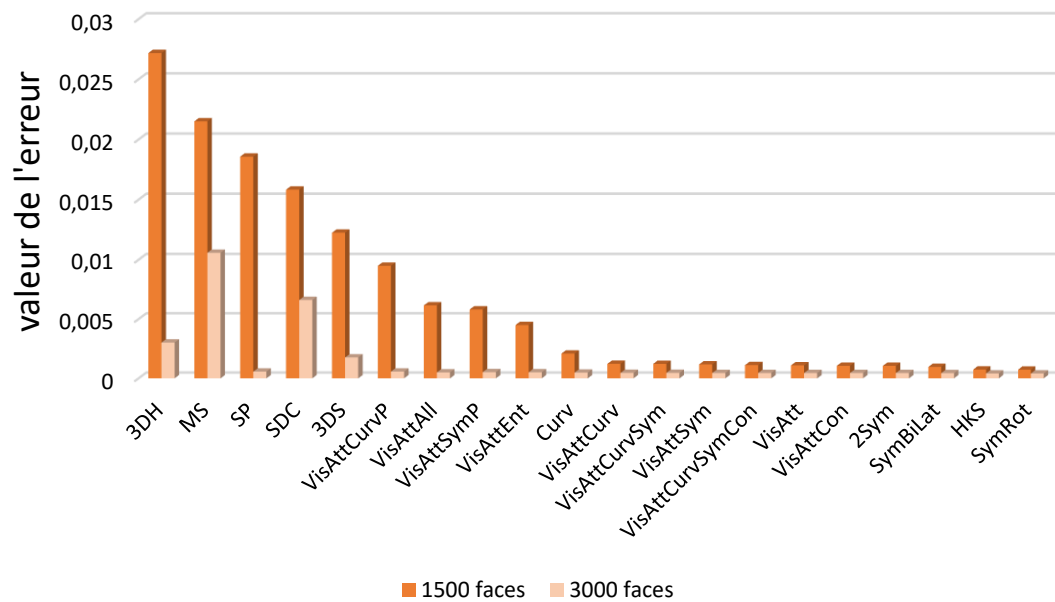


Figure 55- Comparaison de différentes méthodes avec la solution proposée pour un nombre de faces différent.

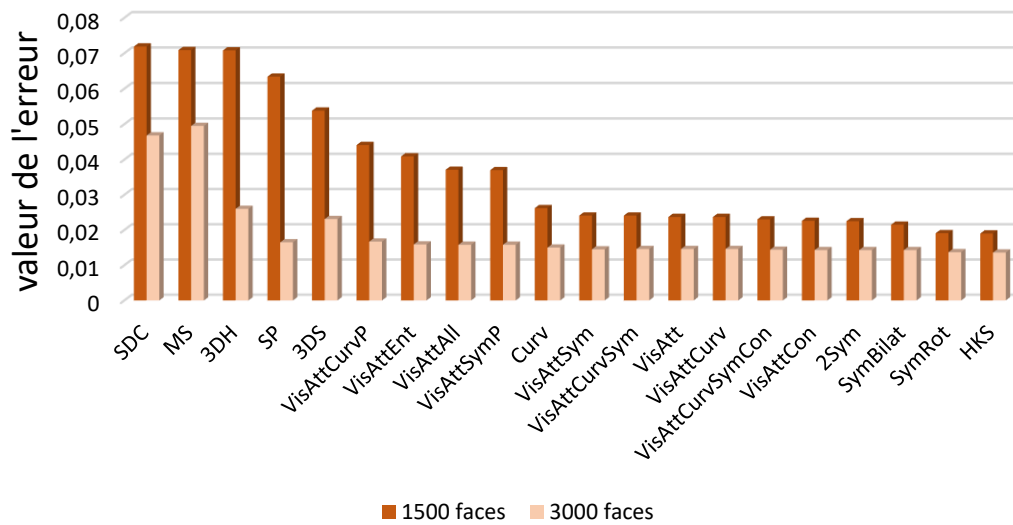


Figure 56- Erreurs perceptuelles basées sur la similarité pour un nombre de faces différent.

En observant les Figure 54 à 57, où les erreurs sont calculées comme une moyenne sur l'ensemble des objets de la base de données, on peut observer que toutes les solutions basées sur l'attention visuelle mènent en général à une meilleure performance pour une simplification avec densité sélective, à l'exception de la méthode HKS, qui obtient de meilleurs résultats. Une certaine corrélation existe entre le nombre de points d'intérêt, montré à la Figure 59, et les mesures d'erreur. Un plus haut nombre de points d'intérêt semble en général mener à des erreurs plus élevées.

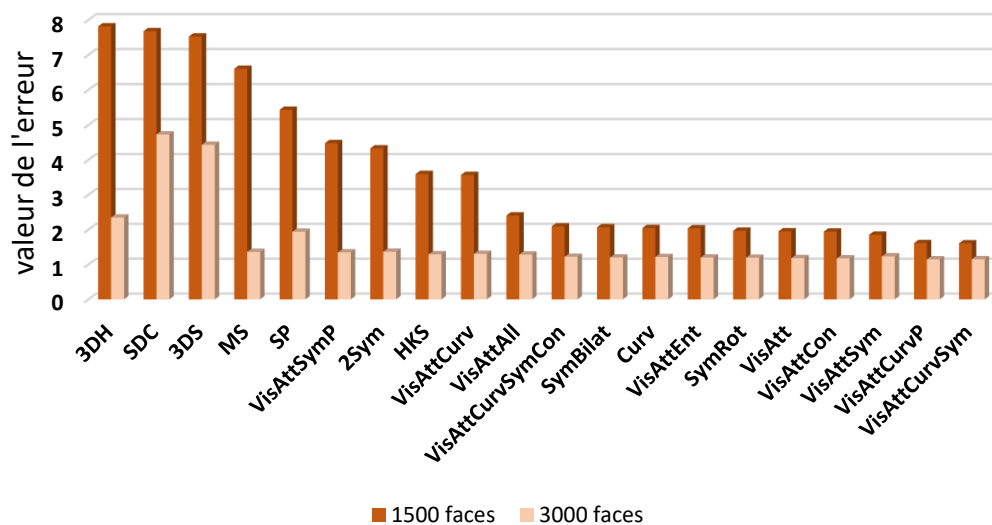


Figure 57- Distance dans le domaine laplacien pour un nombre différent de faces.

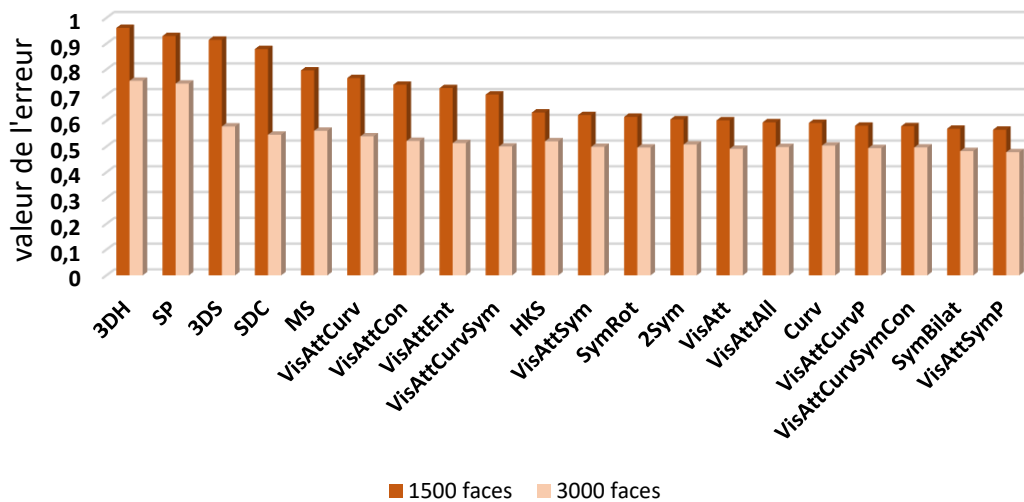


Figure 58- Distance dans le domaine laplacien normalisé.

Cela peut être expliqué par le fait que l’algorithme de simplification préserve les régions autour des points d’intérêt; par conséquent, un grand nombre de faces est nécessaire dans ces régions, ce qui en laisse un nombre réduit pour les régions simplifiées. Dans les réductions à de très basses résolutions, on pourrait alors observer des distorsions dans les régions simplifiées des maillages pour les méthodes qui retournent un très grand nombre de points d’intérêt, comme SDC et 3DH (voir Figure 67 pour un exemple de distorsion causée par un trop grand nombre de points d’intérêt).

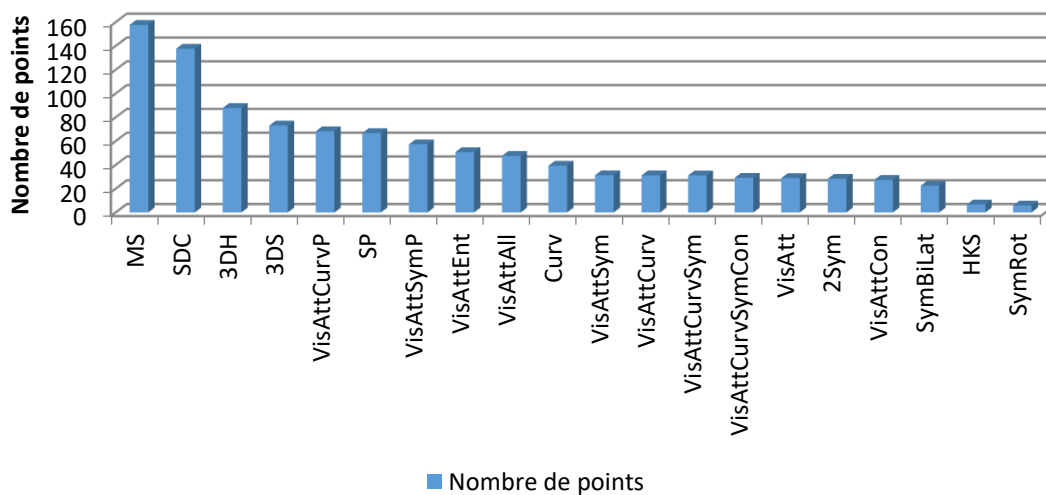


Figure 59- Nombre moyen de points d’intérêt retournés par différentes méthodes et par la solution proposée

Cependant, une réduction drastique du nombre de points d'intérêt ne mène pas nécessairement à une réduction drastique du taux d'erreur (par exemple pour la méthode HKS ou la méthode SymRad). Il est intéressant de remarquer que la méthode de l'attention visuelle avec courbure (VisAttCurvP) donne des erreurs plus faibles que la méthode SP, malgré un nombre de points d'intérêt similaire. Cela est dû à une meilleure distribution des points d'intérêt, comme on peut le voir à la Figure 61b et à la Figure 61n. On peut faire une dernière remarque au sujet de la Figure 60, qui démontre qu'un grand nombre de points d'intérêt ne mène pas nécessairement à de meilleurs résultats.

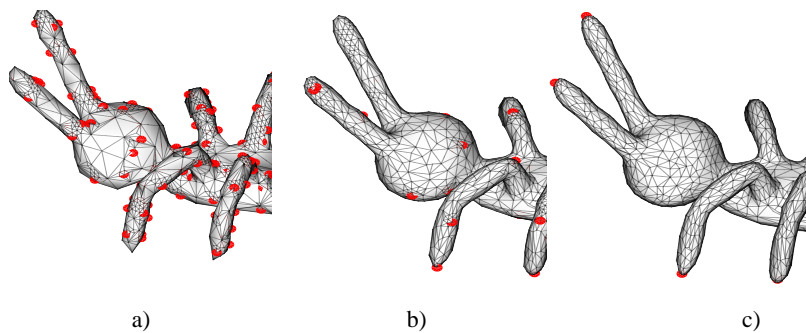
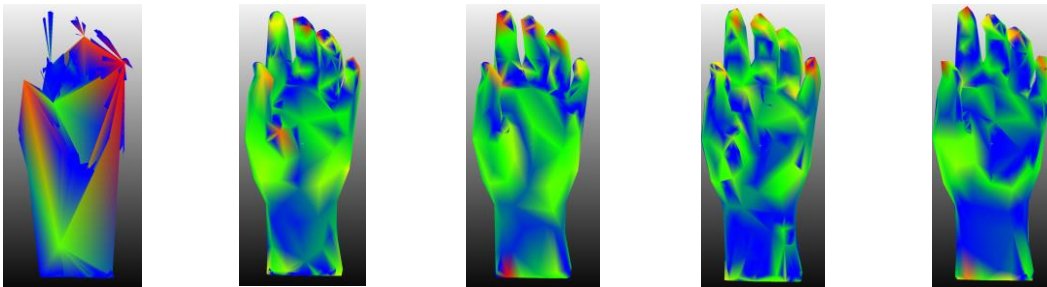


Figure 60- Résultats d'une simplification pour : (a) un grand nombre de points d'intérêt (SDC), (b) un nombre moyen de points d'intérêt (VisAttCurvSymCon) et (c) un petit nombre de points d'intérêt (HKS)

Cette figure compare les résultats d'une simplification pour 3 méthodes, soit une méthode retournant un important nombre de points d'intérêt (SDC), une méthode retournant un nombre moyen de points d'intérêt (VisAttCurvSymCon) et une méthode retournant un petit nombre de points d'intérêt (HKS). On peut observer que trop de points d'intérêt mènent à la création de groupes très denses de triangles sur le maillage. D'un autre côté, un trop petit nombre de points mènent à un modèle très près d'une simplification uniforme, faisant perdre l'utilité de la préservation des points d'intérêt, c'est-à-dire le détail dans les régions caractéristiques. Finalement, la Figure 61 montre, pour chacune des méthodes étudiées, les points d'intérêt détectés, alors que la Figure 62 représente l'évaluation Cloud Compare pour une réduction de 1500 faces du modèle de la main. Les régions en bleu représentent une bonne correspondance entre le modèle initial et le modèle réduit, les régions en vert représentent une correspondance moyenne et les régions en rouge représentent les zones les plus affectées par la simplification (où les erreurs sont les plus élevées).



Figure 61- Comparaison de différents détecteurs de points d'intérêt : (a) MS, (b) SP, (c) 3DH, (d) 3DS, (e) SDC, (f) HKS, (g) VisAtt, (h) VisAttCurv, (i) VisAttSym, (j) VisAttCon, (k) VisAttEnt, (l) VisAttAll, (m) VisAttCurvSym, (n) VisAttCurvP, (o) SymbiLat, (p) SymRot, (q) 2Sym, (r) VisAttSymP, (s) Curv et (t) VisAttCurvSymCon



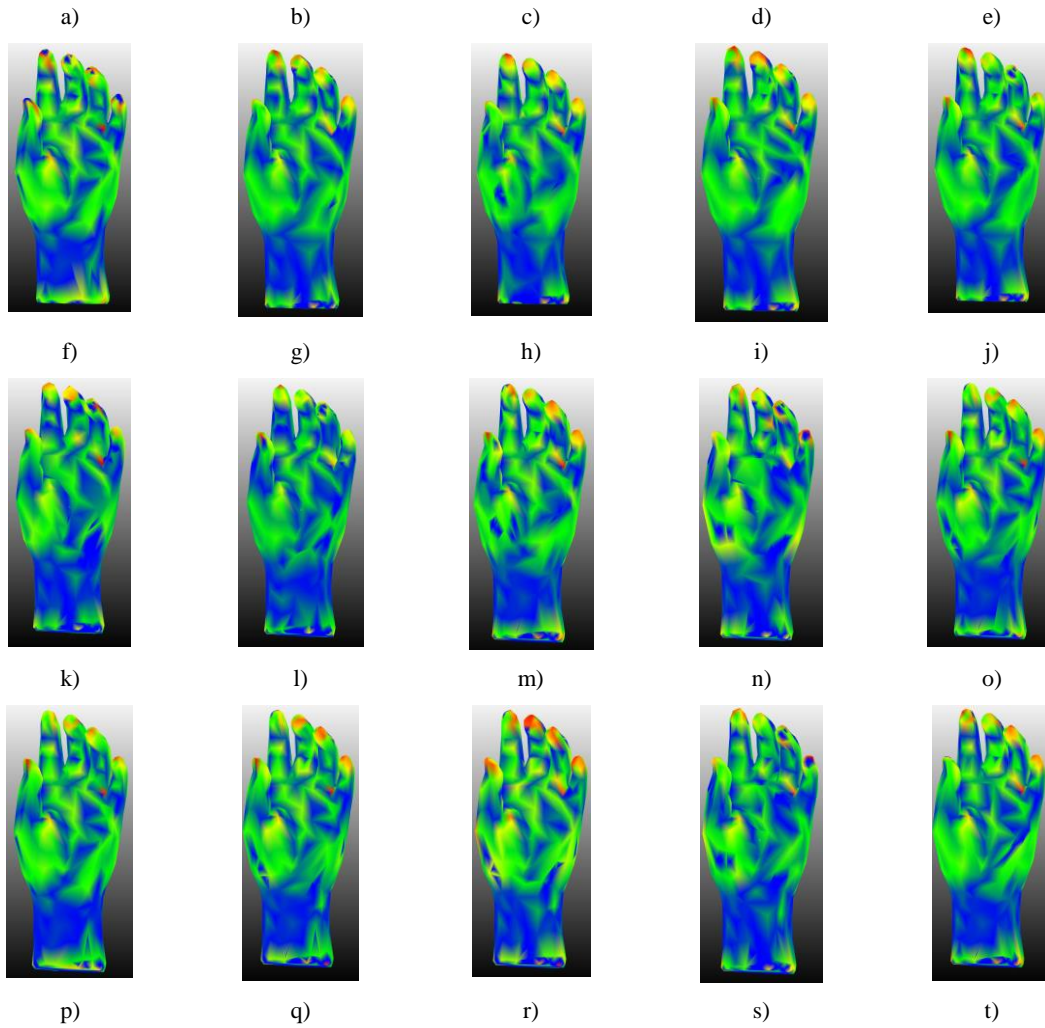


Figure 62- Résultats obtenus en utilisant Cloud Compare pour différents détecteurs de points d'intérêt pour le modèle de la main simplifié à 1500 faces : (a) MS, (b) SP, (c) 3DH, (d) 3DS, (e) SDC, (f) HKS, (g) VisAtt, (h) VisAttCurv, (i) VisAttSym, (j) VisAttCon, (k) VisAttEnt, (l) VisAttAll, (m) VisAttCurvSym, (n) VisAttCurvP, (o) SymBiLat, (p) SymRot, (q) 2Sym, (r) VisAttSymP, (s) Curv et (t) VisAttCurvSymCon

5.11. Évaluation avec la réalité de terrain

L'ensemble d'évaluation pour la réalité de terrain a été construit à partir des données fournies par [1], tel que décrit à la section 4.3.2. L'ensemble de données comporte 43 modèles d'objets, soit les mêmes modèles utilisés pour toutes les expérimentations effectuées dans le cadre de ce travail. L'ensemble de données comprend aussi les points d'intérêt marqués par 16 sujets. Pour la création de la réalité de terrain, le paramètre φ (le facteur de tolérance du rayon) a été laissé aux valeurs par défaut, soit entre 0 et 0,12 avec le pas 0,005.

Pour l'évaluation, deux valeurs du paramètre n_{sujets} ont été utilisées soit $n_{sujets} = 2$ (GT(n=2) à la Figure 68), c'est-à-dire qu'un point fera partie de la réalité de terrain s'il a été identifié par au moins 2 individus, et $n_{sujets} = 9$ (GT(n=9) à la Figure 68) où un point fera partie de la réalité de terrain s'il a été identifié par au moins 9 individus, soit plus de la moitié des participants. Il est important de préciser qu'on considère que deux individus ont marqué un même point si ceux-ci se trouve dans un voisinage de 0,05, soit la valeur par défaut. La Figure 63 présente dans la première colonne tous les points d'intérêt qui ont été marqués par au moins un participant. La deuxième colonne présente les points d'intérêt résultant de l'utilisation du paramètre $n_{sujets} = 2$ et finalement, la troisième colonne présente les points d'intérêt résultant de l'utilisation de $n_{sujets} = 9$. On peut observer que plus la valeur du paramètre n_{sujets} augmente, plus le nombre de points d'intérêt diminue puisqu'on s'attend à ce que plusieurs individus aient marqué les points pour les inclure. On peut également observer cette variation du nombre de points d'intérêt à la Figure 64 qui présente la moyenne (moyenne faite sur le nombre de points retournés pour chaque modèle) du nombre de points d'intérêt retournés par chaque méthode présentée dans ce mémoire ainsi que pour les points provenant de la réalité de terrain (GT(n=2) et GT(n=9)). Il est intéressant de noter que les méthodes VistAtt et GT(n=2) retournent un nombre de points similaire.

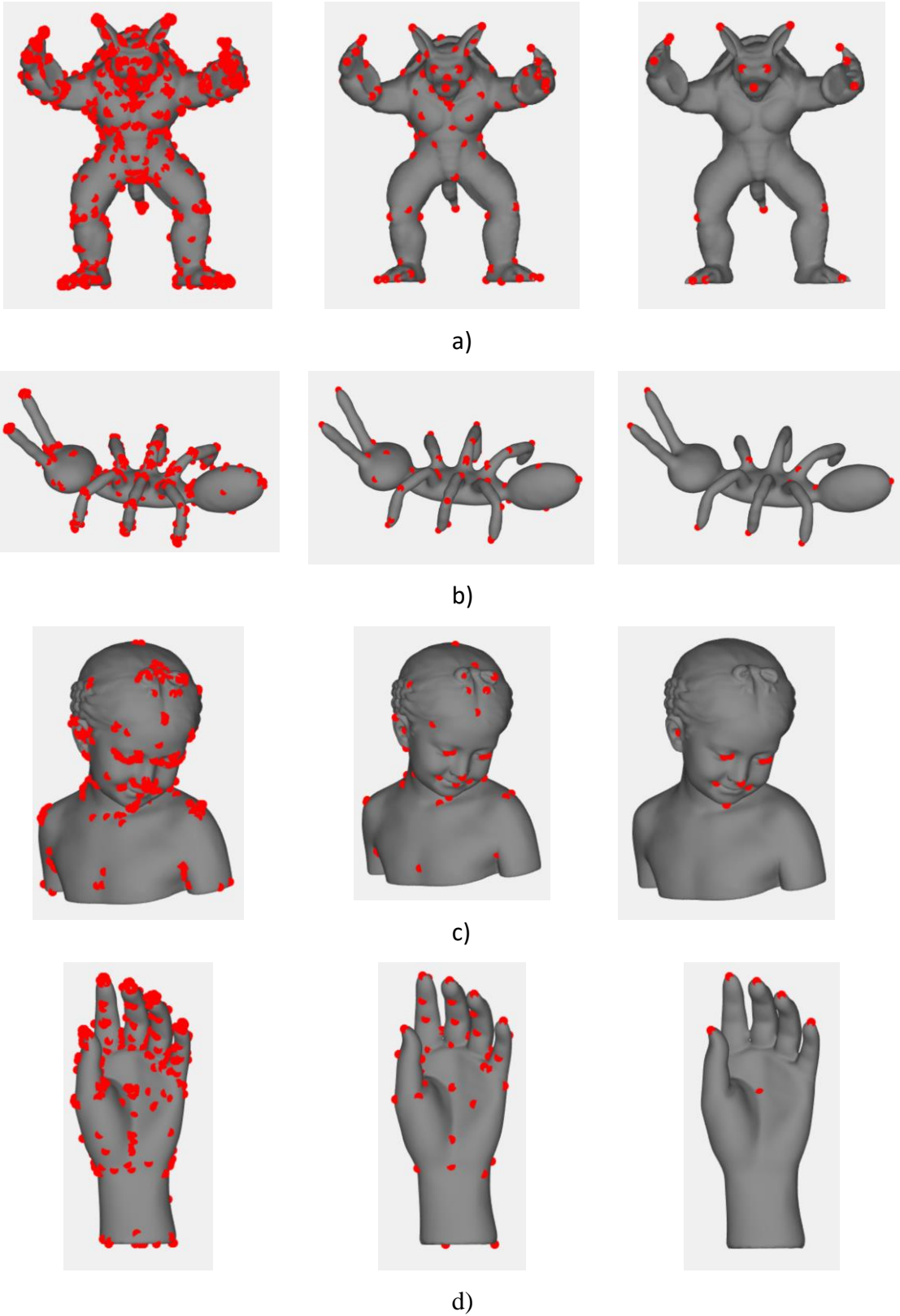


Figure 63- Première colonne : tous les points d'intérêt marqués par les participants, deuxième colonne : les points d'intérêt pour $n_{\text{sujets}} = 2$ et troisième colonne, les points d'intérêt pour $n_{\text{sujets}} = 9$, pour : (a) Armadillo, (b) fourmi, (c) fille et (d) main

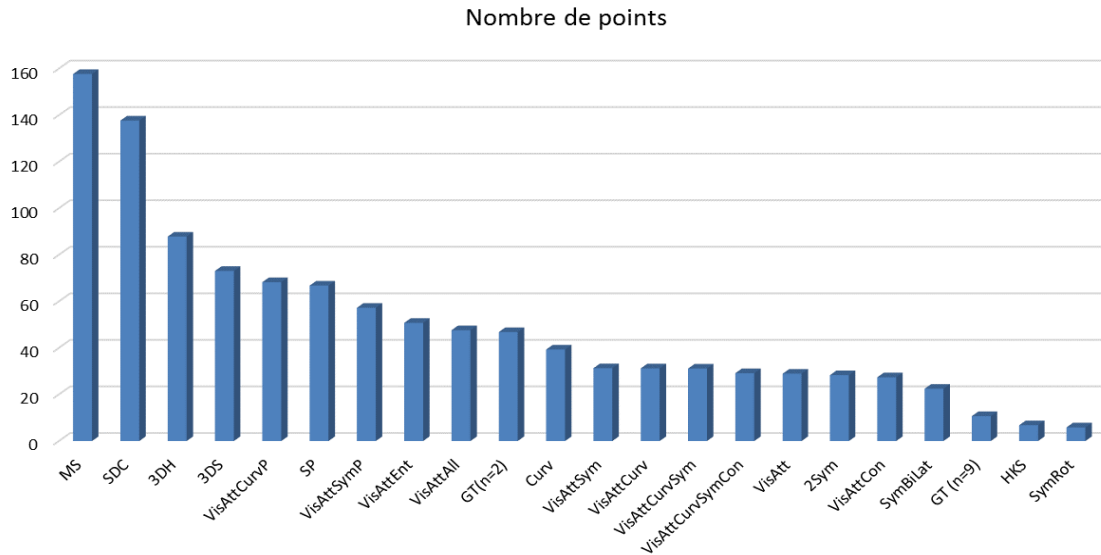
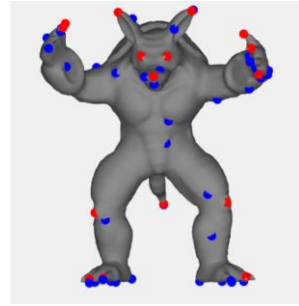
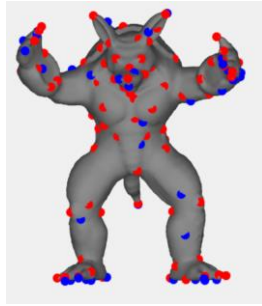
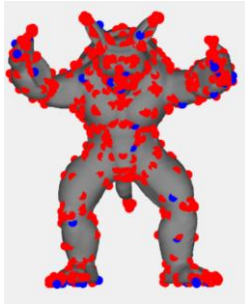


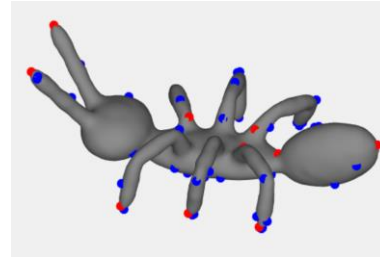
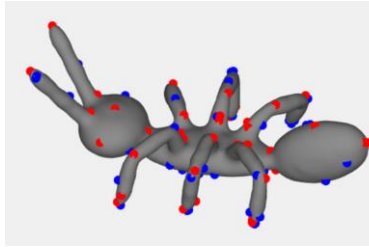
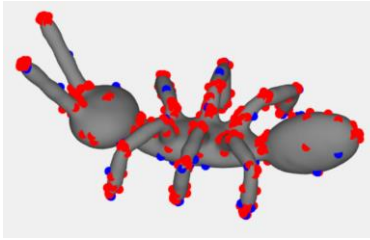
Figure 64- Moyenne du nombre de points d'intérêt retournés par chaque méthode

La Figure 65, la Figure 66 et la Figure 67 présentent une comparaison visuelle des points d'intérêt retournés pour la réalité de terrain (tous les participants, à gauche), GT (n=2) au centre et GT(n=9) à droite) et pour les méthodes VisAtt, VisAttCurvSymCon et VisAttAll respectivement. Il est possible de remarquer que les points d'intérêt retournés par les méthodes basées sur l'attention visuelle se situent en général près de points provenant de la réalité de terrain. Comme on peut l'observer dans la colonne de gauche, l'identification des points par l'humain est un processus extrêmement subjectif qui mène à un modèle presque entièrement recouvert de points. Il est évident que chaque individu utilise son propre schème dans l'identification des points.

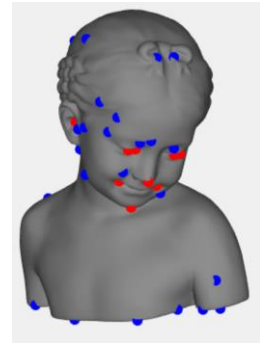
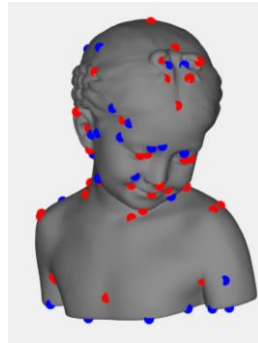
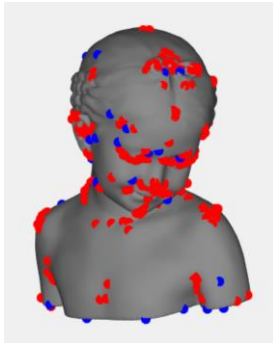
La présence de points provenant de l'attention visuelle dans le voisinage des points d'intérêt provenant de la réalité de terrain permet de remarquer qu'il serait peut-être mieux dans certains cas de ne pas considérer le centroïde (ou pas seulement) de la région saillante comme point d'intérêt. Cette avenue devra être explorée dans de futurs travaux.



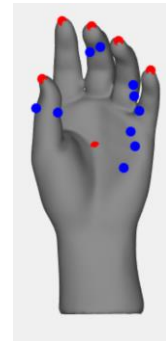
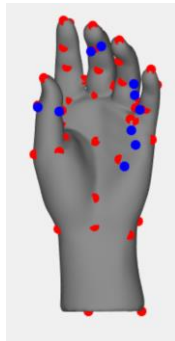
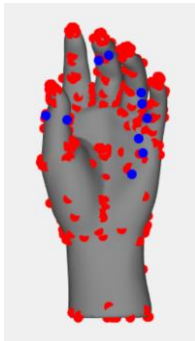
a)



b)

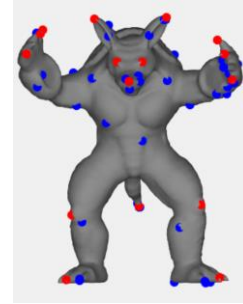
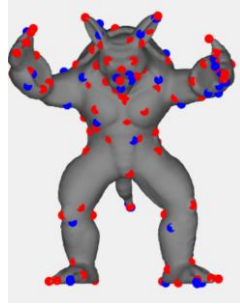
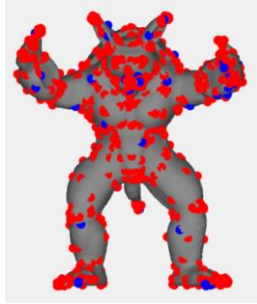


c)

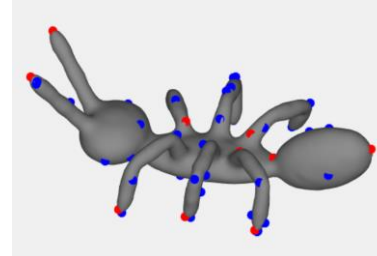
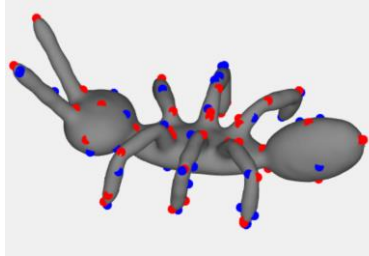
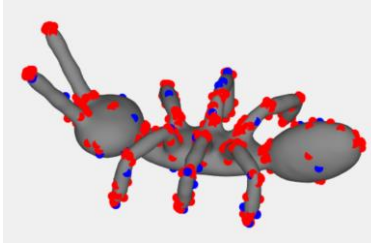


d)

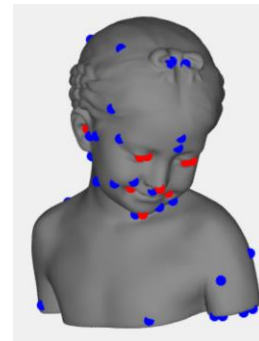
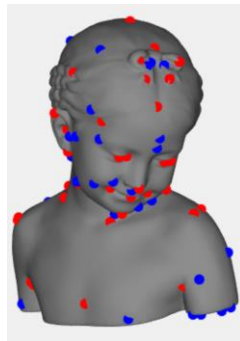
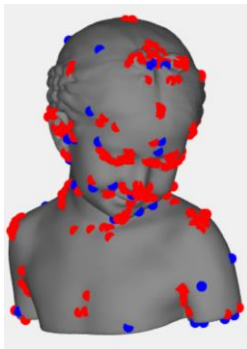
Figure 65- Comparaison des points d'intérêt pour le GT (rouge) et la méthode VisAtt (en bleu). Première colonne : tous les participants, deuxième colonne : $n_{\text{su jets}} = 2$, troisième colonne : $n_{\text{su jets}} = 9$, pour : (a) Armadillo, (b) fourmi, (c) fille et (d) main.



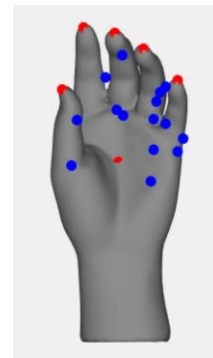
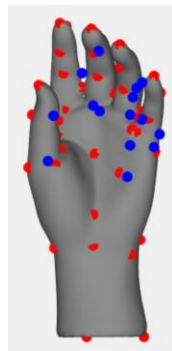
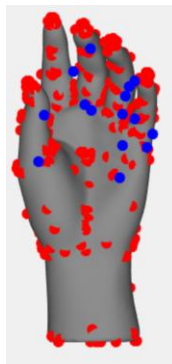
a)



b)

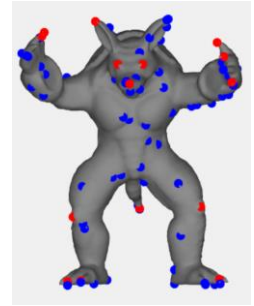
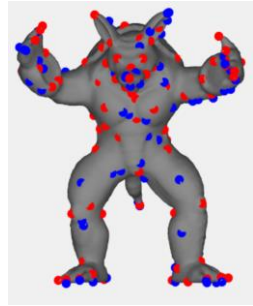
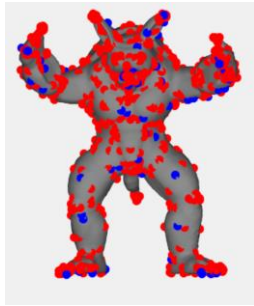


c)

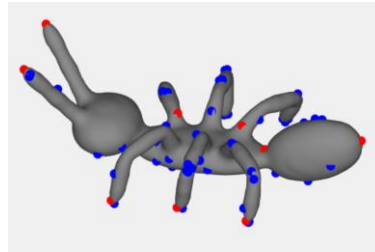
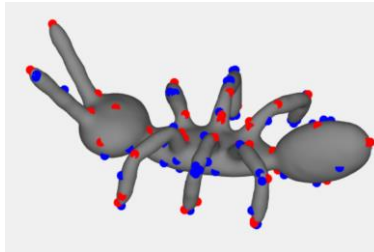
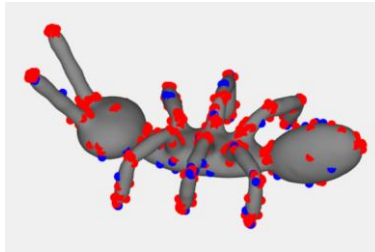


d)

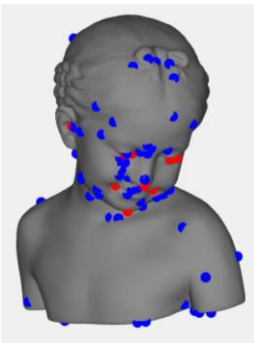
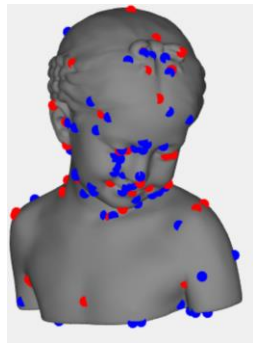
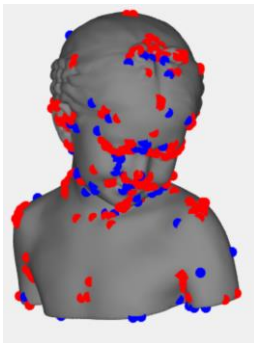
Figure 66- Comparaison des points d'intérêt pour le GT (rouge) et la méthode VisAttCurvSymCon (en bleu). Première colonne : tous les participants, deuxième colonne : $n_{\text{su jets}} = 2$, troisième colonne : $n_{\text{su jets}} = 9$, pour : (a) Armadillo, (b) fourmi, (c) fille et (d) main.



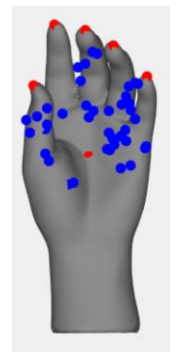
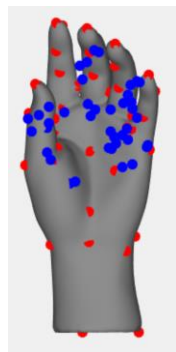
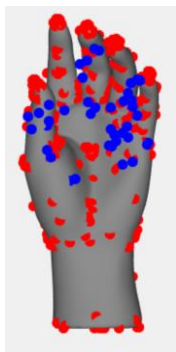
a)



b)



c)



d)

Figure 67- Comparaison des points d'intérêt pour le GT (rouge) et la méthode VisAttAll (en bleu). Première colonne : $n_{\text{Sujets}} = 2$, deuxième colonne : $n_{\text{Sujets}} = 9$, pour : (a) Armadillo, (b) fourmi, (c) fille, et (d) main La troisième colonne est quoi?

Finalement, les points d'intérêt trouvés à l'aide de la réalité de terrain ont été insérés dans l'algorithme de simplification de la même manière que toutes les autres méthodes mentionnées jusqu'à maintenant. Les résultats sont présentés à la Figure 68. On peut observer que pour la plupart des méthodes basées sur l'attention visuelle, les erreurs se situent entre les deux formes de la réalité de terrain (GT(n=2) et GT(n=9)). Pour toutes les méthodes dérivées de la littérature sauf HKS, les erreurs sont plus grandes que celle obtenues en se basant sur les points de la vérité terrain. Pour un nombre de points presque identiques, la méthode VisAttAll obtient des erreurs légèrement plus élevée que GT(n=2) (de l'ordre de 0.003). Les méthodes VisAtt et VisAttCurvSymCon obtiennent des erreurs plus faibles que GT(n=2). Les faibles erreurs obtenues par GT(n=9) peuvent être expliquées par le petit nombre de points que cette méthode retourne, comme c'est le cas pour HKS.

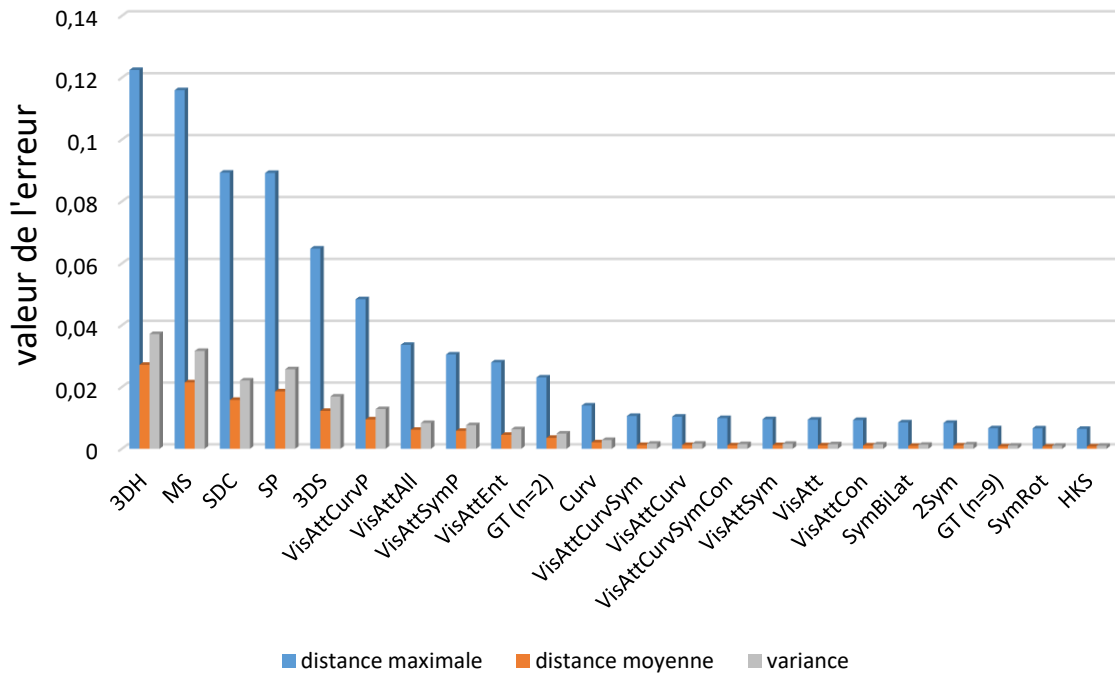


Figure 68- Erreurs Metro pour un nombre de faces de 1500 et un voisinage de n=3

5.12. Compacité des modèles sélectivement simplifiés

Des tests ont aussi été effectués pour démontrer le degré de compacité pouvant être atteint en employant la méthode proposée pour la simplification des objets avec la préservation des régions d'intérêt. Le degré de compacité moyen pouvant être atteint à l'aide de la méthode proposée est de 91,5 % du nombre de faces initial en utilisant un maillage réduit à 1500 faces et 83 % en utilisant un maillage réduit à 3000 faces. Il faut préciser que certaines méthodes produisant un grand nombre de points d'intérêt, comme MS ou SDC, ne peuvent employer un nombre de faces inférieur à 4500 sans provoquer une distorsion dans le maillage, comme il est possible de le voir à la Figure 69; elles peuvent donc effectuer une réduction maximale de 75% du nombre de faces.

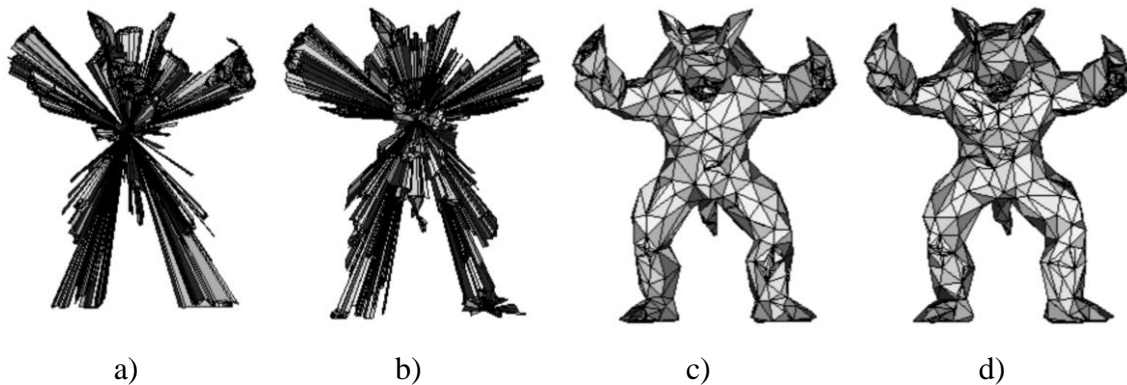


Figure 69 - Modèle de l'Armadillo réduit à 1500 faces en préservant les points d'intérêt pour la méthode (a) MS, (b) SDC, (c) VisAtt, et (d) VisAttCurvSymCon

5.13. Incorporation dans une méthode discrète à multiples niveaux de détails

La Figure 70 montre les modèles de l'Armadillo à différents niveaux de détails créés automatiquement par la méthode proposée. Les nombres de faces choisis pour ce modèle par le réseau de neurones, tel qu'expliqué à la section 4.4, sont les suivants : 925, 1890, 4630, 6219 et 8600 respectivement. Même aux résolutions les plus basses, les détails du modèle comme les oreilles, les pattes ou la queue sont préservés.

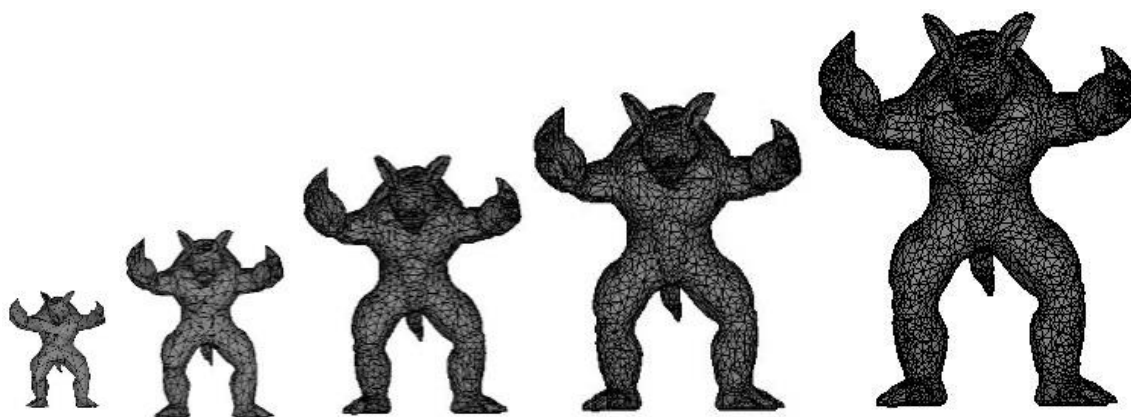


Figure 70- Modèle discret de l'objet à différents niveaux de détails en utilisant la méthode d'identification de points d'intérêt basée sur l'attention visuelle

5.14. Incorporation dans une méthode continue à multiples niveaux de détails

Afin de démontrer l'utilisation des modèles simplifiés dans une modélisation continue à multiples niveaux de détails, la Figure 71 montre un exemple de différents modèles créés par la méthode proposée telle que décrite à la section 4.5. La deuxième ligne de la Figure 71 montre la distribution des erreurs à la surface de l'objet tel que démontré par l'outil Cloud Compare, où les erreurs faibles apparaissent en vert, les erreurs moyennes en jaune et les erreurs importantes en rouge alors que les zones en bleu représentent une adéquation parfaite avec le maillage initial. Les points d'intérêt sont démontrés en rouge sur le maillage. Il est possible de remarquer que même à la plus basse résolution, soit 950 faces dans ce cas, les détails du modèle tels que les oreilles ou les ailes sont préservés.

5.15. Temps de calcul

En termes de temps de calcul, la procédure complète, de la prise d'image à la visualisation des maillages simplifiés, prend en moyenne 120.4 secondes par objet, soit approximativement 69.1 secondes par 10 000 faces. En général, environ 30 % du temps est dédié à la capture des images de prise de vue et 70 % à l'identification des points d'intérêt. La simplification prend au maximum 0,2 seconde par objet. Il faut aussi préciser que tous les calculs, tel qu'expliqué à la section 4, se font hors ligne. Une fois le réseau de neurones entraîné, il produit une estimation du nombre de faces en 0,03 seconde et le temps de

simplification est de 0,2 seconde, ce qui donne un temps total de 0,23 seconde. Les expérimentations ont été réalisées à l'aide de code Matlab sur un ordinateur Intel Core i7 2.2GHz avec 8GB de mémoire.

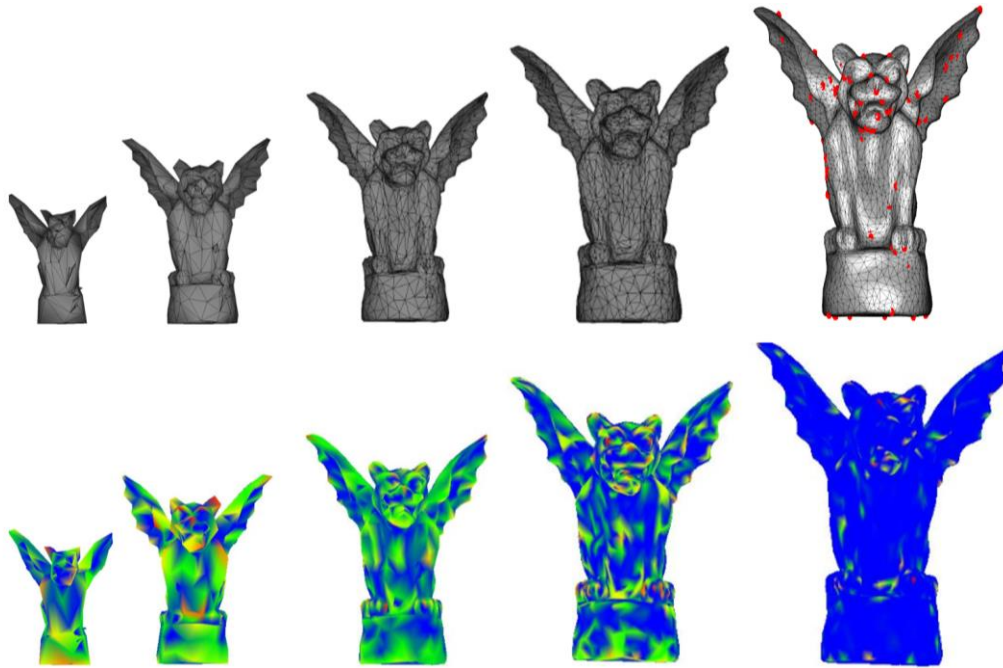


Figure 71- Modèle de l'objet et les erreurs en code de couleur Cloud Compare à différents niveaux de détails en utilisant l'identification des points d'intérêt basée sur l'attention visuelle humaine (méthode VisAttCurvSymCon)

5.16. Généralisation de la méthode proposée pour d'autres objets 3D

Les expérimentations réalisées ont démontré qu'en général, pour la base de données utilisée, la méthode proposée utilisant la combinaison de la courbure, de la symétrie et du contraste (VisAttCurvSymCon) offre le meilleur compromis entre les erreurs quantitatives (Metro et erreurs perceptuelles) et la qualité de la simplification finale. Cependant, il est important de préciser que la nature de l'objet à simplifier et ses caractéristiques pourraient faire en sorte que certains canaux n'auraient pas la même importance que dans le cas de la base de données utilisée dans le cadre de ce travail. Particulièrement, lorsqu'un objet avec texture est testé, des légères différences sont attendues dans les canaux de couleurs, de symétrie et d'entropie. Par conséquent, il est suggéré de tester toutes les combinaisons de canaux, en utilisant les 4 points de vue identifiés et un voisinage de 3 autour des points d'intérêt lors de la simplification, avant d'utiliser le cadre proposé sur une autre base de

données. Les informations pour guider la sélection d'un voisinage sont données à la section 5.2. Les informations provenant de chaque canal de saillance doivent être intégrées directement dans le calcul de la carte de saillance basée sur l'attention visuelle tel que décrit dans le chapitre 4. Les tests réalisés ont montré qu'il ne vaut pas la peine de considérer séparément les points d'intérêt détectés par les différents canaux tels que la courbure ou la symétrie et on s'attend à ce que ce soit le cas peu importe les caractéristiques de l'objet.

Une fois que les différentes combinaisons sont construites, l'utilisateur devrait procéder aux calculs d'erreur pour un nombre variables de faces et, en fonction de l'erreur tolérée pour l'application en question, il pourra retenir une ou plusieurs combinaisons comme solution. En termes de nombre de faces, il est suggéré de commencer avec environ 15% du nombre total de faces du maillage initial et d'augmenter graduellement jusqu'à environ la moitié du nombre de faces du maillage initial, selon les besoins de l'application. Si plusieurs solutions sont retenues en fonction de l'erreur, une comparaison visuelle peut être effectuée avec un outil tel que Cloud Compare pour permettre de sélectionner la meilleure solution.

Il est aussi possible de comparer les résultats avec la réalité de terrain. Cependant, il est conseillé de ne pas se fier à ces mesures en premier lieu, en raison des lacunes exprimées à la section 0. Si l'utilisateur désire intégrer les résultats dans un cadre de modélisation à multiples niveaux de détails, il peut procéder à l'entraînement de réseaux de neurones tel que décrit à la section 4.5. Si une seule meilleure solution est identifiée, un seul réseau de neurones suffira. Si plusieurs solutions sont retenues en fonction de l'erreur, l'utilisateur peut procéder à l'entraînement des séries de réseaux. Finalement, l'algorithme de simplification sélectivement densifié peut être appliqué pour obtenir les différentes versions des modèles simplifiés.

6. Conclusion

L'objectif de ce projet était d'évaluer l'impact de la détection des points d'intérêt basée sur l'attention visuelle humaine dans le contexte de la modélisation d'objets 3D à multiples niveaux de détails pour des applications dans les environnements virtuels. L'influence de différentes combinaisons de canaux de caractéristiques pour améliorer le modèle computationnel de l'attention visuelle a été étudiée expérimentalement et la supériorité de la méthode proposée en termes de qualité est démontrée expérimentalement en comparant la méthode avec différents détecteurs de points d'intérêt provenant de la littérature. De plus, la méthode a été comparée avec les résultats provenant de la réalité de terrain disponible dans la littérature. Bien que les résultats obtenus dans cette section ne soient pas tels qu'attendu, certaines lacunes dans la construction de la réalité de terrain ont été identifiées ce qui suggère que des travaux futurs devront porter sur cet aspect de l'évaluation. Il a aussi été démontré que la méthode peut s'utiliser sur des modèles en couleur ou texturés et ce, sans avoir à lui apporter de changement.

Finalement, ce travail a permis de démontrer l'importance de considérer les capacités visuelles humaine dans la modélisation 3D à multiples niveaux de détails afin d'améliorer la perception des utilisateurs.

En dernier lieu, il est important de préciser que certains aspects de la méthode proposée pourront faire l'objet d'autres expérimentations dans de futurs travaux. Entre autres, il sera intéressant de tester la méthode proposée sur des objets provenant d'autres bases de données, particulièrement sur des objets colorés ou texturés. Aussi, il pourra être possible de vérifier l'effet de l'ajout d'autres canaux de saillance dans le modèle proposé ou encore d'utiliser une pondération différente pour les différents canaux dans le calcul de la carte de saillance. Par exemple, dans le cas des objets texturés, il pourrait être utile de donner un poids plus élevé au canal d'entropie. Il serait en outre intéressant de procéder à un apprentissage par type d'objet.

Une partie des travaux futurs devrait aussi porter sur le choix du voisinage adaptatif. La méthode proposée dans ce mémoire pour la sélection de la taille du voisinage considère une moyenne de la taille des régions d'intérêt identifiées par la saillance sur la totalité de la taille de l'objet. La taille du voisinage sélectionnée est ensuite utilisée autour de chaque

point d'intérêt sur un même objet lors de la simplification. Une avenue intéressante serait de choisir un voisinage adaptatif à chaque point d'intérêt en fonction de la taille de la région d'intérêt où se situe ce point. En effet, comme un même objet peut comporter des régions très petites à très grandes, une telle façon de sélectionner le voisinage permettrait de mieux préserver les régions saillantes dans leur taille originale telle que détectée par la carte de saillance.

Un intérêt pourra aussi être porté sur la sélection des meilleurs points de vue. Par exemple, il pourrait être intéressant de considérer le focus et les saccades. Les régions d'intérêt pour tous les points de vue pourraient être calculés pour les réorganiser dans un ordre décroissant de saillance. Il serait ensuite possible de sélectionner le meilleur point de vue en étant ce qui comporte le plus grand nombre de régions avec un saillance élevée.

Aussi, des travaux devraient porter sur la réalisation d'une nouvelle réalité de terrain. Comme mentionné, la méthode proposée dans la littérature comporte quelques lacunes dans la façon dont sont choisis les points d'intérêt faisant partie de la réalité de terrain et aussi par le petit nombre de participants (16) ayant identifiés les points d'intérêt sur les objets. Il serait intéressant de construire une nouvelle base de données avec un nombre significatif de participants et en établissant une procédure fiable pour la sélection des points d'intérêt sur chaque objet.

Finalement, des tests pourront être menés en utilisant un visiocasque ou directement dans la CAVE (*Chicago Automatic Virtual Environment*), soit la voute immersive à six faces se trouvant au laboratoire de cyberpsychologie de l'Université du Québec en Outaouais pour évaluer de manière plus poussée la qualité des modèles obtenus. En effet, comme l'objectif principal de ce travail concerne l'amélioration de la perception par l'utilisateur, la prochaine étape serait donc d'inviter des participants à faire l'évaluation des maillages simplifiés afin de pouvoir en comparer la qualité finale en termes de perception. Pour se faire, les modèles utilisés devront être transposés dans Unity3D.

7. Références

- [1] H. Dutagaci, C. P. Cheung, and A. Godil, "Evaluation of 3D interest point detection techniques via human-generated ground truth," *The Visual Computer*, vol. 28, pp. 901-917, 2012.
- [2] T. C. Kietzmann, S. Lange, and M. Riedmiller, "Computational object recognition: a biologically motivated approach," *Biological cybernetics*, vol. 100, pp. 59-79, 2009.
- [3] D. P. Luebke, "A developer's survey of polygonal simplification algorithms," *Computer Graphics and Applications, IEEE*, vol. 21, pp. 24-35, 2001.
- [4] D. Walther, L. Itti, M. Riesenhuber, T. Poggio, and C. Koch, "Attentional selection for object recognition—a gentle way," in *Biologically Motivated Computer Vision*, 2002, pp. 472-479.
- [5] S. Frintrop and P. Jensfelt, "Attentional landmarks and active gaze control for visual SLAM," *Robotics, IEEE Transactions on*, vol. 24, pp. 1054-1065, 2008.
- [6] A. Cretu and P. Payeur, "Biologically-inspired visual attention features for a vehicle classification task," *International Journal on Smart Sensing and Intelligent Systems*, vol. 4, pp. 402-423, 2011.
- [7] J. Han, *3D graphics for game programming*: CRC Press, 2011.
- [8] A. Chopine, *3D Art Essentials: The Fundamentals of 3D Modeling, Texturing, and Animation*, 1 ed.: Focal Press, 2011.
- [9] D. P. Luebke, M. Reddy, J. D. Cohen, A. Varshney, B. Watson, and R. Huebner, *Level of detail for 3D graphics*: Morgan Kaufmann, 2003.
- [10] Y. Kho and M. Garland, "User-guided simplification," in *Proceedings of the 2003 symposium on Interactive 3D graphics*, 2003, pp. 123-126.
- [11] T.-C. Ho, Y.-C. Lin, J.-H. Chuang, C.-H. Peng, and Y.-J. Cheng, "User-assisted mesh simplification," in *Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications*, 2006, pp. 59-66.
- [12] E. Pojar and D. Schmalstieg, "User-controlled creation of multiresolution meshes," in *Proceedings of the 2003 symposium on Interactive 3D graphics*, 2003, pp. 127-130.
- [13] R. Song, Y. Liu, Y. Zhao, R. R. Martin, and P. L. Rosin, "Conditional random field-based mesh saliency," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, 2012, pp. 637-640.
- [14] C. H. Lee, A. Varshney, and D. W. Jacobs, "Mesh saliency," in *ACM transactions on graphics (TOG)*, 2005, pp. 659-666.

- [15] L. Itti and C. Koch, "Feature combination strategies for saliency-based visual attention systems," *Journal of Electronic Imaging*, vol. 10, pp. 161-169, 2001.
- [16] S. Frintrop, E. Rome, and H. I. Christensen, "Computational visual attention systems and their cognitive foundations: A survey," *ACM Transactions on Applied Perception (TAP)*, vol. 7, p. 6, 2010.
- [17] V. Navalpakkam and L. Itti, "Modeling the influence of task on attention," *Vision research*, vol. 45, pp. 205-231, 2005.
- [18] J. M. Wolfe and T. S. Horowitz, "What attributes guide the deployment of visual attention and how do they do it?," *Nature reviews neuroscience*, vol. 5, pp. 495-501, 2004.
- [19] P. J. Locher and C. Nodine, "Symmetry catches the eye," *Eye movements: From physiology to cognition*, pp. 353-361, 1987.
- [20] C. Siagian and L. Itti, "Rapid biologically-inspired scene classification using features shared with visual attention," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, pp. 300-312, 2007.
- [21] Y. Hu, X. Xie, W.-Y. Ma, L.-T. Chia, and D. Rajan, "Salient region detection using weighted feature maps based on the human visual attention model," in *Advances in Multimedia Information Processing-PCM 2004*, ed: Springer, 2004, pp. 993-1000.
- [22] S. Frintrop, *VOCUS: a visual attention system for object detection and goal-directed search* vol. 3899: Springer, 2006.
- [23] T. Michalke, J. Fritsch, and C. Goerick, "Enhancing robustness of a saliency-based attention system for driver assistance," in *Computer Vision Systems*, ed: Springer, 2008, pp. 43-55.
- [24] G. Backer, B. Mertsching, and M. Bollmann, "Data- and model-driven gaze control for an active-vision system," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 1415-1429, 2001.
- [25] F. H. Hamker and J. Worcester, "Object detection in natural scenes by feedback," in *Biologically motivated computer vision*, 2002, pp. 398-407.
- [26] M. Z. Aziz and B. Mertsching, "Fast and Robust Generation of Feature Maps for Region-Based Visual Attention," *IEEE Transactions on Image Processing*, vol. 17, pp. 633-644, 2008.
- [27] N. Ouerhani and H. Hügli, "Computing visual attention from scene depth," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, 2000, pp. 375-378.

- [28] S.-H. Lee, J. Moon, and M. Lee, "A region of interest based image segmentation method using a biologically motivated selective attention model," in *Neural Networks, 2006. IJCNN'06. International Joint Conference on*, 2006, pp. 1413-1420.
- [29] L. Wei, N. Sang, and Y. Wang, "A biologically inspired object-based visual attention model," *Artificial Intelligence Review*, vol. 34, pp. 109-119, 2010.
- [30] Y. Hu, D. Rajan, and L.-T. Chi, "Adaptive local context suppression of multiple cues for salient visual attention detection," in *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, 2005, p. 4 pp.
- [31] S.-B. Choi, B.-S. Jung, S.-W. Ban, H. Niitsuma, and M. Lee, "Biologically motivated vergence control system using human-like selective attention model," *Neurocomputing*, vol. 69, pp. 537-558, 2006.
- [32] X. Liu, P. Tao, J. Cao, H. Chen, and C. Zou, "Mesh saliency detection via double absorbing Markov chain in feature space," *Visual Computer*, pp. 1-12, 2015.
- [33] P. Castelló, M. Chover, M. Sbert, and M. Feixas, "Reducing complexity in polygonal meshes with view-based saliency," *Computer Aided Geometric Design*, vol. 31, pp. 279-293, 2014.
- [34] Y. Dong, M. T. Pourasad, and P. Nasiopoulos, "Human Visual System Based Saliency Detection for High Dynamic Range Content," *Multimedia, IEEE Transactions on*, vol. PP, pp. 1-1, 2016.
- [35] L. Zhong, C. Weihai, and W. Xingming, "Salient region detection using high level feature," in *Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on*, 2014, pp. 1213-1217.
- [36] I. Iatsun, M. C. Larabi, and C. Fernandez-Maloigne, "Visual attention modeling for 3D video using neural networks," in *3D Imaging (IC3D), 2014 International Conference on*, 2014, pp. 1-8.
- [37] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 1254-1259, 1998.
- [38] R. Achanta, and S. Strunk, "Saliency detection using maximum symmetric surround," in *2010 IEEE International Conference on Image Processing*, 2010, pp. 2653-2656.
- [39] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, 2007, pp. 1-8.

- [40] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in *Proceedings of the 14th annual ACM international conference on Multimedia*, 2006, pp. 815-824.
- [41] M.-M. Cheng, G.-X. Zhang, N. J. Mitra, X. Huang, and S.-M. Hu, "Global contrast based salient region detection," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011, pp. 409-416.
- [42] Y.-S. Liu, M. Liu, D. Kihara, and K. Ramani, "Salient critical points for meshes," in *Proceedings of the 2007 ACM symposium on Solid and physical modeling*, 2007, pp. 277-282.
- [43] U. Castellani, M. Cristani, S. Fantoni, and V. Murino, "Sparse points matching by combining 3D mesh saliency with statistical descriptors," in *Computer Graphics Forum*, 2008, pp. 643-652.
- [44] X. Lianping, Z. Xiaoting, C. C. L. Wang, and H. Kin-Chuen, "Highly Parallel Algorithms for Visual-Perception-Guided Surface Remeshing," *Computer Graphics and Applications, IEEE*, vol. 34, pp. 52-64, 2014.
- [45] C. Creusot, N. Pears, and J. Austin, "A machine-learning approach to keypoint detection and landmarking on 3D meshes," *International Journal of Computer Vision*, vol. 102, pp. 146-179, 2013.
- [46] M. Feixas, M. Sbert, and F. González, "A unified information-theoretic framework for viewpoint selection and mesh saliency," *ACM Transactions on Applied Perception (TAP)*, vol. 6, p. 1, 2009.
- [47] I. Sipiran, B. Bustos, I. Sipiran, and B. Bustos, "Harris 3D: a robust extension of the Harris operator for interest point detection on 3D meshes," 2011.
- [48] A. Godil and A. I. Wagan, "Salient local 3d features for 3d shape retrieval," in *IS&T/SPIE Electronic Imaging*, 2011, pp. 78640S-78640S-8.
- [49] J. Novatnack and K. Nishino, "Scale-dependent 3D geometric features," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 2007, pp. 1-8.
- [50] J. Sun, M. Ovsjanikov, and L. Guibas, "A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion," in *Computer graphics forum*, 2009, pp. 1383-1392.
- [51] J. Wu, X. Shen, W. Zhu, and L. Liu, "Mesh saliency with global rarity," *Graphical Models*, vol. 75, pp. 255-264, 2013.
- [52] R. Song, Y. Liu, R. R. Martin, and P. L. Rosin, "Mesh saliency via spectral processing," *ACM Transactions on Graphics (TOG)*, vol. 33, p. 6, 2014.

- [53] G. Leifman, E. Shtrom, and A. Tal, "Surface regions of interest for viewpoint selection," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 414-421.
- [54] R. Gal and D. Cohen-Or, "Salient geometric features for partial shape matching and similarity," *ACM Transactions on Graphics (TOG)*, vol. 25, pp. 130-150, 2006.
- [55] Y. Zhao, Y. Liu, K. Song, M. Zhang, and R. Song, "A Retinex theory based point sampling method," in *Soft Computing and Pattern Recognition (SoCPaR), 2011 International Conference of*, 2011, pp. 330-335.
- [56] Y. Zhao, Y. Liu, R. Song, and M. Zhang, "Extended non-local means filter for surface saliency detection," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, 2012, pp. 633-636.
- [57] Y.-B. Yang, T. Lu, and J.-J. Lin, "Saliency regions for 3D mesh abstraction," in *Advances in Multimedia Information Processing-PCM 2009*, ed: Springer, 2009, pp. 292-299.
- [58] M. Garland and P. S. Heckbert, "Surface simplification using quadric error metrics," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, 1997, pp. 209-216.
- [59] A. M. Derrington, J. Krauskopf, and P. Lennie, "Chromatic mechanisms in lateral geniculate nucleus of macaque," *The Journal of Physiology*, vol. 357, pp. 241-265, 1984.
- [60] Z.-L. Lu and B. Doshier, *Visual psychophysics: From laboratory to theory*: MIT Press, 2013.
- [61] D. Cohen-Steiner and J.-M. Morvan, "Restricted delaunay triangulations and normal cycle," in *Proceedings of the nineteenth annual symposium on Computational geometry*, 2003, pp. 312-321.
- [62] G. Loy and J.-O. Eklundh, "Detecting symmetry and symmetric constellations of features," in *Computer Vision-ECCV 2006*, ed: Springer, 2006, pp. 508-521.
- [63] J. Zhang, J. Sun, J. Liu, C. Yang, and H. Yan, "Visual attention model based on multi-scale local contrast of low-level features," in *Proceedings of IEEE 10th International Conference on Signal Processing*, 2010, pp. 902-905.
- [64] J. Harel, C. Koch, and P. Perona, "Graph-based visual saliency," in *Advances in neural information processing systems*, 2006, pp. 545-552.
- [65] T. Kadir and M. Brady, "Saliency, scale and image description," *International Journal of Computer Vision*, vol. 45, pp. 83-105, 2001.

- [66] A. Holzbach and G. Cheng, "A fast and scalable system for visual attention, object based attention and object recognition for humanoid robots," in *2014 IEEE-RAS International Conference on Humanoid Robots*, 2014, pp. 316-321.
- [67] W. Gonzalez and R. E. Woods, "Eddins, Digital Image Processing Using MATLAB," *Third New Jersey: Prentice Hall*, 2004.
- [68] T. Möller and B. Trumbore, "Fast, minimum storage ray/triangle intersection," in *ACM SIGGRAPH 2005 Courses*, 2005, p. 7.
- [69] G. Peyre. (2007). *Toolbox Graph*. Available: <http://www.mathworks.com/matlabcentral/fileexchange/5355-toolbox-graph>
- [70] P. Cignoni, C. Rocchini, and R. Scopigno, "Metro: measuring error on simplified surfaces," in *Computer Graphics Forum*, 1998, pp. 167-174.
- [71] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, pp. 600-612, 2004.
- [72] V. Laparra, J. Ballé, A. Berardino, and E. P. Simoncelli, "Perceptual image quality assessment using a normalized Laplacian pyramid," *Electronic Imaging*, vol. 2016, pp. 1-6, 2016.
- [73] CloudCompare. *3D Point Cloud and Mesh Processing Software, Open Source Project*. Available: <http://www.danielgm.net/cc/>
- [74] H. Monette-Therriault, A.-M. Cretu, and P. Payeur, "3D object modeling with neural gas based selective densification of surface meshes," in *Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on*, 2014, pp. 1354-1359.
- [75] G. Lavoué, M. Tola, and F. Dupont, "MEPP-3D MESH PROCESSING PLATFORM."
- [76] *A Benchmark for 3D Interest Points Marked by Human Subjects*. Available: <http://www.itl.nist.gov/iad/vug/sharp/benchmark/> 3DInterestPoint.
- [77] M. Chagnon-Forget, G. Rouhafzay, A.-M. Cretu, and S. Bouchard, "Enhanced Visual-Attention Model for Perceptually Improved 3D Object Modeling in Virtual Environments," *3D Research*, vol. 7, issue 4, p. 1-18, 2016.