

Impact de la compression des normales sur l'évaluation de qualité subjective de nuages de points éclairés

A. Tious, T. Vigier, V. Ricordel

Nantes Université, École Centrale Nantes, CNRS, LS2N, UMR 6004, F-44000
Nantes, France
amar.tious@univ-nantes.fr

Résumé

L'avènement de nouveaux cas d'usage intégrant des représentations 3D de contenus naturels a mené au développement de nouvelles méthodes pour la compression, le rendu et l'évaluation de qualité de nuage de points (point clouds). Cependant, ces méthodes ne traitent que les coordonnées et la couleur des points, et ne prennent pas en compte d'autres attributs pouvant jouer un rôle dans l'optimisation du rendu de point clouds. Dans cet article, nous nous intéressons aux normales des points qui permettent un rendu avec ombrage des point clouds. Nous ajoutons le traitement des normales au codec V-PCC en considérant 2 approches : encoder les normales comme attributs, ou bien les recalculer après décodage à partir du point cloud décompressé. Nous comparons ces 2 approches en terme de qualité visuelle dans une expérience d'évaluation subjective de point clouds avec ombrage. Nous démontrons alors les avantages que présente la compression des normales pour différents types de contenus et cas d'usages. Nous concluons en identifiant les prochaines étapes vers un rendu physique réaliste des point clouds.

Mots clefs

Nuage de Points, Point cloud, Compression vidéo, Évaluation de qualité subjective, Rendu physique réaliste, V-PCC

1 Introduction

L'imagerie volumétrique permet de représenter des scènes et objets réels en 3D. Ces données peuvent être capturées avec des scanners LiDAR pour de grandes scènes ou des studios et rigs multi-caméras pour les objets et personnes. Un nuage de points (point cloud) décrit ce type de capture comme un ensemble de points caractérisés par leurs coordonnées 3D et divers attributs décrivant la texture et la surface du contenu. C'est une bonne alternative aux maillages 3D (mesh), car plus facile à stocker et moins coûteux en calcul [1]. Suite aux avancées technologiques des réalités étendues, l'utilisation de point clouds est devenue intéressante dans des domaines comme la vidéo 3D immersive, l'architecture, les musées virtuels et la télécommunication [2]. Ces applications nécessitent de nouvelles méthodes de compression des données, de rendu et d'évaluation de qualité. A ce sujet, les groupes de travail MPEG-I [1] et JPEG

Pleno [3] ont défini des normes pour la compression et pour les conditions de tests subjectifs d'évaluation de qualité de point clouds. Les tests subjectifs consistent à faire évaluer la qualité visuelle de point clouds par des observateurs humains afin d'évaluer l'effet des déformations causées par des méthodes de compression, de rendu ou de reconstruction sur le contenu original.

Cependant, la plupart des études d'évaluation de qualité considèrent uniquement la compression des coordonnées et des couleurs des points sans prendre en compte d'autres attributs ou des méthodes de rendu avancées qui peuvent pourtant avoir un impact important sur la qualité perçue du contenu 3D [4]. Rossoni et al. [5] proposent un pipeline théorique tenant compte des propriétés de surface et de matériau pour un rendu physique réaliste (PBR). Cependant, aucune implémentation n'a été réalisée jusqu'à présent, et donc aucune étude d'évaluation de qualité n'a pu valider son intérêt.



FIGURE 1 – Les sources sélectionnées, rendues sans ombrage.

Pour cet article, nous avons mené une étude d'évaluation de qualité subjective visant à comparer deux approches permettant de considérer les normales (coordonnées des vecteurs normaux aux plans de surface des points) en plus des coordonnées et des couleurs dans la compression et le rendu de point clouds. La première approche est celle que nous proposons dans [6] qui est d'encoder les normales en tant qu'attribut comme les couleurs. La seconde approche est de recalculer les normales après compression à partir de la géométrie déformée du point cloud décodé, comme fait précédemment dans [7]. La seconde méthode est à ce jour le cas de figure utilisé dans les rares cas où

TABLEAU 1 – Débits des niveaux de qualité en bits/point (géométrie+normales+couleurs). nE n'a pas de débit car les normales sont calculées et non transmises.

| | | | Normales | | | |
|-----------|----|------|----------|------|------|------|
| | | | n1 | n3 | n5 | nE |
| | | | 0.07 | 0.15 | 0.36 | 0 |
| Géométrie | g1 | 0.03 | 7.20 | 7.28 | 7.49 | 7.13 |
| | g3 | 0.05 | 7.22 | 7.30 | 7.51 | 7.15 |
| | g5 | 0.12 | 7.29 | 7.37 | 7.58 | 7.22 |

les normales sont utilisées après transmission [7]. Le codec V-PCC (video-based point cloud compression) étant le standard de MPEG-I offrant le meilleur rapport qualité visuelle/coût de transmission [8, 9], nous concentrons notre étude sur l'utilisation de ce codec. Notre objectif est d'évaluer l'intérêt d'encoder et transmettre les normales plutôt que de les recalculer dans le cas où les objets seront ensuite rendus en 3D avec un éclairage virtuel.

2 Méthodologie

2.1 Génération de Stimuli

Pour prendre en compte différents cas de contenu et d'applications, nous avons choisi des point clouds sources de différentes catégories sémantiques, parmi ceux de BASICS [10], l'un des datasets d'évaluation de qualité de point cloud le plus riche à ce jour. Six sources ont été sélectionnées parmi celles du dataset selon 3 catégories (2 "Humans", 2 "Objects", 2 "Buildings"). Tous ces modèles sont des numérisations de contenus réels possédant des normales de surfaces en attributs, et ayant été voxelisés dans une grille de 1024x2024x2024 points. La figure 1 présente les six modèles choisis.

Pour chaque source, nous avons généré 12 versions déformées en nous inspirant des niveaux de qualité recommandés par [3]. Pour observer séparément l'impact de la compression des différents paramètres, nous croisons 3 niveaux de qualité des normales ($n1, n3, n5$) avec 3 niveaux de qualité de géométrie ($g1, g3, g5$). Pour chaque niveau de géométrie, nous avons également généré un niveau de qualité nE pour lequel les normales ont été recalculées plutôt que compressées, puisque l'estimation de celles-ci est basée sur la géométrie. Étant donné que la distorsion de la couleur des points impacterait trop fortement la qualité perçue [10], nous maintenons un niveau de compression sans perte pour cet attribut (débit moyen=7.1bpp). Ainsi, les distorsions sont causées seulement par la compression des normales et des coordonnées des points. Le Tableau 1 résume tous les niveaux de qualité utilisés en indiquant, pour chacun, la somme des débits de la géométrie, des normales et des couleurs. Il est important de noter que la couleur, même si ajustée à un niveau de compression juste suffisant, serait toujours bien plus coûteuse en débit que la géométrie. La méthode que nous utilisons pour compresser et estimer les normales est décrite dans notre précédent article [6].



FIGURE 2 – Exemple de point cloud déformé avec des normales compressées, et recalculées.

2.2 Condition de tests

Nous avons suivi les conditions de tests recommandées par ISO/IEC N91058 [3] pour l'évaluation de qualité de point cloud. Les participants étaient placés face à un écran IPS Dell 24" (P2422H) à une distance de 3 fois la hauteur de l'écran (soit 90cm). La luminosité et le gamut de l'écran ainsi que l'éclairage de la salle de test ont été calibrés selon la norme ITU-R BT.500 [11]. La tâche des participants était de noter les déformations des stimuli sur l'échelle DSIS (Double Stimulus Impairment Scale)[11] à 5 niveaux (*Très gênant, Gênant, Légèrement gênant, Perceptible mais pas gênant, Imperceptible*). Les point clouds ont été présentés aux observateurs dans des vidéos de 10 secondes capturées par une caméra virtuelle tournant autour de l'objet. La trajectoire de la caméra est la même que dans l'expérience de [10] : 420° autour de l'axe y (-30° à 390°) et 60° autour de l'axe x (0° à 60°). Pour rendre visible la distorsion des normales, nous avons généré ces vidéos via Unity3D, en utilisant un shader avec une taille de point ajustable et un ombrage de surface [5, 12] afin que les objets soient éclairés et pleins. La scène enregistrée est éclairée par une lumière ambiante à 25% d'intensité et une lumière directionnelle à 100% d'intensité située 30° au-dessus et face à l'objet, comme conseillé par [13] pour simuler un éclairage naturel. L'interface de test est illustrée en Figure 3.



FIGURE 3 – Capture d'écran de l'interface de test.

Pour familiariser les participants avec l'interface et la tâche, la session de test était précédée d'une session d'entraînement pendant laquelle 6 versions déformées d'une source non sélectionnée pour le test étaient montrées. Pendant la session principale, chaque participant devait noter les 72 (6x12) versions déformées de point clouds, résultant en une durée moyenne de test de 15 minutes.

Nous avons collecté les scores de 20 utilisateurs au total (14 hommes et 6 femmes) âgés entre 21 et 49 ans, ayant tous une vision correcte ou corrigée, dont 45% sont non-experts, et 55% sont experts mais non familiers au type de stimulus utilisé.

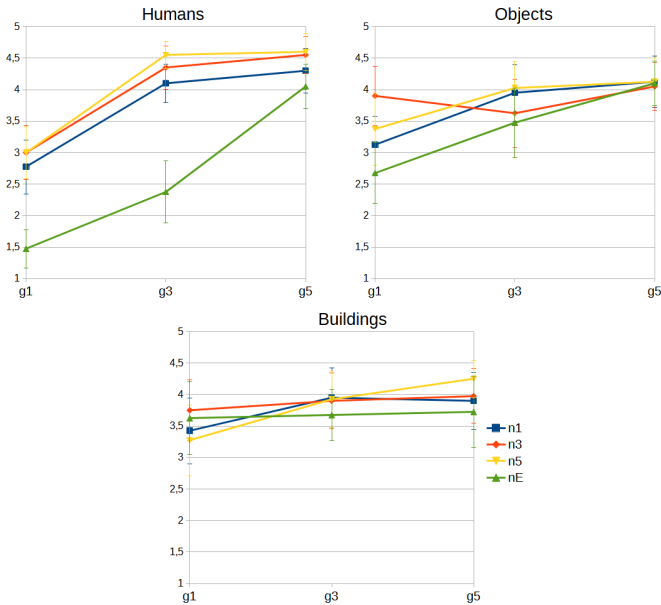


FIGURE 4 – MOS pour tous les niveaux de qualité et chaque catégorie de contenu (avec intervalle de confiance de 95%).

3 Résultats & Discussion

3.1 Scores de qualité et Discriminabilité

Les scores d'opinion moyen (MOS) pour chaque niveau de qualité et pour chaque catégorie de contenu sont représentés en Figure 4, avec les intervalles de confiance ($\gamma = 95\%$). Comme on pouvait s'y attendre, en l'absence de déformation de la couleur (qui est le principal facteur d'influence sur la notation dans les précédentes études), la tâche d'évaluation est plus difficile car l'ambiguïté des déformations est plus élevée. La difficulté est telle que pour les catégories "Objects" et "Buildings", les MOS ne permettent pas de discerner de classement des déformations. Cela semble indiquer que pour ces contenus, le niveau de compression de la géométrie et des normales avec V-PCC n'a pas d'impact sur la qualité perçue. Cela implique qu'utiliser le niveau de qualité le moins coûteux ($g1 + nE$) peut être satisfaisant. Cependant, ces résultats ne sont pas assez significativement fiables pour conclure. Sur la Figure 5 est représentée la variation de la discriminabilité des MOS en fonction du nombre de participants. Ici, la discriminabilité est le taux de paires de déformations avec des scores de qualité significativement différents [14] (test de Student avec $p - value > 0.05$). Nous pouvons voir que pour l'intégralité de nos données, même avec 20 participants, la discriminabilité ne converge pas encore vers un

maximum, ce qui peut suggérer que plus de participants serait nécessaire pour obtenir des résultats significatifs.

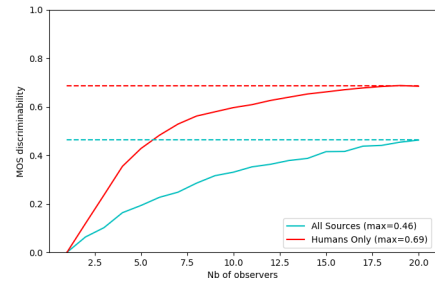


FIGURE 5 – Évolution de la discriminabilité des MOS en fonction du nombre de participants, pour l'ensemble des stimuli et pour seulement la catégorie "Humans".

Cependant, la même conclusion ne s'applique pas aux point clouds de la catégorie "Humans". Le taux de paires de déformations significativement différentes est plus élevé si on ne considère que ces 2 sources là, et il converge à partir de 18 participants, indiquant que les observations que l'on peut faire sur les MOS de ces contenus sont significatives. Pour ces modèles, on voit en Fig.4 que recalculer les normales après compression (nE) rend visible les distorsions de la géométrie pour les niveaux $g1$ & $g3$. Cet effet révélateur est mis en évidence par un rendu anormalement ombré des objets et résulte en une mauvaise qualité perçue par les observateurs. En comparaison, on voit que des normales encodées (quel que soit le niveau de qualité) ont un effet de masquage sur les éventuels défauts de géométrie et permettent donc une meilleure qualité visuelle du rendu. De plus, la différence moindre entre les niveaux $n1$, $n3$, $n5$ de qualité des normales montre que le coût de transmission des normales peut être minimisé. Aussi les combinaisons $g3 + n1$ et $g1 + n1$ ont des débits équivalents et des qualités de rendu supérieures à respectivement $g5 + nE$ et $g3 + nE$, ce qui veut dire que l'encodage des normales peut même permettre d'économiser des bits sur la transmission de la géométrie.

Pour des cas d'usage où les point clouds sont utilisés pour représenter des humains dans des scènes avec éclairage virtuel (télécommunication et vidéo volumétriques), il y a donc un réel avantage à encoder les normales plutôt que de les recalculer. D'autant plus, il faut noter que la version publique du codec V-PCC utilisée dans cette expérience date de 2020. Celui-ci est aujourd'hui bien plus performant à bas débit.

3.2 Limites et Améliorations

Bien que significatives, nos observations sur la catégorie "Humans" ne sont basées que sur deux point clouds sources et ne couvrent pas la diversité de ce type de contenu. Une nouvelle expérience centrée sur cette catégorie sémantique avec plus de sources permettrait de robustifier ces résultats. L'effet de masquage de déformations géométriques qu'on

les normales transmises sur le rendu n'est également pas la seule piste pour améliorer la compression et le rendu de point clouds. La compression des attributs de couleurs était hors du cadre de cet article, mais de futurs tests s'y intéressant pourraient révéler quel niveau de masquage existe entre l'ombrage et les niveaux de déformations de couleurs. Mis à part les normales, il y a aussi d'autres propriétés de matériaux, de texture ou de surface qui pourraient être ou représentés par d'autres attributs [5] (réflectance, transparence, aspect métallique, ...) et leur traitement en transmission mériterait d'être étudié.

Vu que nos résultats semblent concerner principalement des cas d'usage impliquant du contenu dynamique et des applications immersives, Il sera important d'étudier comment d'autres facteurs d'influence, propre au cas de vidéos volumétriques et de réalité étendue [15, 16] (degrés de liberté, interaction, attention visuelle, ...), affecte la perception de ces déformations de compression et rendus.

Enfin, l'impact de ces mêmes déformations sur des métriques de qualité objective pour le point clouds et sur leur performance en termes de corrélation avec des scores d'évaluation subjectives est également à considérer. Le développement des nouvelles métriques, ou l'adaptation de métriques existantes, adaptés à l'évaluation de qualité de rendu est donc nécessaire, comme nous suggérons dans un précédent article [6]. Concernant les autres types de contenus que les humains, des normales compressées, même si elle ne semble pas faire de différence en qualité à priori, pourrait malgré tout s'avérer utile pour la vision par ordinateur. Elles sont plus fidèles à la surface du point cloud avant compression et pourraient servir à corriger des déformations géométriques.

4 Conclusion

Nous avons montré que compresser les normales peut être avantageux en termes de qualité et/ou de coût de transmission pour des cas d'usages tels que la télécommunication et le divertissement immersifs où les point clouds représentent des humains. Nous encourageons donc l'utilisation de cette méthode. Des tests supplémentaires avec plus de stimuli de cette catégorie viendrait robustifier ces résultats. Plus d'études et expériences seraient également intéressantes afin d'évaluer la compression de nouveaux attributs et le rendu physique réaliste de point clouds dans des contextes d'usages, en réalité étendue.

Références

- [1] Sebastian Schwarz et al. Emerging mpeg standards for point cloud compression. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 9(1) :133–148, 2018.
- [2] Stuart Perry. JPEG Pleno Point Cloud Use Cases and Requirements, v1.6. *ISO/IEC JTC 1/SC 29/WG1*, 100096, 2022.
- [3] Stuart Perry. Jpeg pleno point cloud common test conditions v3.6. *ISO/IEC JTC1/SC29/WG1 N*, 86044, Avril 2021.
- [4] Kenneth Vanhoey et al. Visual Quality Assessment of 3D Models : On the Influence of Light-Material Interaction. *ACM Transactions on Applied Perception*, 15(1) :1–18, Octobre 2017.
- [5] Marco Rossoni et al. Physically-based Rendering of Animated Point Clouds for eXtended Reality. *Journal of Computing and Information Science in Engineering*, pages 1–9, Septembre 2023.
- [6] Amar Tious et al. Impact of point cloud normals compression on objective quality assessment. Dans *EU-SIPCO*, Lyon, France, 2024. To be published.
- [7] Alireza Javaheri et al. Point cloud rendering after coding : Impacts on subjective and objective quality. *IEEE Transactions on Multimedia*, 23 :4049–4064, 2021.
- [8] Emin Zerman, Cagri Ozcinar, Pan Gao, et Aljosa Smolic. Textured Mesh vs Coloured Point Cloud : A Subjective Study for Volumetric Video Compression. Dans *QoMEX 2020*, pages 1–6, Mai 2020.
- [9] Cheng-Hao Wu et al. Quantitative Comparison of Point Cloud Compression Algorithms With PCC Arena. *IEEE Transactions on Multimedia*, 25 :3073–3088, 2023.
- [10] Ali Ak et al. BASICS : Broad Quality Assessment of Static Point Clouds in a Compression Scenario. *IEEE Transactions on Multimedia*, pages 1–13, 2024.
- [11] ITU-R Recommendation BT.500-14. Methodology for the subjective assessment of the quality of television pictures, Octobre 2019.
- [12] Amar Tious et al. Physically-based Lighting of 3D Point Clouds for Quality Assessment. Dans *ACM IMX '23*, pages 423–426, New York, NY, USA, 2023.
- [13] James P. O'Shea et al. The assumed light direction for perceiving shape from shading. Dans *APGV '08*, pages 135–142, New York, NY, USA, 2008.
- [14] Andreas Pastor et al. Comparison of conditions for omnidirectional video with spatial audio in terms of subjective quality and impacts on objective metrics resolving power. Dans *ICASSP 2024*, pages 8210–8214, 2024.
- [15] Jesús Gutiérrez et al. Subjective Evaluation of Dynamic Point Clouds : Impact of Compression and Exploration Behavior. Dans *EUSIPCO*, pages 675–679, Septembre 2023. ISSN : 2076-1465.
- [16] Silvia Rossi et al. Behavioural Analysis in a 6-DoF VR System : Influence of Content, Quality and User Disposition. Dans *ACM IXR '22*, pages 3–10, New York, NY, USA, Octobre 2022.