

Enjeux de l'IA pour l'aide à la modélisation 3D d'intérieur d'installation industrielle construite et leur exploitation en réalités virtuelle et augmentée

■ Guillaume TERRASSE – Jean-François HULLO – Guillaume THIBAUT

C'est au cours d'un planning contraint lors des arrêts de fonctionnement des bâtiments réacteur qu'EDF réalise leurs travaux de maintenance. EDF a entrepris la réalisation de maquettes numériques "tel que construit" à partir d'acquisitions de photographies panoramiques, de nuages de points et leur reconstruction 3D. La quantité de données à exploiter est très importante avec au total, pour un unique bâtiment, 500 milliards de pixels, 30 milliards de points et 90 000 géométries 3D reconstruites. Dans cet article, nous présentons les travaux menés par EDF R&D pour enrichir cette donnée géométrique avec du contenu sémantique. Nous avons développé un outil, appelé Léonard qui détecte automatiquement des équipements sur des images à partir de deux algorithmes d'apprentissage profond. Une des contraintes à l'utilisation de ce type de méthode est qu'elle nécessite un nombre important d'exemples labélisés. Or, Léonard intègre dans sa chaîne de traitements une méthode capable de construire automatiquement cette base d'exemples labélisés. Pour cela, il exploite la maquette numérique dans laquelle les photographies sont recalées sur le modèle 3D. EDF étudie également l'utilisation de Léonard sur un nuage de points. Cela permettrait notamment d'automatiser la segmentation du nuage, étape actuellement principalement manuelle, et in fine une reconstruction 3D automatique. Pour EDF, ces travaux d'automatisation permettraient de réduire le temps de reconstruction 3D qui est actuellement prépondérant dans la création de la maquette numérique. De plus, la recherche automatique de contenus sémantiques dans des photographies permettrait également la réalisation d'inventaires rapides d'équipements présents dans des bâtiments complexes et de grande dimension. Les évolutions récentes des systèmes d'acquisition et de production rapide de données 3D vont démocratiser leur utilisation. De plus, de nouveaux moyens d'interaction Homme-3D grand public émergent. Nous présentons également les approches développées par EDF pour que les opérateurs des moyens de production d'électricité exploitent efficacement les nouvelles technologies de réalités virtuelle, augmentée ou mixte.

Introduction

Une centrale nucléaire est une installation industrielle complexe et de grande taille. Elle se compose de plusieurs unités de production qu'on appelle "tranches". Le parc de production d'électricité nucléaire d'EDF est constitué de

58 tranches. Chacune d'elle regroupe 10 bâtiments dont le bâtiment réacteur. Ce dernier est lui-même complexe :

- 15 étages et demi-étages,
- dense en équipements avec plus de 10 000 équipements fonctionnels, répartis dans près de 200 locaux,

MOTS-CLÉS

Maquette numérique, photographies, nuages de points, apprentissage automatique, réalité virtuelle, augmentée, mixte.

– de forme cylindrique avec un diamètre de 40 m et une hauteur de 45 m.

De plus, le bâtiment réacteur n'est accessible que lors des arrêts de production. C'est pendant ces périodes d'arrêt de fonctionnement de l'ordre de quelques semaines que de très nombreuses opérations de maintenance sont réalisées. Ainsi, la complexité des lieux couplée au planning contraint de travaux de maintenance demande une planification rigoureuse des interventions. Il est alors nécessaire de disposer des meilleures connaissances quant à la réalité des lieux. En effet, la réduction des incertitudes permet de mieux maîtriser les coûts et les plannings contraints. Pour ce faire, EDF a notamment entrepris la numérisation photo et topographique de ses installations en vue de la constitution de maquettes numériques. La consultation de celles-ci est alors réalisée au moyen d'un logiciel de visite virtuelle, développé sur mesure, appelé VVProPrépa. L'intervenant peut ainsi y consulter, en préparation des travaux de maintenance, la configuration de l'installation à partir de la maquette numérique :

- les plans masse 2D ;
- les photographies panoramiques 360° haute-définition ;
- les stations laser et le nuage de points consolidé. Les photographies panoramiques sont recalées sur le nuage de points ;
- le Modèle numérique 3D (M3D) construit au plus près du nuage de points.

La chaîne de production des données assure une maîtrise de la précision et de



l'exactitude géométrique des données (nuage de points et modèle 3D). La maquette numérique représente une quantité importante de données avec, pour un unique bâtiment réacteur :

- 1 100 photographies panoramiques, soit 500 milliards de pixels ;
- 1 100 stations scanner, soit 33 milliards de points ;
- 90 000 primitives géométriques 3D.

Ici, l'étape de reconstruction du modèle 3D est une procédure majoritairement manuelle. Actuellement, le reconstituteur identifie visuellement une géométrie dans le nuage de points, la segmente, puis choisit le type de géométrie (un cylindre par exemple) correspondant et enfin réalise l'ajustement de cette géométrie sur les points. Seule cette dernière étape est automatisée. La *Figure 1* illustre cette étape de reconstruction de géométries 3D à partir du nuage de points.

EDF R&D travaille depuis plusieurs années à faire évoluer la production de la maquette numérique pour que cette tâche soit facilitée et plus riche avec des informations métiers complétant les informations géométriques :

- La part du temps alloué à la reconstruction du modèle représente un peu moins des $\frac{3}{4}$ du temps global de traitement (incluant l'acquisition et le post-traitement). Pour un bâtiment réacteur, cela correspond approximativement à 10 mois de travail pour la reconstruction 3D avec plusieurs équipes travaillant en parallèle. Compte tenu du nombre de bâtiments présents

dans le parc nucléaire français, il est actuellement difficile de réaliser une reconstruction pour l'ensemble du parc. Nous souhaitons alors développer **une méthode automatique de reconstruction d'un modèle 3D à partir d'un nuage de points**.

- La modélisation 3D étant complexe, dans certains cas nous pourrions ne disposer que des photographies panoramiques et/ou du nuage de points sans son reconstruit 3D. Or cette donnée ne contient que très peu d'information sémantique. Nous souhaiterions alors l'enrichir en lui associant des informations métiers. Cela permettrait aux utilisateurs de faire facilement le lien entre une requête sur un équipement et sa représentation géométrique. C'est pourquoi, nous souhaitons **détecter automatiquement des équipements dans les photographies**. Exemple : un utilisateur peut alors rechercher l'ensemble des vannes et obtenir une liste des photographies panoramiques où au moins une vanne est présente.
- Dans le cadre de la déconstruction des centrales nucléaires, les bâtiments sont amenés à évoluer très rapidement. Les équipes ont besoin de réaliser régulièrement l'inventaire des équipements présents. Nous cherchons à détecter automatiquement les équipements encore présents à différentes étapes de la déconstruction pour suivre les travaux.

Afin de répondre à ces problématiques, EDF R&D (France et Royaume-Uni) a étudié des méthodes par apprentissage

de réseau de neurones profonds (*deep learning* en anglais). Nous présentons, dans la suite de l'article, l'outil baptisé Léonard ([1]) basé sur deux réseaux de neurones profonds. Léonard permet de détecter des équipements dans les photographies panoramiques par segmentation de pixels selon une liste prédéfinie d'objets. Plus récemment, des travaux ont adapté les réseaux de neurones pour les appliquer directement à des nuages de points. Nous verrons également dans la partie suivante des exemples pour appliquer ces technologies aux nuages de points et les difficultés que cela implique.

Dans la dernière partie de l'article, nous revenons sur nos usages des technologies récentes d'interaction 3D : réalités virtuelle, augmentée et mixte. Face aux promesses et fantômes technologiques, il nous semble important de proposer une vision de leur utilisation dans des usages professionnels d'ingénieurs et de topographes.

Réseaux de neurones profonds pour les maquettes numériques

Nous souhaitons faire évoluer la production de maquettes numériques en :

- accélérant les tâches fastidieuses, notamment la reconstruction du modèle 3D ;
- ajoutant de l'information sémantique aux photographies panoramiques dans le cas où nous ne disposerions pas du modèle 3D.

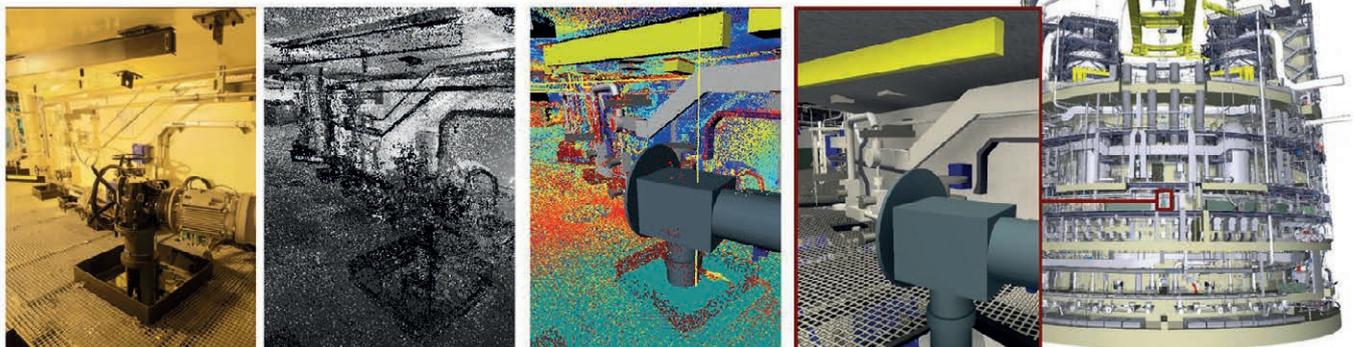


Figure 1. Illustration de la reconstruction d'un modèle 3D d'un bâtiment réacteur à partir d'un nuage de points. De gauche à droite : un extrait d'une photographie panoramique, un nuage de points, l'étape de reconstruction à partir du nuage de points et le modèle 3D reconstruit

Pour cela, nous avons mené des travaux tirant profit de réseaux de neurones profonds. Dans cette partie, nous introduisons les réseaux de neurones profonds. Puis, nous présentons Léonard qui allie deux réseaux de neurones profonds pour détecter automatiquement des équipements dans les photographies panoramiques. Finalement, nous verrons des travaux récents pour appliquer ces technologies aux nuages de points et quelles sont les difficultés liées à leur utilisation.

■ Les réseaux de neurones profonds

Les réseaux de neurones font partie de la famille des algorithmes d'apprentissage automatique. Dans cette famille, nous nous intéressons à l'apprentissage automatique supervisé, c'est-à-dire que l'on dispose d'une base de données labélisée pour estimer les paramètres d'une fonction de décision représentée par le réseau de neurones. Les paramètres sont estimés suivant une métrique qui minimise l'erreur de décision.

La Figure 2 illustre le principe de fonctionnement d'un réseau de neurones qui se décompose en deux phases : la phase d'apprentissage et la phase d'inférence. Dans cette figure, on prend l'exemple d'un réseau de neurones capable de prédire quel type d'équipement est présent dans une image. On dispose

d'une base de données labélisée qui contient des images d'équipements auxquelles sont associées leur dénomination (vanne, échelle...). Le réseau est composé d'un ensemble de neurones représentés par des cercles, et répartis en couches. Ici chaque neurone représente un paramètre de la fonction de décision que l'on cherche à estimer. Lors de la phase d'apprentissage, chaque exemple est utilisé en entrée du réseau de neurones qui produit une décision. S'il se trompe, les paramètres sont modifiés à hauteur de l'erreur de prédiction. Cette étape est répétée jusqu'à ce que la qualité de prédiction converge (i.e. l'erreur de prédiction soit inférieure à une valeur choisie). Une fois la phase d'apprentissage terminée, c'est-à-dire que les paramètres du réseau sont établis, le réseau est utilisé sur des images non labélisées : c'est la phase d'inférence. Pour une image, le réseau fournit un score entre 0 et 1 pour chaque classe. La prédiction finale correspond alors à la classe ayant obtenu le score le plus élevé.

Depuis 2009, les réseaux de neurones profonds ont amélioré de nombreuses méthodes précédemment à l'état de l'art dans des domaines variés. Ces progrès rapides sont la conjonction de l'apparition de trois nouveautés dans la communauté des sciences des données :

- calculs accélérés par GPU ;
- bases de données labélisées publiques pour la phase d'apprentissage (exemple : Imagenet [2]) ;
- codes et bibliothèques de développement accessibles.

La force des réseaux de neurones dits profonds réside dans leur capacité à décomposer l'information et trouver des invariants, c'est-à-dire des redondances dans les images, qui vont leur permettre d'être robustes aux légers changements et, par extension, d'exprimer les concepts sous-jacents d'une classe. Cette capacité se trouve dans les filtres qu'ils apprennent. Ces filtres correspondent à un ensemble de paramètres interdépendants formant une vignette. Ils sont mis en correspondance avec l'image dans le réseau. C'est le nombre de similarités entre les filtres et l'image qui permet d'en déduire une prédiction. Par exemple, l'image d'un visage n'est plus uniquement un ensemble de pixels mais la combinaison de concepts propres aux visages : ses contours, les yeux, le nez, etc. La Figure 3 illustre la capacité d'un réseau de neurones à apprendre à identifier ces concepts pour la reconnaissance d'un visage. Les filtres appris dans les premières couches correspondent aux concepts généraux, c'est-à-dire des contours. Plus le réseau est profond, plus ce dernier apprend des concepts spécifiques aux visages,

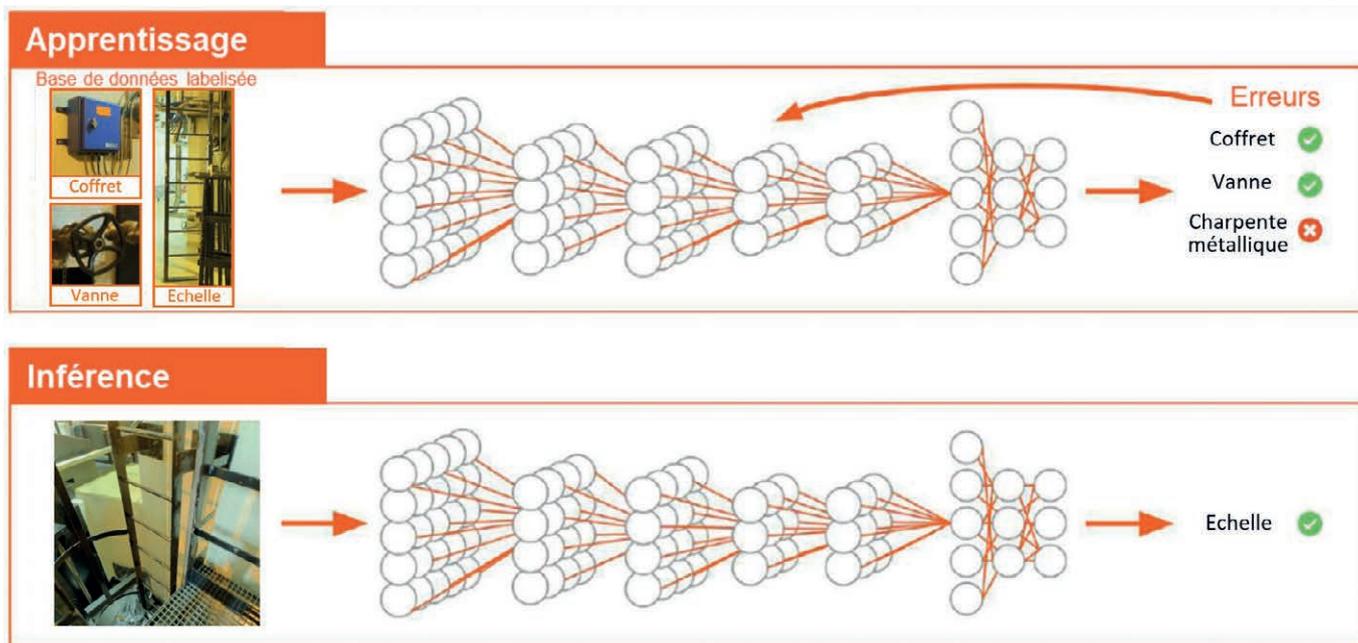


Figure 2. Illustration du fonctionnement d'un réseau de neurones lors de deux phases, d'abord la phase d'apprentissage (figure du haut) puis la phase d'inférence (figure du bas)

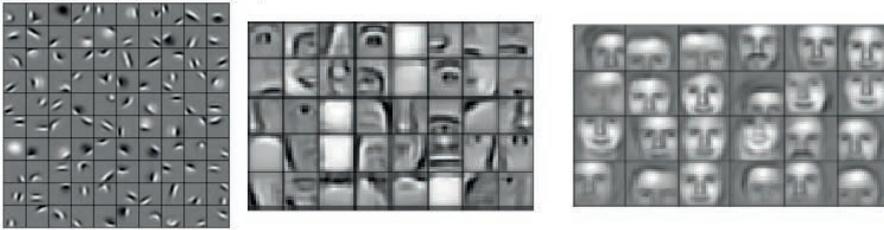


Figure 3. Décomposition de l'information dans l'image selon la profondeur des couches dans le réseau de neurones. De gauche à droite, des exemples de filtres issus des premières couches jusqu'aux couches les plus profondes (extrait de [3])

c'est-à-dire les yeux ou le nez, jusqu'au concept de visage.

En contrepartie, plus un réseau est profond et plus le nombre de paramètres à estimer est important. La complexité d'apprentissage croît alors rapidement. Par exemple, le réseau Resnet ([4]), l'un des plus utilisés et performants, dispose de 152 couches pour environ 60 millions de paramètres à estimer. Par conséquent, il est nécessaire de disposer d'une base de données conséquente labélisée, afin d'assurer une estimation fiable des paramètres.

■ Les réseaux de neurones profond pour des photographies d'intérieur de bâtiments industriels

Nous souhaitons détecter automatiquement les équipements présents dans une image. La Figure 4 présente, à gauche, un extrait de photographie panoramique et à droite, le résultat souhaité de la prédiction par un réseau de neurones. Dans ce résultat souhaité, les pixels sont associés à une classe d'équipement. Ici, nous avons

une détection des équipements au pixel près : on parle de segmentation sémantique. Si de plus la segmentation distingue deux objets d'une même classe qui se chevauchent, on parle de segmentation par instance.

Les acteurs majeurs de la communauté en science des données publient leurs codes en open source. Néanmoins, la complexité des réseaux de neurones nécessite une base de données importante, afin que ceux-ci puissent apprendre les concepts nécessaires à la généralisation. Or, comme expliqué en introduction, EDF dispose d'une quantité importante de données au travers des maquettes numériques de ses bâtiments réacteurs. Les photographies panoramiques sont recalées sur le nuage de points et donc sur le modèle 3D. Or les géométries qui composent le modèle 3D contiennent des informations sur la nature des équipements : échelle, sol, vanne, éclairage, etc. À partir du modèle 3D, on peut construire des images synthétiques dont les pixels correspondent à une classe (Figure 4 image de droite). On appelle ces images

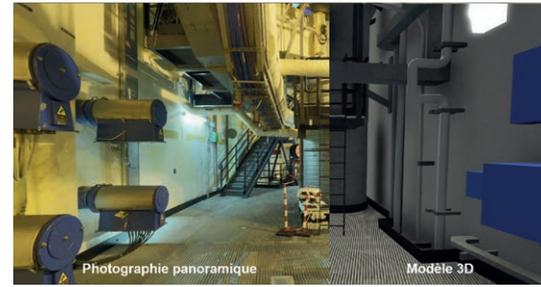
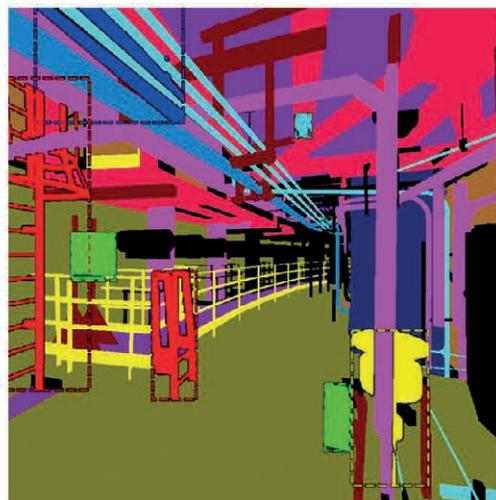
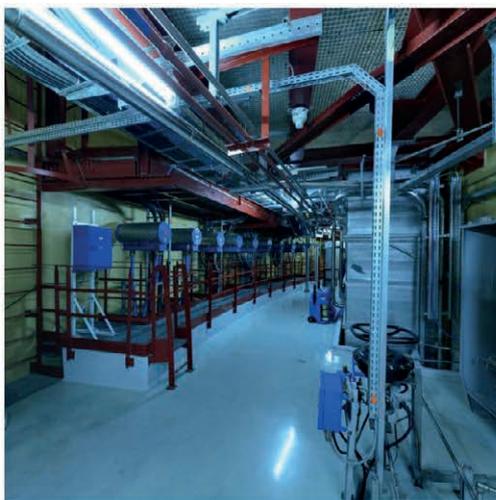


Figure 5. Illustration du recalage des photographies panoramiques avec le modèle 3D de la maquette numérique

des "masques". De plus, on peut associer directement les masques à une photographie panoramique du fait de leur recalage. La Figure 5 illustre le recalage entre les photographies panoramiques et le modèle 3D. EDF R&D a mis au point une chaîne de traitement complète appelée Léonard intégrant la création automatique d'une base de données labélisée pour la segmentation par instance à partir de la maquette numérique jusqu'à l'apprentissage et l'inférence du réseau de neurones.

La création de masques du modèle 3D pour des points de vue correspondant aux positions des photographies panoramiques est assurée à partir d'une image de rendu sans ombre de la scène 3D (à partir du logiciel Unity). Pour chaque pixel de l'image de rendu, on encode la classe de l'équipement associé dans le premier canal (le canal rouge par exemple) et un identifiant dans les deux canaux suivants (les canaux vert et bleu). La Figure 6 présente une image de rendu sans ombre et le masque



- Echelle
- Charpente métallique
- Sol caillebotis
- Tuyauterie
- Garde-corps
- Vanne
- Boîtier
- Ventilation
- Eclairage
- Sol béton

Figure 4. Résultat souhaité par le réseau de neurones profond, ici la segmentation par instance. À gauche un extrait d'une photographie panoramique, à droite la classe d'équipement associée à chaque pixel.

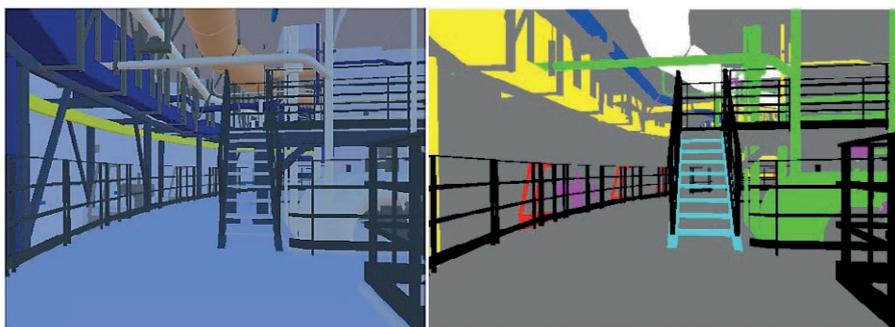


Figure 6. Création de masque à partir du modèle 3D. À gauche : une image de rendu sans ombre ; à droite : le masque obtenu à partir de l'encodage des classes d'équipement dans chaque pixel.

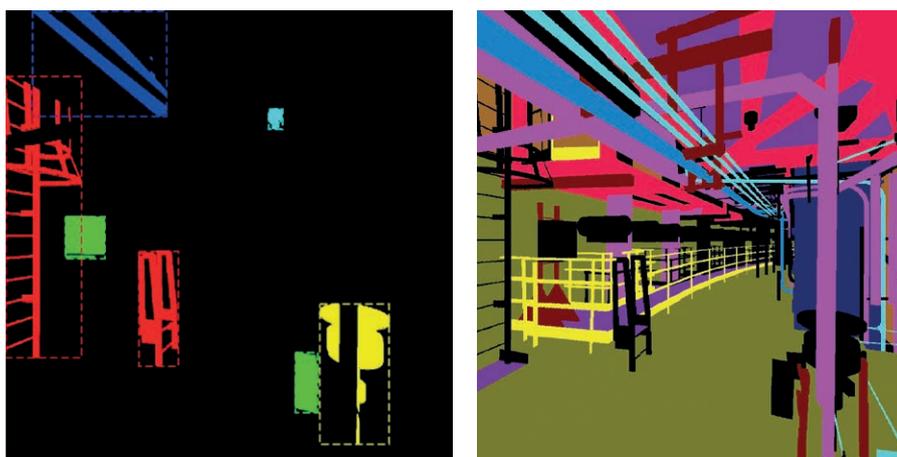


Figure 7. Illustration de la séparation en deux typologies de classe, *cars* et *roads*, à partir de l'exemple de masque donné en Figure 4

obtenu à partir de l'encodage des classes d'équipement.

Maintenant que nous disposons d'une base de données labélisée, nous avons besoin d'une méthode de segmentation par instance. Pour cela, Léonard scinde le problème de segmentation en deux :

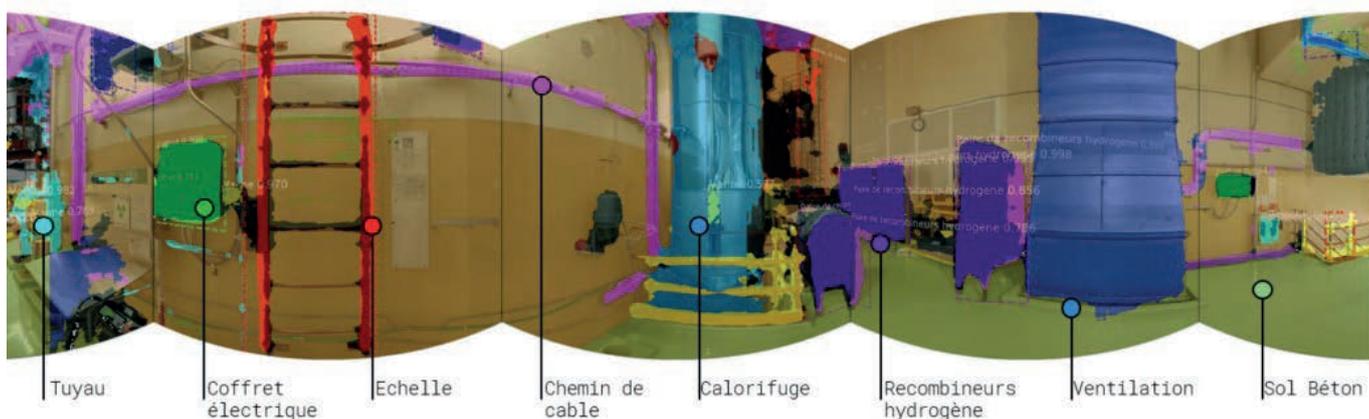
- la segmentation des *cars* : les objets de dimensions finies dans l'image, telles les vannes, échelles, etc.
- la segmentation des *roads* : les objets de dimensions non finies, comme les tuyauteries, sols, etc.

Pour chaque problème, Léonard bénéficie de réseaux de neurones publiés dans la communauté scientifique : Mask R-CNN ([4]) pour la segmentation des *cars* et DeepLab ([5]) pour la segmentation des *roads*. Le premier est un réseau de neurones pour la segmentation par instance alors que le second réalise une segmentation sémantique.

Nous avons réalisé des essais sur une machine dédiée au jeu vidéo (CPU Intel Core I7, GPU NVIDIA GeForce GTX1070). Sur celle-ci, le temps consacré à la génération de la base de données labélisée pour un bâtiment



Figure 8. Résultat qualitatif de la détection par Léonard des objets *cars* et *roads*



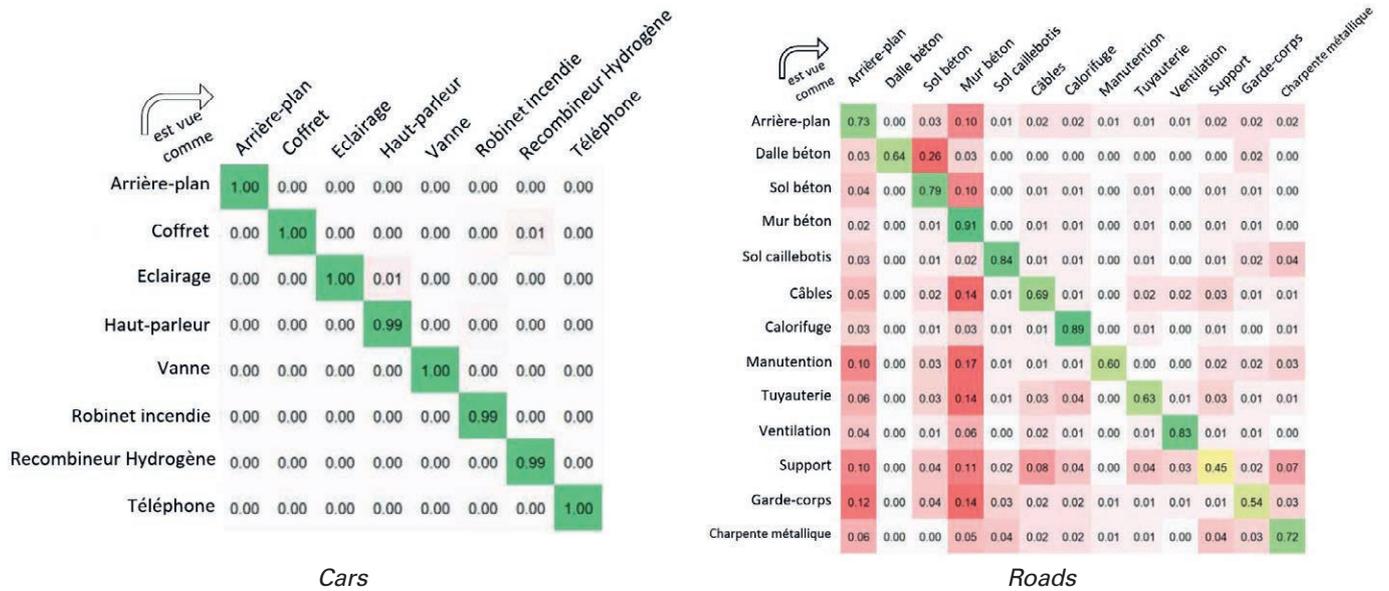


Figure 9. Matrices de confusion des ratios de prédiction de chaque classe pour la segmentation des cars et des roads

avec 1 255 photographies panoramiques est de 55 heures. La phase d'apprentissage pour les objets cars avec Mask R-CNN est de 61 heures et 42 heures pour les objets roads avec DeepLab. La Figure 8 présente un résultat qualitatif de détection des cars et roads par Léonard. Nous avons également évalué quantitativement les performances de Léonard sur 30 % de la base de données, les 70 % complémentaires ayant été utilisés pour la phase d'apprentissage. La Figure 9 présente deux matrices de confusion résumant les performances des deux méthodes de segmentation. La première méthode (figure de gauche) présente les ratios de prédiction selon la boîte englobante (et non en termes de pixel). On constate que Léonard ne

mélange que très rarement les classes cars. Néanmoins, cette matrice de confusion ne répertorie pas les objets non détectés. La figure de droite présente la matrice de confusion des prédictions pour les objets roads. On constate que certaines classes comme les murs en béton et les calorifuges sont très bien détectées par Léonard avec un ratio proche des 0.90. D'autres classes, comme les supports et les garde-corps, présentent des scores plus faibles avec respectivement des ratios de bonne prédiction de 0.45 et 0.54. Cela peut s'expliquer par la forme allongée de ces équipements : les erreurs de prédiction se produisent à la frontière des objets qui sont alors prédits comme arrière-plan ou mur en béton dans respectivement 10 % des cas environ.

Enfin, sur la Figure 10, nous présentons un graphe de la répartition des scores de prédiction en fonction du score de recouvrement entre la boîte englobante prédite et la boîte englobante de la vérité terrain (IoU, Intersection of Union). Pour le cas de l'éclairage, on constate que lorsque Léonard est confiant sur sa prédiction (score élevé), il se trompe rarement (peu d'IoU à zéro). Dans le cadre des échelles, on constate que dans certains cas Léonard se trompe : IoU à zéro et un score de prédiction > 0,5. Toutefois dans ces cas-là, son score de prédiction associé est généralement faible. Les fausses détections pourront alors être facilement filtrées en rehaussant les exigences sur le score de détection (à 0.8 par exemple).

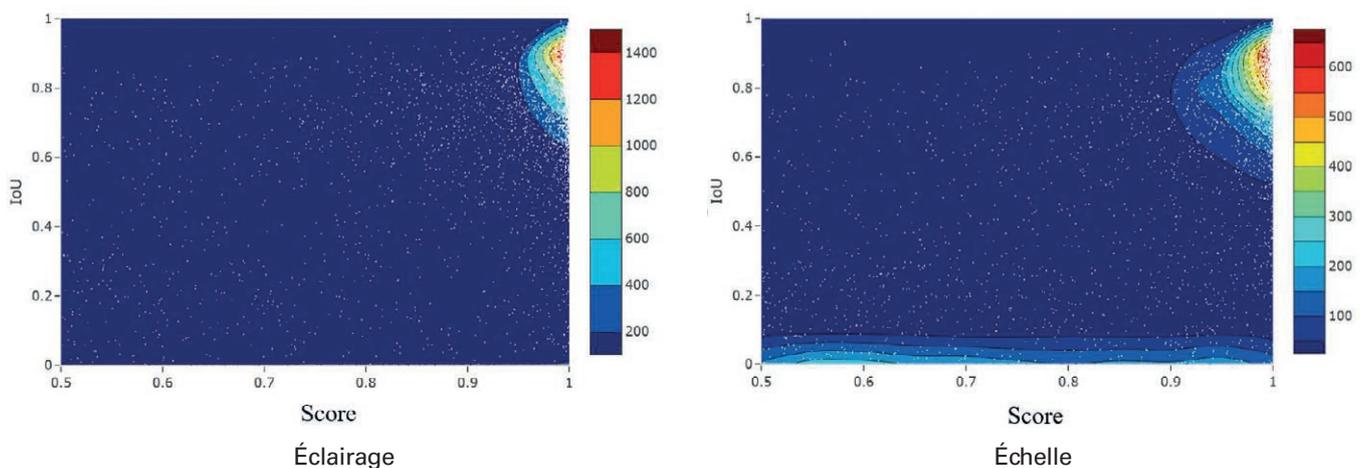


Figure 10. Répartition des scores de prédiction en fonction de l'indice de l'IoU (indice de recouvrement entre la boîte englobante prédite et la boîte englobante vraie) pour la segmentation des cars.



■ Le passage aux nuages de points

Nous avons vu que les méthodes de réseaux de neurones profonds s'appliquent avec succès aux images. Comme dans le cas de Léonard, certaines classes de réseaux réalisent des prédictions au niveau du pixel d'une image. Nous pourrions alors envisager de transposer ces réseaux aux nuages de points, afin de les segmenter automatiquement. Néanmoins dans le cadre des nuages de points, si l'information est plus riche, l'apprentissage en est plus complexe. La force des réseaux de neurones est leur capacité à décomposer l'information à partir de filtres, c'est-à-dire à trouver des motifs redondants dans des voisinages de l'image. Dans une image, l'expression d'un voisinage est explicite : le voisin d'un pixel se trouve à la même position incrémentée de 1 selon la ligne ou la colonne du tableau de pixels. Or, pour les nuages de points, l'expression d'un voisinage est moins directe. Cela suppose de calculer une distance entre le point d'intérêt et les autres points constituant le nuage. On comprend alors ici qu'il sera plus difficile aux réseaux de neurones de décomposer efficacement la structure du nuage de points. Pour pallier ce problème, plusieurs travaux ont tenté de se replacer dans un cadre proche de celui du traitement d'images en ayant une expression plus explicite d'un voisinage. Par exemple, [7] exprime le nuage de points sous forme d'une série d'images synthétiques pour plusieurs poses ou encore [8] applique les réseaux de neurones à des voxels. Certains travaux, comme [9], utilisent quant à eux une structure en graphe.



Néanmoins, ce n'est que récemment que la communauté d'apprentissage automatique a montré un fort intérêt pour les nuages de points. En 2017, une équipe de Stanford a notamment présenté une avancée importante dans la segmentation automatique de nuage de points avec l'algorithme PointNet ([10]) qui exploite directement le nuage de points sans transformation en amont des données (pour identifier des voisins). En 2019, de nombreux travaux sur la segmentation automatique de nuage de points ont été présentés dans les plus importantes conférences de vision par ordinateur et apprentissage automatique (CVPR, ICCV, etc.).

Nous avons appliqué Léonard aux nuages de points. Dans un premier temps, nous l'avons appliqué à des images synthétiques générées à partir de nuage de points (de manière analogue à [7]). Cette étape est réversible. Nous pouvons exprimer le nuage de points sous forme d'image synthétique et ensuite retrouver un nuage de points à partir de l'image synthétique. Ainsi, les résultats de Léonard obtenus sur une image synthétique peuvent être convertis en nuage de points. La Figure 11 présente un exemple d'un nuage de points et une image synthétique obtenue à partir de ce même nuage de points.

Les premiers résultats obtenus avec Léonard sont relativement intéressants (Figure 12). Néanmoins, les performances sont en deçà de celles obtenues sur des photographies panoramiques. Il y a plusieurs éléments qui pourraient expliquer cela :

- La dynamique et le bruit de l'image synthétique sont différents de ceux des photographies panoramiques.



Figure 11. Génération d'une image synthétique à partir d'un nuage de points.

À gauche : un nuage de points en vue du dessus.

À droite : une image synthétique obtenue à partir du nuage de points

Léonard n'est pas encore assez robuste pour supporter ces changements. Un apprentissage spécifique sur ces données est nécessaire.

- La complexité des lieux. Les environnements industriels sont très complexes avec une diversité importante de géométries d'objets. Certains sont compacts, d'autres imposants ou avec une forte asymétrie (ex. : tuyauterie). De plus dans les nuages de points, les objets sont parfois moins bien définis. Certains objets sont incomplets avec peu de points (notamment les objets avec un matériau noir brillant).

Une des principales différences avec l'approche image est la décomposition efficace de l'information. Toutefois, cette différence est en train d'être comblée par les nombreuses équipes de chercheurs travaillant sur le sujet. Un autre aspect qui est actuellement très différent entre la communauté Image et Nuage de points est le manque de bases de données publiques pour entraîner les réseaux. Actuellement, il existe plusieurs bases de données publiques comme ModelNet ([11]) ou ShapeNet ([12]). Toutefois, ces dernières sont assez limitées et composées d'environnements simples issus du tertiaire. La complexité des environnements industriels et l'absence d'une base de données publique d'environnements industriels expliquent pourquoi actuellement il n'existe pas de réseaux adaptés aux nuages de points dont les performances soient comparables aux applications basées image.

Réalités virtuelle, augmentée et mixte

Les méthodes récentes d'interprétation de données exploitent aujourd'hui de nouveaux algorithmes d'apprentissages profonds ainsi que des cartes graphiques, utilisées pour les calculs vectoriels, de plus en plus performantes. Un bon nombre de tâches, comme les inventaires de matériels, sont en passe d'être complètement automatisées. Mais une majorité des usages de ces grands environnements 3D reposent encore sur l'exploration et l'interaction par un opérateur humain. Depuis des décennies, les opérateurs spécialisés se sont formés à utiliser

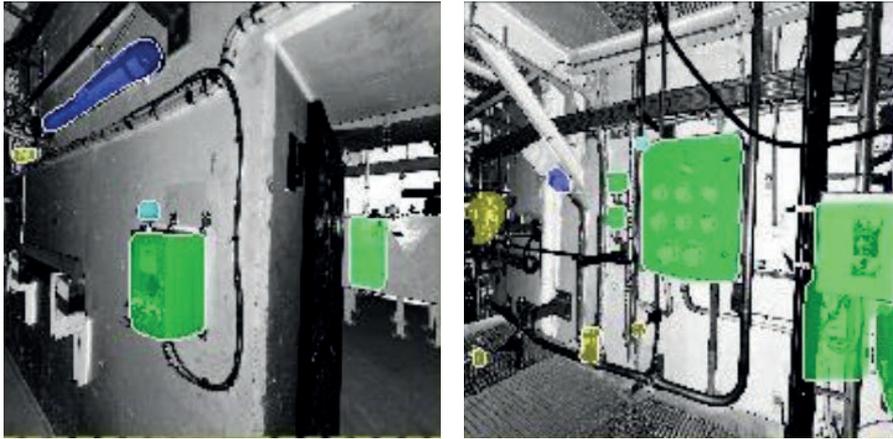


Figure 12. Deux exemples des résultats de Léonard sur des images synthétiques obtenus à partir de nuages de points

efficacement des périphériques d'interaction avec un écran, souris, claviers puis manettes. Le recyclage des capacités cognitives dans ces interactions a un coût important d'appropriation (plusieurs années) et a des limites claires : faible immersion, pas d'interaction *in situ*, fatigue visuelle. De nouveaux systèmes de visualisation et d'interaction, pour la plupart issus du jeu vidéo, forment, avec les écrans classiques, une panoplie d'outils disponibles pour l'interaction homme-3D. On distingue 3 grandes familles d'équipements, qui sont à la 3D ce qu'une carte est à la 2D :

- réalité virtuelle : casques opaques, avec suivi 3D de la tête, parfois des mains via l'intermédiaire de contrôleurs,
- réalité augmentée : utilisation du flux caméra d'un smartphone ou d'une tablette pour le sur-affichage

in situ de modèles numériques, qu'il s'agisse d'informations géolocalisée ou de modèles géométriques 3D,

- réalité mixte : casque semi-transparent de réalité augmentée, à l'intersection entre réalités virtuelle et augmentée.

Chacune de ces technologies trouve un champ d'usage bien distinct et aura des cas d'usages particuliers. En raison du coût (complexité de mise en œuvre et coût cognitif pour l'utilisateur) de ces technologies, il est crucial de s'assurer du besoin de leur recours. L'approche communément développée à EDF, et illustrée dans la *Figure 13*, est centrée sur les tâches que l'utilisateur cherche à réaliser, et ses contraintes d'utilisation ; elle consiste d'abord à répondre aux questions suivantes, avant de poursuivre une démarche de développement aux côtés de designers et spécialistes de la cognition : la tâche peut-elle être

réalisée sur un écran classique ? Si oui, le recours à d'autres moyens est inutile. Si non, la tâche nécessite-t-elle une immersion à échelle humaine, avec prise en compte des déplacements de l'utilisateur ?

- S'il s'agit d'une interaction active, avec un environnement purement numérique : réalité virtuelle souhaitable.
- S'il est nécessaire de superposer et interagir une information 2D ou 3D à un environnement existant complexe : réalité mixte.
- Si l'utilisateur peut utiliser ses deux mains, que l'environnement est géométriquement simple et que l'affichage est informatif : réalité augmentée souhaitable.

Conclusion

Les progrès technologiques dans le domaine de la numérisation permettent de réaliser des acquisitions de photographies et de nuage de points, recalées entre elles, à grande échelle dans un planning contraint. Toutefois la grande quantité de données acquise souffre du manque d'information sémantique. Le passage de cette donnée géométrique à une modélisation 3D labélisée est encore fastidieux avec des traitements essentiellement manuels. Néanmoins, les avancées spectaculaires dans le domaine de l'apprentissage automatique offre l'opportunité d'enrichir efficacement les données géométriques avec des infor-



Hinkley Point C in Mixed reality
Réalité Mixte



Revue d'ingénierie d'un design réacteur
Écran 3D



Visualisation de fermes éoliennes
Réalité Augmentée



Visualisation de vibrations de turbines
Réalité Mixte



Serious game « build the reactor »
Réalité Virtuelle



Suivi de procédures in-situ
Réalité Mixte

Figure 13. Exemples récents d'utilisation des technologies de réalités virtuelle, augmentée et mixte à EDF.



mations sémantiques. L'apprentissage automatique montre d'excellents résultats sur les images, mais peine encore à avoir les mêmes performances sur les nuages de points. Les communautés académique et industrielle ont récemment publié de nombreux travaux sur le traitement automatique du nuage de points de manière analogue aux travaux en Image (conférences renommées (CVPR, ICCV, etc.), codes open-source, base de données publique).

Les réalités virtuelle, augmentée ou mixte offrent de nouveaux moyens de visualisation et d'interaction avec la 3D. Tout comme la CAO et les scanners laser auparavant, les topographes doivent se saisir de ces nouvelles technologies qui vont changer la profession. Les nouveaux moyens de numérisation laser combinés aux avancées de l'apprentissage automatique profond vont démocratiser encore plus l'usage de la 3D. Dans ce contexte, l'interaction Homme-3D aura évolué avec le développement de nouvelles technologies de visualisation 3D. Le topographe restera le garant de la mesure de par sa connaissance du terrain et des outils qu'il utilise pour numériser son environnement. ●

Remerciements

Les auteurs souhaitent remercier Adrien NIVAGGIOLI qui a participé au développement de Léonard.

Contact

Guillaume TERRASSE - EDF Lab Saclay
guillaume.terrasse@edf.fr

Jean-François HULLO - EDF Energy R&D UK Centre
jean-francois.hullo@edfenergy.com

Guillaume THIBAUT - EDF Lab Saclay
guillaume.thibault@edf.fr

Bibliographie

[1] Nivaggioli, A., Hullo, J. F., & Thibault, G. (2019). *Using 3d models to generate labels for panoptic segmentation of industrial scenes*. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 4.

[2] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). *Imagenet: A large-scale hierarchical image database*. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255).

- [3] Lee, H., Grosse, R., Ranganath, R., & Ng, A. Y. (2011). *Unsupervised learning of hierarchical representations with convolutional deep belief networks*. Communications of the ACM, 54(10), 95-103.
- [4] He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
- [5] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). *Mask r-cnn*. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).
- [6] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). *DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs*. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848.
- [7] Su, H., Maji, S., Kalogerakis, E., & Learned-Miller, E. (2015). *Multi-view convolutional neural networks for 3d shape recognition*. In Proceedings of the IEEE international conference on computer vision (pp. 945-953).
- [8] Maturana, D., & Scherer, S. (2015, September). *Voxnet: A 3d convolutional neural network for real-time object recognition*. In 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 922-928). IEEE.
- [9] Defferrard, M., Bresson, X., & Vandergheynst, P. (2016). *Convolutional neural networks on graphs with fast localized spectral filtering*. In Advances in neural information processing systems (pp. 3844-3852).
- [10] Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). *Pointnet: Deep learning on point sets for 3d classification and segmentation*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 652-660).
- [11] Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., & Xiao, J. (2015). *3d shapenets: A deep representation for volumetric shapes*. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1912-1920).
- [12] Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z. & Xiao, J. (2015). *Shapenet: An information-rich 3d model repository*. arXiv preprint arXiv:1512.03012.

ABSTRACT

Key words: Digital model, photographs, point clouds, deep learning, virtual, augmented or mixed realities.

Following a busy schedule, EDF carries out its maintenance work during reactor building operating interruption. EDF undertook the implementation of an "as-built" digital model based on the acquisition of panoramic photographs, point clouds and their 3D model reconstruction which represents a total of 500 billion pixels, 30 billion points and 90,000 reconstructed 3D geometries for a single building. The amount of data to be exploited is very large. In this article, we present the work carried out by EDF R&D to enrich this geometric data with semantic content. We have developed a tool, called Léonard, which automatically detects equipment on images based on two deep learning algorithms. One of the constraints to the use of this type of method is that it requires a large number of labelled examples. Yet, Léonard also includes in its processing chain a method to automatically generate the training set. To do this, it leverages the digital model in which photographs are registered on the 3D model. EDF is also studying the implementation of Léonard on a point cloud. This would notably allow the automation of the cloud segmentation, which is currently mainly a manual step and ultimately an automatic 3D reconstruction. For EDF, this automatization work would reduce the 3D reconstruction time that is currently predominant in the production of the digital model. In addition, the automatic search for semantic contents in photographs would also enable rapid equipment inventories in large and complex buildings to be carried out. Recent developments in rapid acquisition devices and production of 3D data will democratize their use. In addition, new devices for human-3D interaction for the general public are emerging. We also present the approaches developed by EDF to enable operators to effectively use these new technologies of virtual, augmented or mixed realities.