Computer Vision : Introduction au calcul du volume des objets

La Revue 3E.I
Culture Sciences de l'Ingénieur

Simon PLAYE[1]

Édité le 12/03/2025

[1] *Data Scientist chez Sicara (Theodo Data & AI)*

Cette ressource fait partie du N°115 de La Revue 3EI du deuxième trimestre 2025.

Avez-vous déjà travaillé sur le calcul du volume des objets à partir de vidéos en utilisant l'IA ? Cela s'est-il avéré soit peu pertinent, soit extrêmement difficile à réaliser ? Si vous avez répondu « oui » à ces deux questions et que vous êtes passionné par le calcul du volume ou l'IA, cet article est fait pour vous. En utilisant deux vidéos, je vais vous montrer comment calculer le volume des objets.

Les cas d'utilisation concernent principalement la gestion des stocks. Alors que de plus en plus d'entreprises cherchent à réduire leur impact environnemental, un tel outil peut permettre d'améliorer l'efficacité des chaînes d'approvisionnement et optimiser la gestion des stocks. Cet outil peut également être utilisé pour le placement de meubles : design d'intérieur, planification d'événements, construction ou rénovation...

Le calcul du volume des objets soulève plusieurs défis :

- Localiser les objets dans la vidéo
- Définir leurs limites
- Mettre à l'échelle la vidéo pour convertir une distance en pixels, en mètres

Cet article passera en revue toutes les étapes ci-dessous pour expliquer comment effectuer le calcul du volume des objets :

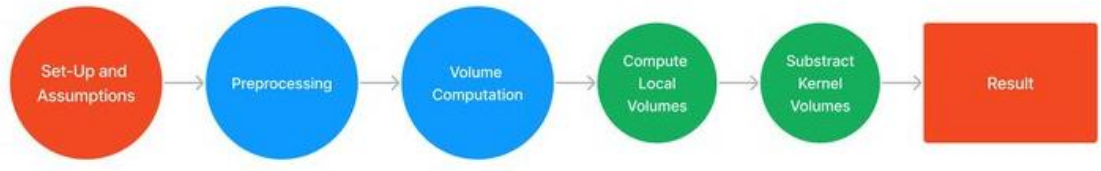

Figure 1 : Étapes du processus de calcul du volume.

1 - Configuration sous-jacente et hypothèses

Pour effectuer le calcul du volume des objets, je m'appuie sur quelques hypothèses :

- L'espace filmé est une pièce fermée.
- Cette pièce est filmée deux fois depuis le même point de vue : une fois sans les objets et une fois avec eux.
- Les vidéos sont identiques, à l'exception des objets.
- Les objets sont placés directement contre un mur, sans aucun espace.

Par souci de simplicité, dans le reste de cet article, la pièce filmée est vide, et les objets sont des boîtes rectangulaires. Par conséquent, le calcul du volume sera effectué sur ces boîtes. Voici un exemple de séquence vidéo respectant ces hypothèses :

Figure 2 : Un exemple de séquence vidéo respectant ces hypothèses. Ici, les objets étudiés sont les deux boîtes.

L'outil le plus important pour effectuer le calcul du volume est la caméra. En effet, une caméra 2D normale n'affiche que des valeurs de couleur pour tous les pixels filmés. Dans cet article, la configuration consiste en une caméra spéciale nommée Intel RealSense D435. Cette caméra est composée de deux caméras infrarouges qui offrent une représentation 3D de ce qui est filmé. Une API Python, pyrealsense2, permet de récupérer, pour chaque pixel, une coordonnée 3D. Ainsi, il est possible d'obtenir la distance en mètres par rapport à la caméra sur l'axe des x (de droite à gauche), sur l'axe des y (de haut en bas), et sur l'axe des z (de près à loin) pour tous les pixels. N'importe quelle autre caméra peut être utilisée tant qu'elle fournit ces coordonnées.

1.1 - Quelques mots sur pyrealsense2

Pyrealsense2 est l'enveloppe Python pour le SDK Intel RealSense 2.0, qui est une bibliothèque pour les caméras Intel RealSense. Le SDK peut normalement être utilisé avec librealsense, un package C++. Selon la page Github de librealsense : « Le SDK permet la diffusion de profondeur et de couleur, et fournit des informations sur la calibration intrinsèque et extrinsèque. La bibliothèque offre également des flux synthétiques (nuage de points, profondeur alignée avec la couleur et vice-versa), ainsi qu'un support intégré pour l'enregistrement et la lecture des sessions de diffusion. » Pour fournir ces services, le SDK utilise un modèle déterministe qui repose sur les entrées des deux caméras infrarouges.

1.2 - Prétraitement

Entrons maintenant dans le processus de calcul du volume. Voici la profondeur des images de ces deux vidéos :

Figure 3a : Image brute avec boîte

Figure 3b : Image brute sans boîte

Ces images illustrent un problème statistique classique : les valeurs aberrantes (outliers) :

- Les zones noires sur les images des vidéos correspondent à des pixels sans coordonnées (la profondeur est définie à 0).
- Les zones colorées correspondent à des coordonnées erronées (la profondeur est de 10 mètres, alors que le mur est à 1,5 mètre de la caméra).

1.3 - Traiter les valeurs aberrantes : bonnes pratiques

Tout d'abord, définissons plus précisément ce qu'est une valeur aberrante. Ici, une valeur aberrante est un pixel dont la valeur de la coordonnée z (correspondant à la profondeur) est égale à 0 ou supérieure à 1,6 mètre. Puisque le mur est à environ 1,5 mètre de la caméra, aucun pixel ne peut avoir une valeur z supérieure à 1,6 mètre.

Cette méthode permet également de sélectionner les pixels ayant des valeurs aberrantes pour les coordonnées x et y. En effet, un contrôle rapide montre que les pixels avec une valeur de coordonnée z égale à 0 ou supérieure à 1,6 ont aussi des valeurs aberrantes pour les coordonnées x et y. En revanche, aucun pixel n'a été trouvé avec des coordonnées x et y aberrantes, mais une valeur de coordonnée z comprise entre 0 et 1,6. Par conséquent, seule la valeur de la coordonnée z définit si un pixel est une valeur aberrante.

Prenons un pixel spécifique d'une image d'une vidéo et imaginons que ce pixel soit une valeur aberrante.

La meilleure manière de remplacer la valeur de la coordonnée z est d'utiliser la valeur de la coordonnée z du pixel non aberrant le plus proche. Ici, le « pixel non aberrant le plus proche » est le pixel ayant la distance euclidienne la plus faible par rapport à notre pixel.

La méthode utilisée pour récupérer les coordonnées x et y est plus compliquée. En nous concentrant d'abord sur x, utilisons la grille ci-dessous pour visualiser la méthode employée. Cette grille donne les coordonnées x des pixels situés sur une portion 7x7 de l'image. Les cellules rouges représentent des valeurs aberrantes, les cellules vertes représentent des pixels non aberrants. La valeur de la coordonnée x est écrite à l'intérieur des pixels non aberrants.

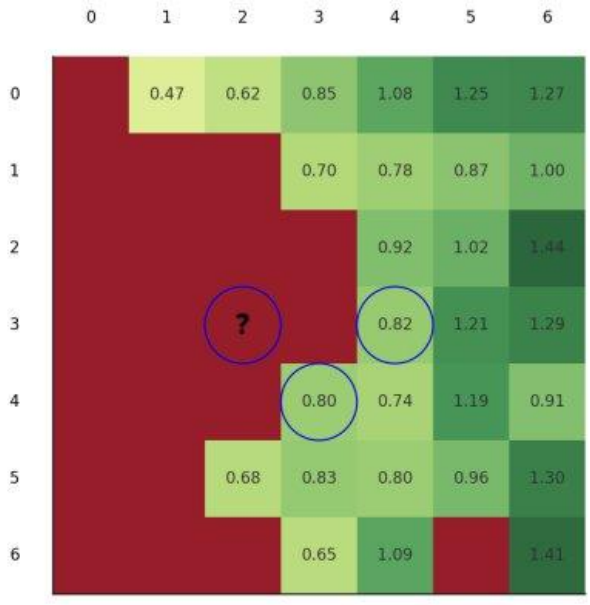

Figure 4 : Sous-ensemble de l'image : les zones rouges sont des valeurs aberrantes, les zones vertes sont des pixels avec des valeurs

Ici, je souhaite calculer la coordonnée x pour l'outlier entouré « ? » situé en (3,2). Pour ce faire, j'ai d'abord sélectionné les deux points les plus proches de cet outlier (entourés dans la grille) qui ne se trouvent pas dans la même colonne. La différence de coordonnée x entre ces deux pixels est 0.82 - 0.80 = 0.02. Ces deux pixels sont distants d'une colonne. Par conséquent, on peut estimer qu'en se déplaçant d'une colonne, la valeur de x change de 0.02. Ainsi, puisque l'outlier est situé à deux colonnes de ce pixel entouré (0.82), sa valeur estimée est 0.78 (= 0.82 - 0.02 * 2).

La même méthode est utilisée pour remplacer les valeurs des coordonnées y pour les outliers.

Ces méthodes de prétraitement permettent d'obtenir une bien meilleure représentation 3D de ce que la caméra a filmé (ici pour la profondeur) :

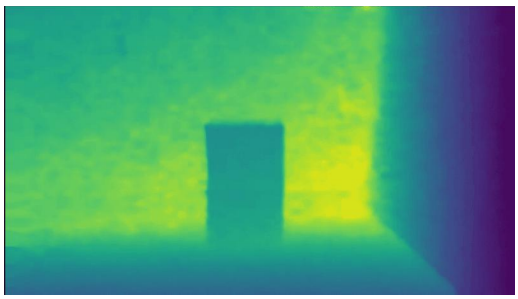

Figure 5a : Image brute avec boîte

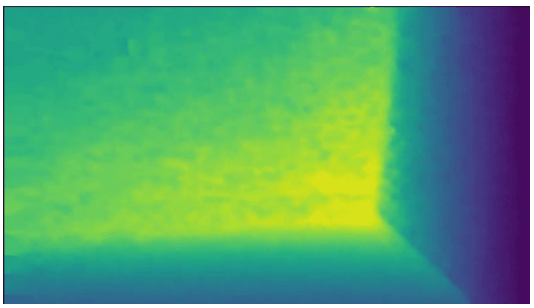

Figure 5b : Image brute sans boîte

2 - Processus de calcul du volume

2.1 - Technique pour effectuer le calcul du volume à l'aide de deux vidéos

En disposant de vidéos avec des coordonnées 3D cohérentes, je peux effectuer le calcul du volume. En lisant cet article, vous vous êtes peut-être demandé : « Pourquoi insister sur l'enregistrement d'une vidéo sans boîtes alors que seul le volume des boîtes est nécessaire ? » L'astuce pour effectuer le calcul du volume des boîtes est de calculer le volume d'une image de la vidéo avec les boîtes, puis de soustraire ce volume du volume de la même image sans boîtes. Cette différence correspond au volume calculé des boîtes.

2.2 - Calcul des volumes locaux à l'aide de noyaux

Comment effectuer le « calcul du volume » sur une seule image ? En regardant l'image ci-dessous, on peut la diviser en petits carrés, appelés noyaux :

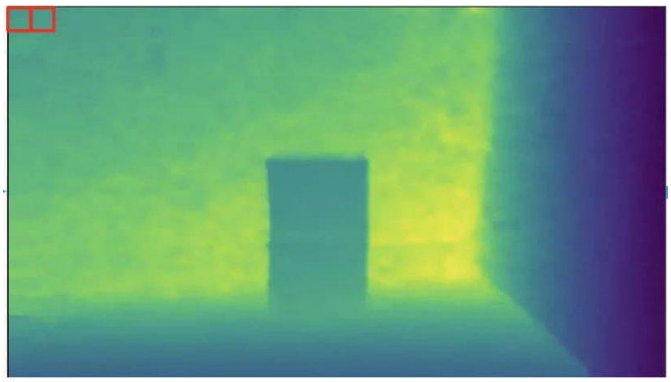

Figure 6 : Image traitée avec boîtes et noyaux

Simplifions ce problème en commençant par essayer de calculer les aires des noyaux, puis leurs volumes correspondants.

Ayant les coordonnées 3D des quatre pixels qui composent un noyau, on peut calculer l'aire du noyau de différentes manières. Pour être plus précis, nous avons approximé un noyau à un parallélogramme. Ensuite, j'ai éliminé les coordonnées z des points et utilisé cette formule :

L'aire d'un parallélogramme peut être calculée en utilisant la formule suivante :

$$A = \left\| \overrightarrow{AB} \wedge \overrightarrow{AD} \right\|$$

Formule vectorielle de l'aire d'un parallélogramme

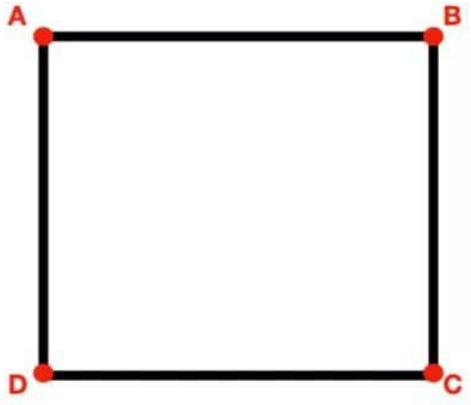

Figure 7 : Un noyau

Cette formule ignore un point (ici C), et les quatre points sont considérés comme étant dans le même plan 2D, car les coordonnées z de chaque point sont ignorées.

En se concentrant sur un seul noyau, on peut considérer ce noyau comme la base d'un prisme rectangulaire, l'autre base étant située aux coordonnées z=0 et étant identique. Une manière de l'imaginer est de penser que la caméra est située sur un plan perpendiculaire à l'axe des z. Par conséquent, le prisme rectangulaire aura une base sur ce plan et l'autre sera le noyau sur votre image. La hauteur de ce prisme rectangulaire sera la moyenne des coordonnées z de tous les points à l'intérieur du noyau, comme les points A, B, C, D, E, F et G dans l'image ci-dessous :

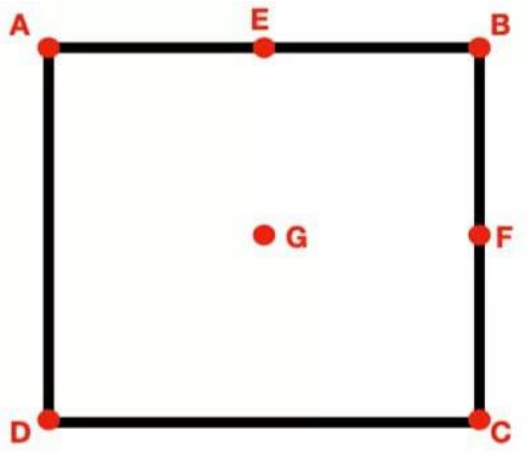

Figure 8 : E, F et G sont également utilisés pour calculer la hauteur moyenne

Ainsi, en calculant l'aire d'un noyau et en obtenant la hauteur du prisme rectangulaire correspondant, on peut effectuer le calcul du volume pour un noyau. En couvrant une image avec des noyaux et en additionnant leurs volumes correspondants, il est possible d'obtenir le volume total de l'image. Voici une représentation visuelle du volume : un pixel de chaque image représente le volume calculé avec un noyau de 2x2 pixels :

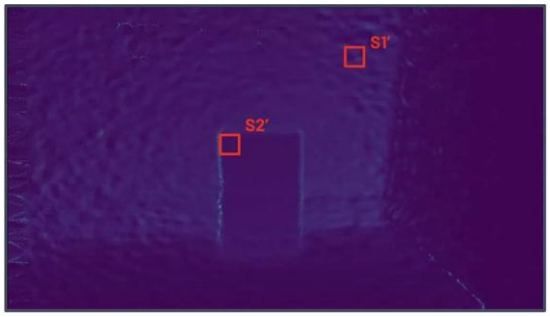

Figure 9a : Calcul du volume- Image brute avec boîte

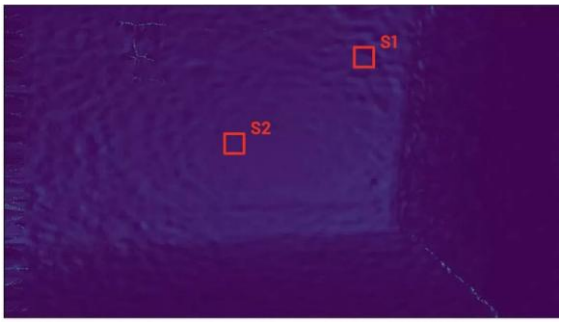

Figure 9b : Calcul du volume - Image brute sans boîte

Dans les images ci-dessus, j'ai ajouté deux carrés : S1 et S2 sur l'image sans boîtes, et S1' et S2' sur l'image avec boîtes. En comparant S1 et S1', les volumes devraient être approximativement les mêmes, car ils couvrent la même surface et ont la même profondeur. Ainsi, la différence entre S1 et S1' devrait être proche de 0. Cependant, en comparant S2 et S2', le volume de S2 devrait être plus grand que celui de S2' car S2' se trouve sur une boîte. Ainsi, même si S2 et S2' ont la même surface, la profondeur moyenne des pixels dans S2' est inférieure à celle de S2. La différence entre S2 et S2' correspond au volume occupé par la boîte. Par conséquent, en recouvrant les deux images de carrés et en calculant la différence de volume pour chaque carré correspondant, on peut obtenir le volume des boîtes.

3 - Conclusion

Pour conclure, le calcul du volume en utilisant cette technique donne de très bons résultats pour estimer le volume des objets. Pour deux boîtes, l'erreur absolue entre le volume calculé et le volume réel est inférieure à 1 %. Pour une et trois boîtes, cette erreur est inférieure à 3 %. Malgré les nombreuses hypothèses sous-jacentes concernant la pièce et les objets, cet algorithme se révèle efficace pour le calcul du volume des objets. De plus, il fournit une base pour des applications intéressantes en vision par ordinateur, telles que le calcul du volume de différents types d'objets en utilisant des algorithmes de détection d'objets.

4 - Références :

[1]: https://data-ai.theodo.com/en/technical-blog/mastering-volume-computation-of-objects-from-videos