

Évaluation de la qualité visuelle des maillages 3D en utilisant un réseau neuronal convolutif et la saillance visuelle

Ilyass Abouelaziz¹

Aladine Chetouani²

Mohammed El Hassouni¹

Hocine Cherifi³

¹ LRIT-CNRST, URAC 29, Rabat IT Center, Faculty of Sciences, Mohammed V University, Rabat, Morocco.

² University of Orleans – PRISME laboratory - Orleans, France .

⁴ LE2I UMR 6306 CNRS, University of Burgundy, Dijon, France.

Adresse du ou des auteurs

ilyass.abouelaziz@gmail.com , aladine.chetouani@univ-orleans.fr,
mohamed.elhassouni@gmail.com, hocine.cherifi@u-bourgogne.fr

Résumé – Dans ce travail, nous proposons une méthode d’estimation de la qualité des maillages 3D sans référence exploitant l’information de saillance visuelle et un réseau de neurones convolutionnels (CNN). Une fois le maillage de saillance calculé, des vues 2D sont extraites du modèle 3D et du maillage de saillance correspondant. La saillance locale de chaque vue 2D est ensuite analysée pour sélectionner les régions (patches) les plus pertinents. Les patches retenus sont envoyés en entrée d’un modèle CNN pour prédire le score de qualité de chacun d’eux. La note de qualité finale est donnée par la moyenne des scores. La méthode proposée a été évaluée sur la base *general purpose* et les résultats obtenus ont montré la pertinence de la méthode.

Abstract – In this work, we propose a method for estimating the quality of 3D meshes without reference using visual saliency information and a convolutional neural network (CNN). Once the saliency mesh is calculated, 2D views are extracted from the 3D model and the corresponding saliency mesh. The local saliency of each 2D view is then analyzed to select the most relevant regions (patches). The selected patches are used as input to our CNN model in order to predict the quality score of each one. The final quality score is given by the average score. The proposed method was evaluated on the *general purpose* dataset and the results obtained showed the relevance of the method.

1 Introduction

De même que les images et les vidéos, la qualité perceptuelle des maillages 3D peut être affectée par plusieurs facteurs externes (tatouage, compression, simplification, etc.) [8]. Afin d’estimer l’impact de ces traitements, plusieurs mesures de qualité ont été proposées dans la littérature. Ils existent actuellement trois grandes familles de métriques: référence complète (FR) qui suppose que le maillage de référence est disponible, référence réduite (RR) qui exploite uniquement des caractéristiques du maillage de référence et les mesures sans référence (NR), appelées également aveugle, qui n’ont accès à aucune information du maillage de référence.

Les premiers travaux développés ont utilisé des similarités simples entre le maillage de référence et sa version déformée telle que l’erreur quadratique moyenne (RMS) [4] et la distance de Hausdorff (HD) [2]. Cependant, ce type de mesures n’est pas totalement en accord avec notre perception humaine car elles n’exploitent pas les spécificités du Système Visuel Humain (SVH). Afin de pallier à cela, de nombreux auteurs ont axé leur travaux sur l’élaboration de mesures intégrant des caractéristiques perceptuelles. Dans [6], les auteurs ont proposé la métrique MSDM qui est basée sur l’analyse des courbures. Corsini et *al.* ont développé une métrique permettant d’évaluer la qualité des maillages

tatoués à travers les variations de la rugosité du maillage [5]. Dans [12], une mesure perceptuelle, nommée FMPD (Fast Mesh Perceptual Distance) a également été proposée. Cette métrique est basée sur une mesure de rugosité locale dérivée de la courbure gaussienne. Ces méthodes avec référence obtiennent une corrélation élevée avec la perception humaine. Cependant, leur principal inconvénient réside dans l’utilisation du maillage de référence, ce qui n’est pas le cas dans la plupart des applications réelles.

Dans cet article, nous présentons une méthode d’estimation de la qualité des maillages 3D sans référence basée sur l’utilisation d’un réseaux de neurones convolutifs (CNN) et de la saillance visuelle. L’idée développée ici est d’estimer la qualité en se focalisant sur les régions qui attirent notre attention visuelle. Ainsi, la saillance est ici employée pour sélectionner les patches les plus saillants et les utiliser comme entrée d’un modèle CNN. La méthode proposée a été évaluée sur la base *general purpose* et les résultats obtenus ont montré la pertinence de la méthode proposée.

Le reste de ce document est organisé comme suit: La méthode proposée est décrite en détails dans la section 2. Les résultats expérimentaux sont présentés dans la section 3. La conclusion et les perspectives sont données dans la section 4.

2 Méthode proposée

La figure 1 présente les différentes étapes de la méthode proposée. A partir du maillage 3D d'un objet, la saillance de chaque sommet est tout d'abord calculée. Une série de vues 2D de l'objet 3D et de son maillage saillant correspondant sont ensuite extraites. La saillance de chaque vue 2D est ensuite analysée localement afin de sélectionner les plus pertinents perceptuellement. Ces derniers sont finalement utilisés comme entrée d'un modèle CNN pour estimer la qualité du maillage.

2.1 Saillance visuelle des objets 3D

L'attention visuelle joue un rôle important lorsque l'humain analyse un objet. En effet, les observateurs attachent plus d'importances à certaines régions que d'autres. Ces régions, communément appelées régions saillantes, possèdent certaines caractéristiques (couleur, orientation, intensité, etc.) qui attirent notre regard. Dans ce travail, nous proposons d'exploiter cette information de saillance visuelle pour sélectionner les régions les plus perceptuellement pertinentes. Pour ce faire, nous avons utilisé la méthode de saillance proposée par Lee et al [7] qui est basée sur les courbures moyennes. Soit un objet 3D, la courbure moyenne de chaque sommet est tout d'abord calculée. Le maillage obtenu est ensuite filtré par deux filtres gaussiens pour calculer la saillance à différentes échelles. Le maillage de saillance 3D est finalement obtenu par la somme des valeurs de saillance obtenues sur les différentes échelles.

2.2 Projections 2D

Des vues 2D sont ensuite extraites à partir du modèle 3D et du maillage de saillance correspondant. Les vues sont ici obtenues en suivant la procédure de projection décrite dans [3]. Pour ce faire, le centroïde de l'objet 3D est placé à l'origine du système de coordonnées sphériques. Les coordonnées $(\theta_{az}, \theta_{el})$ des caméras virtuelles sont obtenues en divisant $[0, 2\pi]$ par 8 pour les valeurs de θ_{az} , et diviser $[0, \pi/2]$ par 8 les valeurs de θ_{el} . Ainsi, 64 vues sont utilisées pour représenter un objet 3D.

2.3 Sélection des patches basée sur la saillance

Les vues 2D du maillage et du maillage saillant correspondant sont décomposées en patches de taille 28×28 sans chevauchement. Pour chaque patch de la carte de saillance, le niveau de saillance moyen (LoS) est calculé et est utilisé pour sélectionner les patches les plus pertinents selon le critère suivant: $LoS \geq S_t$. Le seuil de saillance S_t a été ici fixé expérimentalement à 0.4. Cette étape de sélection des patches saillants nous permet d'obtenir en moyenne de 13787 patches par modèle.

2.4 Estimation de la qualité visuelle

L'étape suivante consiste à utiliser un réseau de neurones convolutifs pour estimer la qualité visuelle. L'architecture du modèle CNN utilisé est composée de différentes couches (voir

figure 2). La première couche est une couche de convolution de 50 noyaux de taille 5×5 . La deuxième couche est une couche de max-pooling de taille 2×2 avec un pas de 2. La troisième couche est une couche de convolution avec 50 noyaux de taille 5×5 . La quatrième couche est une couche de max-pooling avec une fenêtre locale de taille 8×8 . En sortie de cette couche, nous obtenons un vecteur de caractéristiques de taille 1×50 . Les cinquième et sixième couches sont deux couches entièrement connectées de 250 neurones chacune. La dernière couche est une couche de régression linéaire dont la sortie correspond au score de qualité.

2.5 Apprentissage et prédiction de la qualité

Durant la phase d'apprentissage, la cible de chaque patch correspond au score de qualité subjectif de l'objet 3D. De même que dans [9], nous adoptons la fonction objective d'entraînement définie comme suit:

$$L = \frac{1}{N} \sum_{n=1}^N \|S(p_n; \omega) - MOS_n\|_{l1}$$

$$\hat{\omega} = \min_{\omega} L$$

où MOS_n est le score d'opinion moyen attribué à un patch p_n et $S(p_n; \omega)$ est le score prédit de p_n avec les poids ω . La descente de gradient stochastique (SGD) a été utilisée comme fonction d'optimisation et le nombre d'epochs a été fixé à 40.

3 Résultats expérimentaux

3.1 Bases de données et protocole de validation

Pour évaluer les performances de la méthode proposée, la base de données *General purpose* a été utilisée. Elle est constituée de 84 modèles déformés issus de 4 modèles différents. Le bruit et le lissage sont les deux types de dégradations considérés. Les notes subjectives de chaque objet et ses versions dégradées sont également fournies.

La méthode leave-one-out-cross-validation (LOOCV) a été appliquée durant l'apprentissage. Les patches des objets déformés existants dans la base de données sont partitionnés en base d'apprentissage-validation et en base de test. Lors de chaque partition, 3 modèles et ses versions dégradées sont utilisés pour l'apprentissage-validation et le reste pour le test. Ainsi, les modèles 3D qui constituent le jeu de test sont totalement différents de ceux de l'apprentissage. Ainsi, $3 \times 21 \times 13787$ échantillons (75%) de la base de données sont utilisés pour l'apprentissage et le reste pour le test ($1 \times 21 \times 13787$, i.e. 25%). Deux types de coefficients de corrélation sont couramment utilisés pour évaluer les mesures de qualité, à savoir le coefficient de corrélation linéaire de Pearson (r_p) qui mesure la précision de la prédiction et le coefficient de corrélation des rangs de Spearman (r_s) qui mesure la monotonie de prédiction

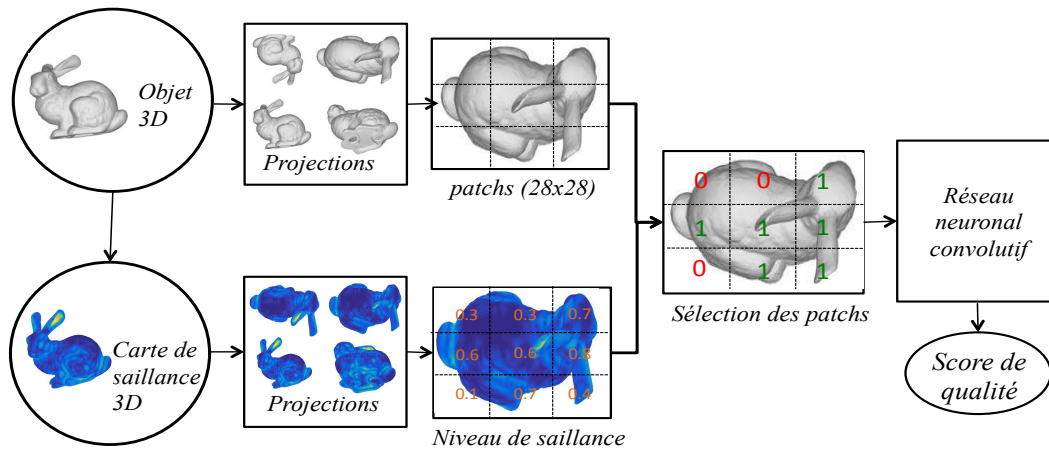


Figure 1 – Schéma général de la méthode proposée.

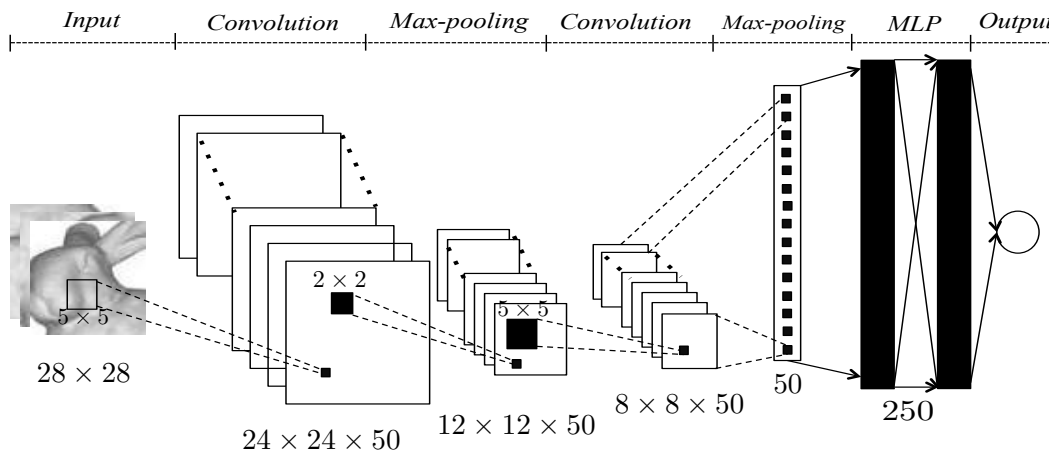


Figure 2 – Configuration du réseau neuronal convolutif pour l'évaluation de la qualité des maillages 3D.

[13]. Pour ces deux critères, une corrélation de 1 indique une corrélation parfaite.

3.2 Résultats et discussion

Dans cette section, nous évaluons les performances de notre méthode et comparons les résultats obtenus à l'état de l'art (FR, RR en NR). Les résultats obtenus sont listés respectivement dans le tableau 1. Les valeurs en gras représentent les coefficients de corrélation les plus élevés. Il convient de préciser que les corrélations sur l'ensemble du corpus sont calculées entre les scores objectifs obtenus de tous les objets dans le corpus et le MOS correspondant.

Les métriques classiques HD et RMS ont obtenu des corrélations très faibles, tandis que les méthodes basées sur la perception ont obtenu des scores de corrélation élevés. Comparé aux méthodes sans référence, notre méthode a obtenu les meilleures performances avec un gain de corrélation allant de 8% à 13%. Comparé aux autres approches (FR et RR), la méthode propo-

sée a obtenu la corrélation de Pearson la plus élevée (90.2) et la métrique FMPD a obtenu la corrélation de Spearman la plus élevée (90.3), soit un gain de corrélation par rapport à notre méthode de 2,26%. Cependant, contrairement à notre méthode, la mesure FMPD nécessite des caractéristiques du modèle 3D.

4 Conclusion

Dans cet article, nous avons proposé une méthode d'évaluation de la qualité visuelle sans référence des maillages 3D basée sur l'utilisation d'un réseau de neurones convolutif et une étape de sélection de patches à partir de la saillance visuelle. La méthode a été évaluée et comparée aux méthodes de l'état de l'art. Les résultats obtenus ont montré la pertinence de l'approche proposée.

Comme perspectives, nous envisageons d'utiliser les réseaux de neurones basés sur les graphes, ce qui permettrait de s'affranchir de l'étape d'extraction des vues 2D.

Table 1 – Coefficients de corrélation r_s (%) et r_p (%) obtenues sur la base general-purpose.

Type	Méthode	Armadillo		Dyno		Venus		Rocker		All	
		r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p	r_s	r_p
FR	HD [4]	69.5	30.2	30.9	22.6	1.6	0.8	18.1	5.5	13.8	1.3
	RMS [2]	62.7	32.3	0.3	0.0	90.1	77.3	7.3	3.0	26.8	7.9
	MSDM2 [6]	81.6	85.3	85.9.4	85.7	89.3	87.5	89.6	87.2	80.4	81.4
	TPDM [11]	84.5	78.8	92.2	89.0	90.6	91.0	92.2	91.4	89.6	89.2
RR	3DWPM1 [5]	65.8	35.7	62.7	35.7	71.6	46.6	87.5	53.2	69.3	38.4
	3DWPM2 [5]	74.1	43.1	52.4	19.9	34.8	16.4	37.8	29.9	49.0	24.6
	FMPD [12]	75.4	83.2	89.6	88.9	87.5	83.9	88.8	84.7	90.3	89.4
NR	NR-SVR [1]	76.8	91.5	78.6	84.1	85.7	88.6	86.2	86.6	81.5	87.8
	BMQI [10]	20.1	NA	83.5	NA	88.9	NA	92.7	NA	78.1	NA
	Our method	85.6	86.9	85.6	88.4	95.3	92.7	88.6	84.9	88.4	90.2

References

- [1] I. Abouelaziz, M. El Hassouni, and H. Cherifi. No-reference 3d mesh quality assessment based on dihedral angles model and support vector regression. In *International Conference on Image and Signal Processing*, pages 369–377. Springer, 2016.
- [2] N. Aspert, D. Santa-Cruz, and T. Ebrahimi. Mesh: Measuring errors between surfaces using the hausdorff distance. In *Multimedia and Expo, 2002. ICME'02. Proceedings. 2002 IEEE International Conference on*, volume 1, pages 705–708. IEEE, 2002.
- [3] S. Bai, X. Bai, Z. Zhou, Z. Zhang, Q. Tian, and L. J. Latecki. Gift: Towards scalable 3d shape retrieval. *IEEE Transactions on Multimedia*, 19(6):1257–1271, 2017.
- [4] P. Cignoni, C. Rocchini, and R. Scopigno. Metro: Measuring error on simplified surfaces. In *Computer Graphics Forum*, volume 17, pages 167–174. Wiley Online Library, 1998.
- [5] M. Corsini, E. D. Gelasca, T. Ebrahimi, and M. Barni. Watermarked 3-d mesh quality assessment. *IEEE Transactions on Multimedia*, 9(2):247–256, 2007.
- [6] G. Lavoué. A multiscale metric for 3d mesh visual quality assessment. In *Computer Graphics Forum*, volume 30, pages 1427–1437. Wiley Online Library, 2011.
- [7] C. H. Lee, A. Varshney, and D. W. Jacobs. Mesh saliency. In *ACM transactions on graphics (TOG)*, volume 24, pages 659–666. ACM, 2005.
- [8] D. P. Luebke. A developer’s survey of polygonal simplification algorithms. *IEEE Computer Graphics and Applications*, 21(3):24–35, 2001.
- [9] A. Mittal, A. K. Moorthy, and A. C. Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012.
- [10] A. Nouri, C. Charrier, and O. Lézoray. 3d blind mesh quality assessment index. *Electronic Imaging*, 2017(20):9–26, 2017.
- [11] F. Torkhani, K. Wang, and J.-M. Chassery. A curvature tensor distance for mesh visual quality assessment. *Computer Vision and Graphics*, pages 253–263, 2012.
- [12] K. Wang, F. Torkhani, and A. Montanvert. A fast roughness-based approach to the assessment of 3d mesh visual quality. *Computers & Graphics*, 36(7):808–818, 2012.
- [13] Z. Wang and A. C. Bovik. Modern image quality assessment. *Synthesis Lectures on Image, Video, and Multimedia Processing*, 2(1):1–156, 2006.