DIASI/SIALV/25-036/Rév. 0 21 août 2025

Rapport de stage

Amélioration de la reconstruction 3D par Gaussian Splatting et stéréovision

Dorian Geraldes Pereira

Commissariat à l'énergie atomique et aux énergies alternatives
Institut List | CEA Paris-Saclay (Nano-INNOV) | Bât. 861-PC173
91191 Gif-sur-Yvette Cedex - FRANCE
T. +33 1 69 08 00 99
list.cea.fr

Direction: Institut List
Département: DIASI
Service: SIALV

Titre :	Amélioration de la reconstruction 3D par Gaussian Splatting et stéréovision	DIASI/SIALV/25-036
Auteur :	Dorian Geraldes Pereira	
Unité :	DIASI/SIALV/LVML	Page 2
Projet :	Rapport de stage	

Nombre de pages : 34

Résumé :	Amélioration de la reconstruction 3D grâce à des techniques d'intelligence artificielle, en particulier par l'intégration de méthodes de Gaussian Splatting et de stéréovision dans un pipeline robuste.

1		Validation électronique	Validation électronique	Validation électronique	Validation électronique
0		Dorian Geraldes Pereira	Régis Vinciguerra	Romain Dupont	Quoc-Cuong PHAM
Rév.	Date	Rédacteur	Vérificateur	Émetteur	Approbateur

Commissariat à l'énergie atomique et aux énergies alternatives
Institut List | CEA Paris-Saclay (Nano-INNOV) | Bât. 861-PC173
91191 Gif-sur-Yvette Cedex - FRANCE
T. +33 1 69 08 00 99
list.cea.fr

Direction: Institut List
Département: DIASI
Service: SIALV

Liste de diffusion

Destinataires	Société	Adresse(s) Email
Quoc-Cuong Pham	DRT/LIST/DIASI/SIALV	quoc-cuong.pham@cea.fr
Régis Vinciguerra	DRT/LIST/DIASI/SIALV/LVML	Regis.VINCIGUERRA@cea.fr
Romain Dupont	DRT/LIST/DIASI/SIALV/LVML	romain.dupont@cea.fr

Table des matières

Remerciements

Je tiens à exprimer ma profonde gratitude à l'ensemble des personnes qui ont contribué au bon déroulement de ce stage au sein du CEA List, et plus particulièrement au sein du laboratoire LVML.

Je remercie tout d'abord Romain Dupont, mon tuteur, pour son accompagnement tout au long du stage. Sa disponibilité, ses conseils techniques avisés et sa vision claire des enjeux m'ont été d'une aide précieuse, tant pour orienter mes travaux que pour approfondir mes compétences en vision par ordinateur. Je remercie également Régis Vinciguerra, chef du laboratoire, pour m'avoir accueilli au sein du LVML et permis d'évoluer dans un environnement stimulant, propice à la recherche appliquée.

Mes remerciements vont aussi à Pierre Elis et Olivier Gomez, pour leur aide leurs conseils. Enfin, je remercie l'ensemble de l'équipe du LVML pour leur accueil chaleureux, leurs échanges enrichissants et l'atmosphère conviviale qui a largement contribué à la qualité de cette expérience.

1 Présentation de l'entreprise, du laboratoire d'accueil

1.1 Présentation de l'entreprise

Le **Commissariat à l'Énergie Atomique et aux Énergies Alternatives (CEA)** est un organisme public de recherche scientifique fondé en 1945 par le général de Gaulle. Sa mission initiale était de fournir à la France la technologie atomique à des fins militaires et civiles. Aujourd'hui, le CEA est un acteur majeur de la recherche, du développement et de l'innovation dans de nombreux domaines scientifiques et technologiques : l'énergie (y compris les énergies renouvelables), la défense et la sécurité, les technologies de l'information, la santé, les sciences du vivant, le climat, l'environnement, ainsi que les sciences fondamentales (matière et univers).

Le CEA emploie près de 22 000 personnes réparties sur neuf sites en France : Paris-Saclay, Grenoble, Marcoule, Cadarache, DAM Île-de-France, Valduc, Cesta, Gramat et Le Ripault.

1.1.1 Organisation structurelle

Le CEA possède une organisation matricielle structurée autour de départements opérationnels et fonctionnels. En ce qui concerne les directions opérationnelles, on distingue quatre grandes entités :

- **DAM (Direction des Applications Militaires)** : en charge des activités liées à l'armement nucléaire, y compris les sous-marins nucléaires.
- **DES (Direction des Énergies)** : se consacre à l'innovation pour un système énergétique décarboné.
- **DRF (Direction de la Recherche Fondamentale)** : mène des recherches fondamentales en biologie, physique, chimie, etc., au service de l'ensemble des programmes scientifiques du CEA.

- **DRT (Direction de la Recherche Technologique)**, aussi appelée **CEA Tech** : focalisée sur le développement de technologies innovantes et leur transfert vers l'industrie. C'est dans cette direction que s'est déroulé mon stage.

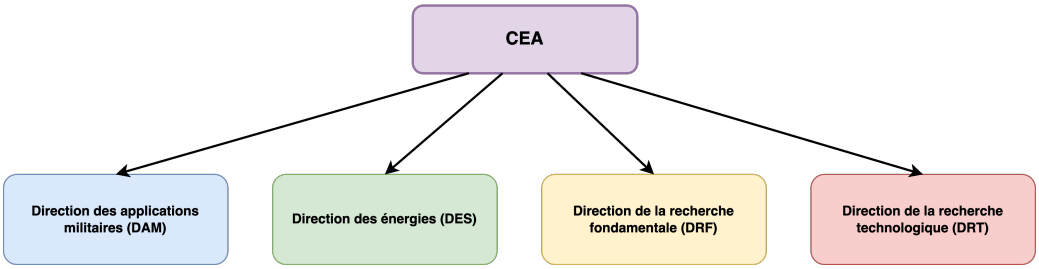


FIGURE 1 – Organisation structurelle du CEA

La division numérique de la DRT comprend deux instituts :

- **CEA-Leti**, basé à Grenoble, spécialisé dans les micro/nano-technologies (photonique, nanoélectronique, biotechnologie, etc.).
- **CEA-List**, situé principalement à Paris-Saclay, spécialisé dans les systèmes à forte intensité logicielle (systèmes embarqués, robotique, big data, capteurs, vision par ordinateur...).

1.2 Présentation du laboratoire d'accueil (LVML)

Dans le cadre de mon stage de fin d'études, j'ai intégré le laboratoire **LVML** (Laboratoire Vision pour la Modélisation et la Localisation), rattaché au **SIALV** (Service Intelligence Artificielle pour le Langage et la Vision), lui-même placé au sein du **DIASI** (Département Intelligence Ambiante et Systèmes Interactifs).

Le laboratoire LVML mène des travaux de recherche en vision par ordinateur et en apprentissage machine, initiés dès les années 1990. Il se concentre sur des applications telles que la localisation et le suivi d'objets, la reconstruction 3D, la réalité augmentée/diminuée et le rendu réaliste.

Le Laboratoire Vision pour la Modélisation et la Localisation, où s'est déroulé le stage, est en charge d'une activité de recherche initiée dès les années 90 sur les thèmes de la vision par ordinateur et du machine learning pour des applications traitant de :

- localisation dans l'environnement,
- localisation et suivi précis d'objets,
- reconstruction 3D,
- et plus généralement, vision pour la robotique.

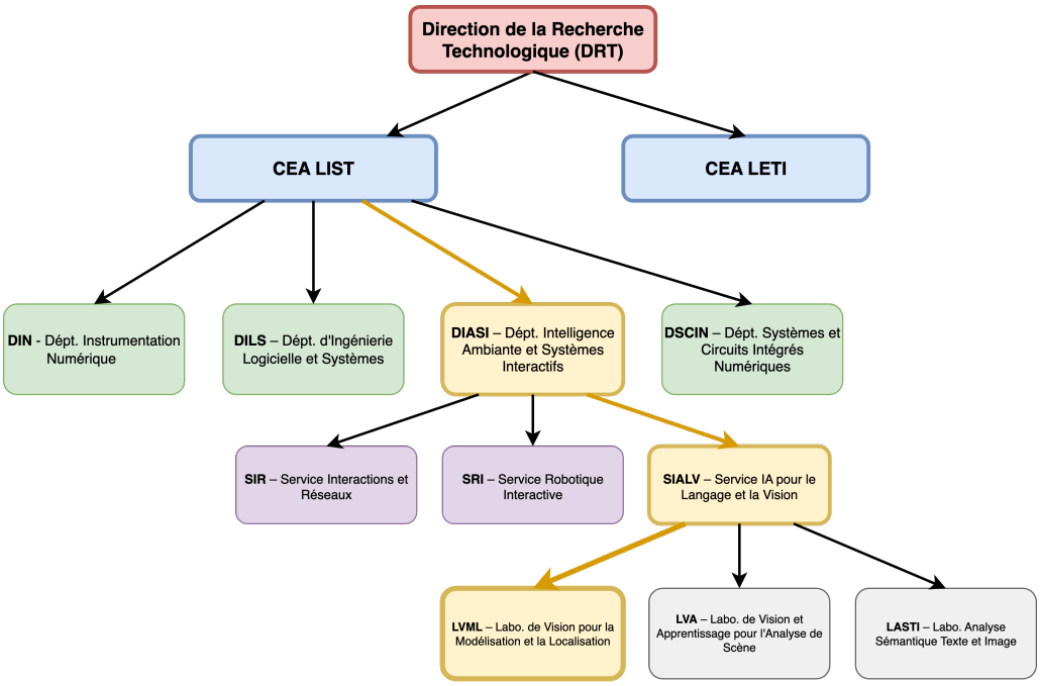


FIGURE 2 – Architecture organisationnelle de la DRT - Instituts, Départements, Services et Laboratoires

En tant que composant de l'Institut CEA LIST, le laboratoire a pour mission le transfert de son capital technologique vers le monde industriel. Ce transfert se construit en amont, souvent en collaboration avec des acteurs académiques, par le développement de méthodes d'analyse à la pointe de l'état de l'art ; et en aval, en partenariat avec des industriels, par l'adaptation de ces technologies à des contextes applicatifs spécifiques et innovants, en travaillant depuis la preuve de concept jusqu'aux prototypes préindustriels.

Ce positionnement a permis de développer une solide expérience dans la maturation des technologies pour passer du laboratoire (TRL 2-3) au démonstrateur en condition d'exploitation (TRL 7).

Le laboratoire est ainsi partenaire de grands groupes industriels tels que **KNDS**, **SIEMENS**, **IN GROUP** ou encore **TNS**, et adresse des projets dans un grand nombre de secteurs d'activité : manufacturing avancé, défense et sécurité, transports, énergie ...

Cette activité de transfert s'illustre également par la création de startups : **ActiCM** (2000, rachetée ensuite par *CREAFORM* – contrôle industriel), **Diota** (2009 – réalité augmentée), et **Tridimeo** (2017 – capteur 3D & multispectral). Ces partenariats sont sécurisés par un portefeuille brevet conséquent (10 brevets internationaux actifs).

Composé d'une vingtaine d'ingénieurs et chercheurs, dont 13 permanents et 7 doctorants, le LVML produit annuellement :

- une quinzaine de publications dans des conférences et journaux internationaux,
- 2 à 3 dépôts de brevets,

- et des recherches sur diverses technologies :
  - Géolocalisation et navigation 3D,
  - Modélisation 3D et sémantique de scènes
  - Détection de matériaux,
  - Compression de réseaux de neurones,
  - Vision pour la robotique de manipulation.

1.3 Politique de Responsabilité Sociétale des Entreprises

Le **CEA LIST** intègre la Responsabilité Sociétale des Entreprises (RSE) dans ses activités de recherche et d'innovation technologique. La direction technologique du CEA se veut être le pont entre la recherche et l'industrie. En tant qu'institut de recherche appliquée, il se consacre à des projets ayant un impact direct sur la société, tels que le développement de technologies pour la santé, la transition énergétique, et la sécurité numérique.

Le CEA LIST promeut des pratiques durables en minimisant son empreinte écologique, notamment à travers :

- l'optimisation de l'efficacité énergétique de ses infrastructures,
- la gestion responsable des déchets,
- la promotion de modes de transports bas carbone, tels que le train pour les voyages professionnels.

Par ailleurs, il valorise le bien-être de ses employés en mettant en place des initiatives pour favoriser la diversité, l'inclusion et le développement professionnel.

Grâce à sa démarche RSE, le CEA LIST contribue activement à l'innovation responsable, alignant ses objectifs scientifiques et technologiques avec les grands enjeux environnementaux et sociaux contemporains.

2 Introduction

La reconstruction 3D à partir d'images constitue un enjeu majeur dans de nombreux domaines technologiques, tels que la réalité augmentée, la robotique ou encore la cartographie. Elle vise à générer une représentation numérique fidèle d'un environnement ou d'un objet en trois dimensions, à partir d'un ensemble d'images capturées sous différents angles.

Si la qualité du *rendu visuel* est souvent mise en avant, il est tout aussi crucial d'obtenir une représentation géométrique explicite, notamment sous forme de **maillage 3D**. En effet, un tel maillage permet :

- de **réaliser des mesures physiques** (distances, volumes, angles) à partir d'une surface explicite,
- de **caractériser la topologie** d'objets ou de scènes (composantes connexes, trous, cavités, etc.),

- de **comparer géométriquement** différentes scènes ou objets (détection de changements, recalage, suivi d'évolution),
- de **garantir une cohérence structurelle** lors d'une visualisation à distance ou sous des angles non optimaux,
- de servir de support à des simulations physiques ou à des traitements plus avancés (Planification robotique, FEM, etc.).

Ce stage s'inscrit dans un cadre plus large de recherche sur les techniques d'intelligence artificielle appliquées à la reconstruction 3D, avec un objectif d'amélioration des pipelines existants. Après une phase d'exploration des différentes approches, le sujet a été recentré sur l'optimisation de la reconstruction à partir de représentations implicites, en tirant parti des avancées en stéréovision dense et en Gaussian Splatting.

Contrairement aux approches monoculaires, où l'estimation de la profondeur repose essentiellement sur des indices d'apparence et des modèles statistiques entraînés, la stéréovision permet d'exploiter directement les **contraintes géométriques fortes entre vues**. Cette redondance multi-vues rend l'estimation de profondeur plus fiable et mieux contrainte, notamment dans des scènes complexes ou faiblement texturées, où les méthodes monoculaires montrent rapidement leurs limites.

Mon travail de stage s'inscrit dans ce contexte, avec pour objectif d'extraire une *géométrie maillée exploitable* à partir de représentations continues issues de méthodes récentes de Gaussian Splatting. La chaîne de traitement repose initialement sur des outils classiques de Structure-from-Motion (SfM) pour estimer les poses de caméras et générer un nuage de points 3D initial, avant d'optimiser une représentation plus dense et précise de la scène sous forme de splats gaussiens.

Cependant, le cadre de ce stage se distingue par le fait que les données utilisées proviennent de captations en conditions réelles, ce qui introduit plusieurs défis pratiques, rarement abordés dans les pipelines idéalisés de la littérature. On rencontre notamment :

- des effets de spécularité et de réflexions, qui nuisent à la correspondance entre images,
- des problèmes de calibration, les paramètres des caméras ne sont pas toujours connus
- la nécessité de rectifier les paires stéréo pour assurer une correspondance géométrique fiable,
- et une variabilité importante dans la qualité des données (flou, éclairage, bruit capteur).

Bien que 3D Gaussian Splatting (3DGS) permette un rendu photoréaliste en temps réel, sa représentation volumique pose des limites pour l'extraction de surfaces : les splats sont évalués selon la vue courante, ce qui peut induire des incohérences multi-vues et compromettre la qualité de la géométrie reconstruite.

Pour contourner cette limitation, ce stage explore l'approche alternative du 2D Gaussian Splatting (2DGS), qui projette directement les splats dans l'espace image. Cette formulation permet une *meilleure cohérence entre vues*, facilite la génération de *cartes de profondeur fiables*, et simplifie l'intégration avec des méthodes d'extraction de surface comme la fusion TSDF et l'algorithme Marching Cubes.Toutefois, cette approche peut entraîner une légère dégradation de

la qualité visuelle du rendu final, du fait de la perte de certaines optimisations spécifiques à la 3D.

L'un des objectifs spécifiques de ce travail est d'exploiter les contraintes géométriques induites par la stéréo, pour guider l'estimation de profondeur et améliorer la qualité des reconstructions. Cette exploitation directe de la disparité entre vues offre un avantage décisif par rapport aux approches monoculaires, en apportant une robustesse accrue face aux artefacts des données réelles et une meilleure cohérence géométrique globale.

Ainsi, ce stage vise à construire un pipeline cohérent allant des images d'entrée jusqu'au maillage final, en évaluant l'apport de chaque étape, et en proposant des améliorations spécifiques à l'extraction de géométrie à partir des représentations issues de 2DGS, dans un cadre réaliste, bruité, et structurellement contraint par la stéréo.

3 État de l'art

La reconstruction 3D à partir d'images est un domaine central en vision par ordinateur, avec des applications en réalité virtuelle, en cartographie, ou encore en robotique. Plusieurs approches ont été proposées au fil des années pour répondre à ce besoin, allant de méthodes géométriques classiques à des approches plus récentes basées sur l'apprentissage automatique.

3.1 Méthodes classiques de reconstruction 3D

Les méthodes traditionnelles de reconstruction 3D reposent principalement sur la *Structure-from-Motion* (SfM) et la *Multi-View Stereo* (MVS). Ces techniques permettent de reconstruire une scène tridimensionnelle à partir d'un ensemble d'images 2D, en estimant d'abord les poses des caméras, puis en triangulant les points correspondants pour obtenir un nuage de points dense.

La SfM consiste à estimer simultanément la position des caméras et la géométrie de la scène à partir de correspondances entre images. Une fois les poses des caméras obtenues, la MVS est utilisée pour densifier le nuage de points initial en exploitant les informations de profondeur issues de multiples vues.

Dans un pipeline typique, on distingue plusieurs étapes successives :

- **SfM** : calcul des poses et d'un nuage de points clairsemé à partir des correspondances d'images ;
- **Estimation de cartes de profondeur** : pour chaque image ou paire d'images, une carte de disparité ou de profondeur est générée, tenant compte des poses connues ;
- **Fusion des cartes de profondeur** : pour produire un **nuage de points dense** représentatif de la scène ;
- **Reconstruction de surface** : génération d'un **maillage 3D** à partir du nuage dense via des techniques comme Poisson Surface Reconstruction ou Delaunay triangulation ;
- **Texturation** : application de textures photoréalistes issues des images d'origine pour recouvrir le maillage.

Plusieurs outils implémentent efficacement ce pipeline, chacun avec ses spécificités. **OpenMVS** est une solution open-source couramment utilisée en recherche, qui s'intègre facilement avec des frameworks SfM comme COLMAP. Des alternatives propriétaires comme **RealityCapture** ou **Metashape** (anciennement Agisoft) offrent des pipelines plus optimisés et intégrés, utilisés notamment dans les domaines de la photogrammétrie commerciale, de l'architecture ou de la conservation du patrimoine. Ces outils proposent des interfaces utilisateur avancées et une automatisation complète, allant de l'import des images jusqu'à l'export de modèles texturés haute qualité.

Ces approches géométriques ont longtemps constitué l'état de l'art pour la reconstruction 3D, et servent aujourd'hui encore de fondation à de nombreuses méthodes récentes, notamment en fournissant une estimation initiale fiable de la scène.

3.2 Neural Radiance Fields (NeRF)

Avec l'essor du deep learning, de nouvelles représentations dites *implicites* ont vu le jour. Les plus connues sont les Neural Radiance Fields (NeRF) [7], qui modélisent une scène comme une fonction continue, apprise par un réseau de neurones, reliant chaque point 3D à une densité et une couleur. Bien que cette approche permette des rendus photoréalistes, elle présente plusieurs limitations : un temps d'apprentissage long, une difficulté d'interprétation géométrique directe, et une complexité d'extraction de surface.

3.3 3D Gaussian Splatting (3DGS)

Pour surmonter certaines limitations de NeRF, notamment son coût d'inférence élevé, *3D Gaussian Splatting* [4] propose une approche de rendu photoréaliste en temps réel sans avoir recours à un réseau de neurones à l'inférence. La scène est représentée comme un ensemble de gaussiennes tridimensionnelles, chacune définie par sa position, sa forme (covariance), sa couleur et son opacité.

Le rendu est effectué par rasterisation sur GPU : chaque gaussienne est projetée de manière différentiable dans l'espace image, et son influence est accumulée pour produire une image finale. Cette méthode permet des rendus de haute qualité avec un pipeline bien plus rapide que NeRF, notamment pour des scènes déjà reconstruites.

L'initialisation de la scène repose généralement sur un pipeline SfM tel que COLMAP, qui fournit les poses de caméras ainsi qu'un nuage de points approximatif utilisé pour estimer les gaussiennes initiales.

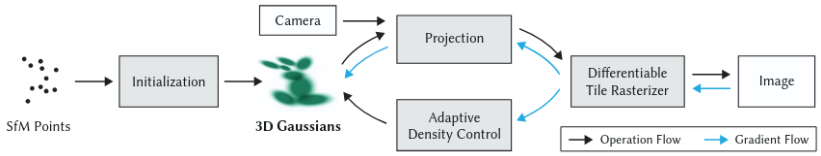


FIGURE 3 – Le processus d'optimisation débute à partir d'un nuage de points sparse issu d'une reconstruction SfM, utilisé pour initialiser un ensemble de gaussiennes 3D. Cet ensemble est ensuite affiné de manière itérative, avec un contrôle adaptatif de la densité. L'optimisation s'appuie sur un moteur de rendu rapide basé sur un découpage en tuiles (tile-based), ce qui permet des temps d'apprentissage compétitifs par rapport aux méthodes rapides de champs de radiance de l'état de l'art. Une fois entraîné, le rendu permet une navigation en temps réel dans une grande variété de scènes.

Bien que 3DGS soit avant tout conçu pour le rendu, sa nature explicite, contrairement à celle de NeRF, offre un point de départ intéressant pour d'éventuelles tentatives d'exploitation géométrique, comme celles explorées dans le cadre de ce stage.

3.4 2D Gaussian Splatting (2DGS)

Bien que 3DGS permette un rendu photoréaliste en temps réel, sa représentation volumique tridimensionnelle présente certaines limites, notamment des incohérences entre les vues dues à la projection sur des plans d'intersection différents selon la position de la caméra. Ces variations peuvent compromettre la cohérence multi-vues, ce qui complique les tâches de reconstruction géométrique.

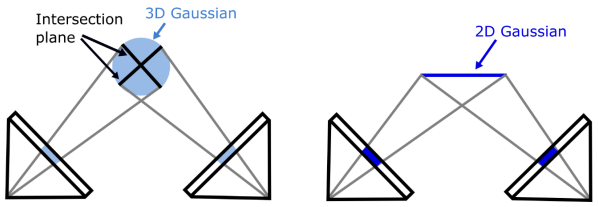


FIGURE 4 – Comparaison entre 3DGS et 2DGS [3]. Dans 3DGS, l'évaluation des splats dépend du point de vue, induisant des incohérences multi-vues. 2DGS assure une évaluation cohérente en projetant directement les splats dans l'espace image.

Pour pallier cette limitation, *2D Gaussian Splatting* [3] propose une représentation alternative basée uniquement sur des splats gaussiens en 2D, définis directement dans l'espace image. Cette approche permet une meilleure cohérence inter-vues et simplifie considérablement le traitement géométrique.

Concrètement, des cartes de profondeur synthétiques (depth maps) sont générées à partir des splats projetés dans chaque vue, en exploitant la rasterisation différentiable. Ces cartes sont ensuite fusionnées à l'aide d'une méthode volumique classique basée sur une TSDF (Truncated Signed Distance Function), comme celle fournie par la bibliothèque Open3D [13]. L'application de l'algorithme Marching Cubes permet alors d'extraire une surface maillée explicite.

Ainsi, 2DGS complète 3DGS en proposant un pipeline plus adapté à la reconstruction géométrique, tout en conservant les avantages du splatting, mais dans un espace image mieux maîtrisé.

3.5 Supervision géométrique : Gaussian splatting supervisé par la profondeur

Afin d'améliorer la stabilité géométrique des splats, plusieurs travaux ont proposé d'introduire une supervision explicite à partir de données de profondeur ou de régularisation géométrique multi-vues. Ces approches visent à éviter les dérives locales de la géométrie, en particulier dans les régions peu texturées ou mal reconstruites.

Self-Evolving Depth-Supervised 3D Gaussian Splatting from Rendered Stereo Pairs [8] propose d'améliorer la précision géométrique des splats en introduisant une supervision de la profondeur auto-supervisée, sans capteur externe. Leur méthode repose sur l'idée d'exploiter dynamiquement les indices de profondeur fournis par un réseau stéréo existant, appliqué sur des paires d'images virtuelles générées pendant l'entraînement. Concrètement, à intervalles réguliers, une vue stéréo est synthétisée à partir du modèle GS lui-même, puis comparée à la vue principale via un estimateur de disparité comme RAFT [11] pour produire une carte de profondeur pseudo-vraie.

Cette profondeur estimée est ensuite comparée à la profondeur rendue par les splats via une fonction de coût de type L1. L'optimisation conjointe permet une amélioration progressive de la représentation géométrique, en particulier dans les zones où la supervision photométrique seule échoue (basse texture, ambigüité de correspondance)

Multiview Geometric Regularization of Gaussian Splatting for Accurate Radiance Fields [5] propose une régularisation géométrique multi-vues en incorporant directement les estimations de profondeur issues d'un pipeline de stéréo multi-vues (MVS). Cette approche s'appuie sur le constat que MVS est particulièrement robuste dans les zones riches en texture, tandis que Gaussian Splatting tend à fournir des profondeurs plus fiables dans les régions à faible variation de couleur.

La méthode introduit une fonction de coût de profondeur relative multi-vues, formulée autour d'une médiane robuste des profondeurs observées depuis plusieurs vues. Une fenêtre de tolérance dégressive est appliquée, afin de ne pénaliser que les écarts significatifs. En complément, une initialisation des gaussiennes guidée par les reconstructions MVS permet d'éviter les placements aberrants en début d'optimisation.

Ces deux approches illustrent une tendance actuelle visant à renforcer la composante géométrique du Gaussian Splatting en intégrant des signaux explicites de profondeur. Elles contribuent à réduire les ambiguïtés multi-vues et à rendre les reconstructions plus précises, tout en conservant la qualité de rendu caractéristique du splatting.

3.6 MILo : Mesh-In-the-Loop Gaussian Splatting for Detailed and Efficient Surface Reconstruction

La plupart des méthodes de Gaussian Splatting produisent une représentation optimisée pour le rendu, mais difficile à convertir directement en surface exploitable. L'extraction de maillage est généralement réalisée en post-traitement à l'aide de techniques volumétriques comme la fusion TSDF suivie de l'algorithme Marching Cubes. Cette étape peut entraîner une fonction de coût

de détails géométriques fins et produire des maillages très denses, peu adaptés à des applications interactives.

MILo (Mesh-In-the-Loop) [2] propose un cadre d'apprentissage différentiable dans lequel un maillage est extrait à chaque itération directement à partir des gaussiennes optimisées. Cette méthode établit une connexion explicite entre la représentation volumique des splats et une surface triangulée, permettant aux deux de co-évoluer pendant l'entraînement.

Le pipeline complet s'effectue en cinq étapes clés :

1. Échantillonnage des sommets de Delaunay à partir des centres des gaussiennes (appelés pivots), définis comme points d'ancrage spatiaux.
2. Construction dynamique d'une triangulation de Delaunay autour de ces sommets.
3. Interpolation de distances signées aux sommets du maillage, apprises de manière différentiable à partir des gaussiennes voisines.
4. Extraction d'une isosurface via une variante différentiable de l'algorithme *Marching Tetrahedra*, entièrement implémentée sur GPU.
5. Rendu simultané du maillage extrait et des gaussiennes pour appliquer des fonctions de coût d'image (photométriques) et de profondeur/normale, puis rétro-propagation vers les paramètres des gaussiennes.

Grâce à cette intégration directe du maillage dans la boucle d'optimisation, MILo permet de produire des reconstructions précises avec un nombre de sommets réduit, tout en conservant la qualité de rendu des méthodes basées sur le splatting

4 Méthodologie

Le pipeline mis en œuvre durant ce stage vise à améliorer la qualité des reconstructions produites par la méthode 2D Gaussian Splatting (2DGS), en exploitant des cartes de profondeur denses générées par stéréovision. L'objectif est de partir d'un ensemble d'images, potentiellement non calibrées, pour aboutir à une reconstruction 3D cohérente et exploitable.

4.1 (Facultatif) Estimation des paramètres de la caméra avec COLMAP

Lorsque les paramètres intrinsèques et extrinsèques des caméras ne sont pas disponibles, une étape de calibration structure-from-motion est réalisée à l'aide de COLMAP. Ce dernier permet d'estimer les poses relatives entre les caméras, ainsi que les paramètres intrinsèques (focale, distorsion, etc.), à partir d'un ensemble d'images.

4.2 Rectification des paires d'images avec OpenCV

Les images sont ensuite rectifiées pour préparer l'estimation des carte de disparité . En s'appuyant sur les paramètres de caméra estimés ou fournis, la bibliothèque OpenCV est utilisée pour effectuer une rectification stéréo. Cette étape aligne les épipolaires, ce qui facilite la correspondance pixel-à-pixel entre deux images.

4.3 Estimation des poses sur images rectifiées

Les images rectifiées ne sont plus strictement compatibles avec les poses initiales (lorsqu'elles existent), en raison des transformations géométriques appliquées lors de la rectification. Il est donc nécessaire d'estimer les poses dans l'espace des images rectifiées.

Ainsi, une estimation des poses caméras est systématiquement réalisée après la rectification, en utilisant COLMAP sur les images rectifiées. Si les paramètres caméras étaient absents au départ, il s'agit d'une seconde estimation ; dans le cas contraire, c'est la seule étape de calibration effectuée, directement sur les données rectifiées.

4.4 (Facultatif) Remise à l'échelle du référentiel COLMAP

COLMAP reconstruit la scène dans un référentiel projectif arbitraire, où l'échelle absolue n'est pas définie. Lorsque des mesures réelles sont nécessaires notamment pour rendre les cartes de profondeur cohérentes avec les unités physiques (mètres) une remise à l'échelle est effectuée.

Cette opération consiste à calculer un facteur d'échelle global, en comparant la baseline estimée par COLMAP à une valeur de référence (issue, par exemple, de métadonnées, d'un capteur, ou de la configuration de la scène). Ce facteur est ensuite appliqué aux poses caméra et aux profondeurs, afin d'obtenir une reconstruction métrique fidèle à la réalité.

4.5 Estimation des cartes de disparité avec FoundationStereo

Des paires d'images rectifiées sont ensuite passées dans le réseau *FoundationStereo*, qui permet de produire des cartes de disparité à partir des correspondances stéréo.

Pour fiabiliser les résultats, un *cross-checking* gauche-droite / droite-gauche est effectué : seules les correspondances réciproques sont conservées, c'est-à-dire celles pour lesquelles la projection d'un point de l'image gauche vers la droite coïncide avec sa reprojection inverse. Cette vérification croisée permet de détecter et d'éliminer les erreurs de correspondance incohérentes, notamment dans les régions ambiguës ou peu texturées, où les réseaux apprennent parfois des *a priori* de surface incorrects (comme des discontinuités lissées ou des plans erronés).

Le résultat est une carte de disparité filtrée, plus robuste aux artefacts, servant de base pour la génération des cartes de profondeur.

4.6 Conversion des disparités en cartes de profondeur

Les cartes de disparité sont converties en cartes de profondeur métriques à l'aide de la baseline entre les caméras. Cette baseline est soit connue a priori, soit estimée à partir des poses COLMAP. On obtient ainsi, pour chaque pixel, une profondeur exprimée dans le repère de la caméra.

4.7 Re-projection en nuage de points dense

En utilisant les cartes de profondeur et les poses caméra issues de COLMAP, chaque pixel est re-projeté dans l'espace 3D pour générer un nuage de points dense. Ce nuage est une représentation géométrique plus complète que celle fournie par COLMAP seul.

4.8 Reconstruction de la scène avec 2D Gaussian Splatting

Enfin, les données produites (cartes de profondeur, nuage de points dense, poses caméra) sont utilisées pour initialiser la méthode 2D Gaussian Splatting [3]. Celle-ci a été modifiée dans le cadre de ce travail afin d'accepter un nuage de points dense comme point de départ, et d'intégrer les cartes de profondeur dans le processus d'entraînement.

4.9 Extraction du maillage 3D à partir de 2DGS

Une fois la reconstruction réalisée via 2D Gaussian Splatting, des cartes de profondeur peuvent être rendues depuis différentes vues en exploitant les poses caméras. Ces cartes sont ensuite utilisées pour fusionner la géométrie dans un volume TSDF (Truncated Signed Distance Function) à l'aide de la bibliothèque Open3D.

Cette étape permet d'intégrer les surfaces visibles dans une représentation volumique, et d'en extraire un maillage 3D explicite à l'aide de l'algorithme de Marching Cubes. Elle permet ainsi d'obtenir un modèle surfacique structuré à partir des données implicites fournies par 2DGS.

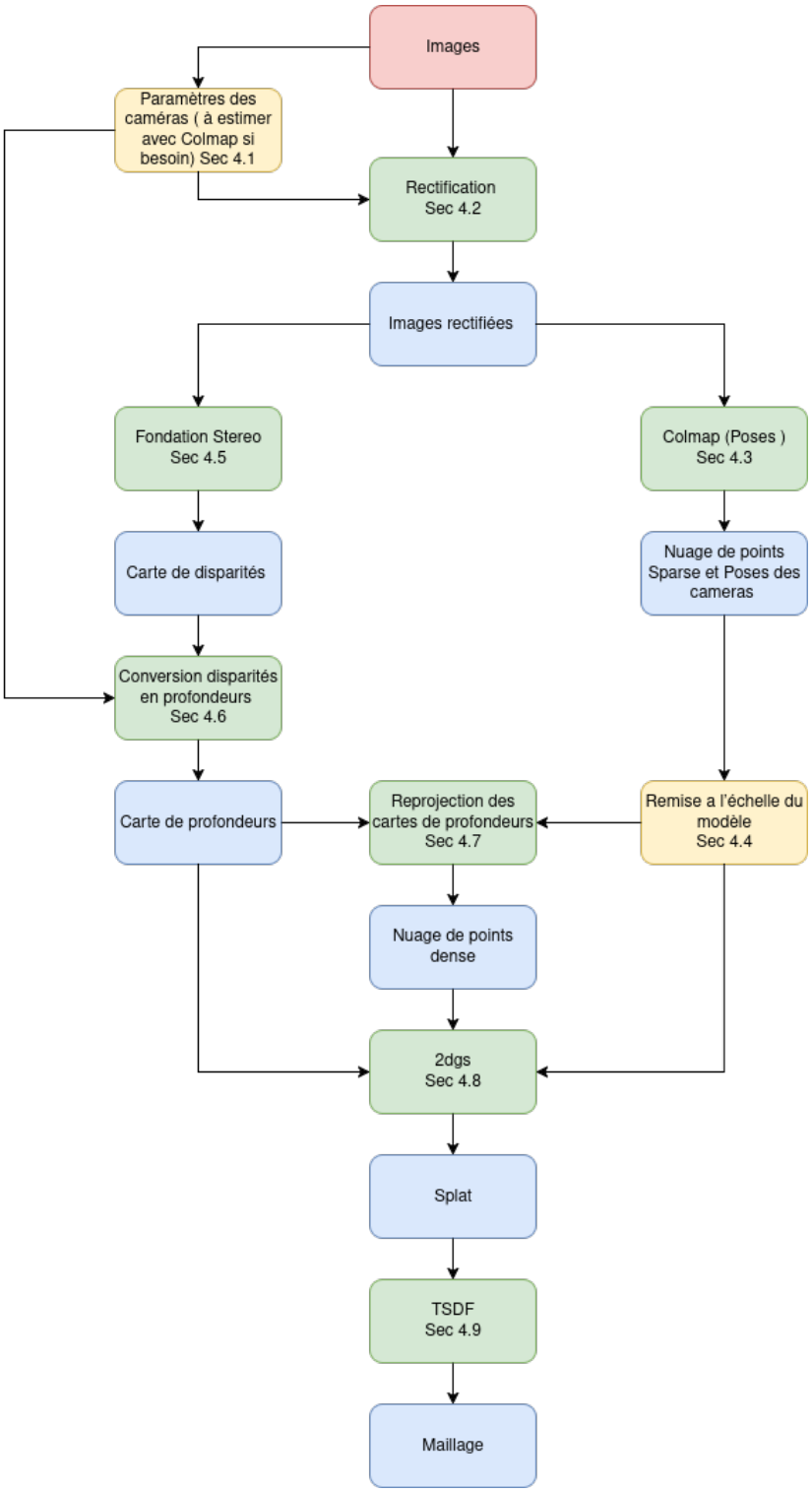


FIGURE 5 – Schéma de la méthodologie proposée. En vert : algorithmes obligatoires, en jaune : modules optionnels, en bleu : résultats intermédiaires ou finaux.

5 Outils

La mise en place du pipeline de reconstruction 3D s'appuie sur plusieurs outils logiciels interconnectés. Cette section présente brièvement les principaux composants utilisés et leur rôle dans le processus.

5.1 COLMAP : un outil pivot de la reconstruction

COLMAP [9, 10] est un outil open-source largement utilisé dans les pipelines de reconstruction 3D. Il implémente une version robuste de la Structure-from-Motion (SfM) combinée à des algorithmes de Multi-View Stereo (MVS), avec des descripteurs locaux (comme SIFT) et une estimation robuste des poses par RANSAC.

COLMAP occupe aujourd'hui une place centrale dans les pipelines hybrides combinant géométrie et apprentissage automatique. Il est notamment utilisé dans :

- L'estimation des poses de caméra à partir d'un ensemble d'images,
- La génération d'un nuage de points initial pour guider les étapes ultérieures,
- L'initialisation de scènes dans des approches modernes comme 3D Gaussian Splatting (3DGS) ou 2DGS.

Sans cette étape de reconstruction initiale, les méthodes comme 3DGS ou 2DGS ne peuvent ni projeter correctement les éléments dans l'espace image, ni effectuer une optimisation géométrique cohérente. COLMAP s'impose donc comme une brique logicielle incontournable dans le pipeline de reconstruction moderne.

5.2 FoundationStereo

FoundationStereo [12] est un estimateur de disparité basé sur des techniques récentes de stéréovision dense. Il prend en entrée des paires d'images rectifiées et génère des cartes de disparité de haute résolution, sans nécessiter d'entraînement préalable (approche *zero-shot*).

Dans le cadre de notre pipeline, ces cartes de disparité sont exploitées à deux niveaux :

- elles sont converties en nuages de points denses via une reprojection 3D fondée sur la calibration des caméras,
- elles servent à définir une fonction de coût géométrique permettant de contraindre l'optimisation des splats dans les méthodes de type Gaussian Splatting.

L'utilisation de FoundationStereo améliore ainsi la précision géométrique globale du pipeline, tout en restant compatible avec des scènes variées et non vues lors de l'entraînement.

5.3 OpenCV

OpenCV [1] est une bibliothèque open-source largement utilisée pour le traitement d'images et de vidéos. Dans notre pipeline, elle intervient dès les premières étapes du traitement stéréo.

En particulier, la fonction `stereoRectify` d'OpenCV est utilisée pour réaliser la rectification stéréo des paires d'images. Cette opération transforme les images afin d'aligner leurs épipolaires, rendant ainsi les correspondances entre pixels plus fiables. Une bonne rectification conditionne directement la qualité des cartes de disparité générées, et par conséquent la précision des cartes de profondeur.

OpenCV constitue donc un outil fondamental pour la préparation des données d'entrée dans notre pipeline de reconstruction.

5.4 Open3D

Open3D [13] est une bibliothèque open-source dédiée au traitement des données 3D, et constitue un composant central du pipeline de reconstruction géométrique.

En fin de pipeline, Open3D est utilisée pour la fusion volumique des cartes de profondeur générées à partir du modèle 2DGS, via son module basé sur la *Truncated Signed Distance Function* (TSDF). Cette intégration spatiale permet de combiner les observations issues de plusieurs vues de manière cohérente, réduisant le bruit et comblant les lacunes. Le résultat est une représentation surfacique dense et explicite de la scène, à partir de laquelle le maillage final est extrait à l'aide de l'algorithme de Marching Cubes.

Ainsi, Open3D joue un rôle crucial dans la consolidation des informations géométriques et la génération du modèle 3D final.

6 Implementations et Contributions

6.1 Modification de l'initialisation du Gaussian Splatting

La première amélioration apportée au pipeline 2D Gaussian Splatting a consisté à modifier l'étape d'initialisation des splats. Plutôt que d'utiliser un nuage de points sparse issu de la reconstruction SfM (Structure-from-Motion), un nuage de points dense a été généré à partir des cartes de profondeur obtenues par stéréovision.

Ce nuage dense est obtenu par reprojection 3D des cartes de profondeur produites par *FoundationStereo* (cf. section 5) sur des paires d'images rectifiées. Les poses caméra, connues ou estimées par COLMAP, permettent cette conversion fiable. Les splats sont ensuite initialisés à partir de ces points 3D, ce qui garantit un positionnement plus dense et précis, notamment sur les structures planes de la scène (sols, murs, bureaux).

Cette initialisation dense présente un double intérêt. D'une part, elle améliore significativement la couverture géométrique de la scène dès le départ, en évitant les trous dans les zones planes ou peu texturées. D'autre part, elle permet de guider l'optimisation dans un espace de solutions plus favorable, en réduisant le risque de convergence vers des minima locaux sous-optimaux. En effet, dans le cadre du 2D Gaussian Splatting, la fonction de coût est fortement non convexe, et une initialisation aléatoire ou sparse peut entraîner une dérive rapide des splats vers des configurations erronées.

Malgré cette amélioration, certaines limites subsistent lors de l'entraînement :

- Dans les zones peu texturées ou contenant des surfaces réfléchissantes, les splats dérivaient fortement au cours de l'optimisation.

- La variation d'éclairage entre les vues créait des incohérences de supervision photométrique, ce qui perturbait l'alignement des splats.

Ces dérives s'expliquent par le fait que la fonction de coût photométrique, utilisée seule pendant l'entraînement, n'est pas suffisante pour garantir une reconstruction géométriquement correcte. Elle repose sur l'hypothèse d'une cohérence d'apparence entre les vues, hypothèse souvent mise à mal par des effets d'éclairage, de brillance ou des occlusions. En l'absence de contrainte explicite sur la géométrie, les splats peuvent ainsi converger vers des configurations erronées qui minimisent le coût photométrique sans respecter la structure 3D réelle.

Ces constats ont conduit à une seconde amélioration, décrite dans la section 6.4 : l'intégration explicite des cartes de profondeur dans la fonction de coût pendant l'apprentissage. Cette supervision géométrique a permis de stabiliser l'optimisation dans les zones problématiques.

6.2 Implémentation de masques pour le traitement des zones problématiques et la suppression d'objets

L'intégration de masques dans le pipeline d'entraînement du 2D Gaussian Splatting a initialement été motivée par le besoin de traiter certaines zones problématiques des images : surfaces spéculaires, variations d'éclairage d'une vue à l'autre, ou encore régions sombres peu informatives. Ces phénomènes introduisent des incohérences photométriques qui perturbent l'optimisation, en particulier dans un cadre multi-vues.

L'idée était donc de désactiver la supervision photométrique dans ces régions, en les masquant explicitement pendant l'apprentissage. Les masques ont été générés soit manuellement, soit automatiquement à l'aide du modèle de segmentation Segment Anything Model (SAM) [6]. En pratique, les pixels masqués ne participent plus à la fonction de coût RGB, ce qui empêche ou limite la formation de splats dans les zones concernées.

Cependant, cette approche ne s'est pas révélée concluante pour corriger les défauts liés aux zones problématiques : dans plusieurs cas, leur absence a conduit à une perte de continuité visuelle et de consistance géométrique. On suppose que cela est en partie dû à la perte de points de vue alternatifs dans les zones masquées, ce qui fragilise la supervision globale.

Dans un second temps, cette capacité à masquer certaines régions a été détournée à des fins exploratoires : nous avons testé la possibilité d'utiliser les masques pour supprimer certains objets visibles dans les images (ex. personnes, mobilier, etc.), dans le but d'évaluer le potentiel du pipeline pour des tâches d'édition de scènes.

Les résultats obtenus sont encourageants : malgré les limitations actuelles — en particulier l'absence de prise en compte des masques dans la supervision géométrique par cartes de profondeur — plusieurs objets masqués ont pu être atténués, voire totalement supprimés dans les rendus finaux.

Cette approche simple mais efficace ouvre des perspectives intéressantes pour l'édition 3D légère, notamment dans des contextes de nettoyage de scène ou d'anonymisation visuelle

6.3 Modification de la rectification stéréo

Dans sa version standard, la fonction `stereoRectify` d'OpenCV ne garantit pas le recentrage du point principal lors de la rectification des images. Or, cette contrainte géométrique est essentielle pour assurer la compatibilité avec le pipeline 2D Gaussian Splatting, qui suppose que le point principal se situe au centre de chaque image rectifiée afin de garantir une projection correcte des splats gaussiens.

Dans le cadre de ce travail, le code source C++ d'OpenCV a été modifié afin d'imposer systématiquement cette recentralisation. Cette adaptation permet de produire des paires rectifiées compatibles avec les exigences géométriques de la méthode 2DGS, tout en conservant l'intégration dans le pipeline général de reconstruction.

6.4 Implémentation des cartes de profondeur dans l'entraînement

Pour stabiliser l'optimisation des splats, notamment dans les zones peu texturées ou soumises à de fortes variations d'éclairage, une supervision basée sur les cartes de profondeur a été intégrée au processus d'apprentissage du 2D Gaussian Splatting.

Dans les travaux de *Self-Evolving Depth-Supervised 3D Gaussian Splatting from Rendered Stereo Pairs* [8], les auteurs proposent une supervision explicite en utilisant une fonction de coût $\mathcal{L}_1$ classique entre la profondeur rendue et la profondeur issue de la stéréo. Bien que cette formulation améliore la stabilité dans certaines régions, elle ne prend pas en compte la non-uniformité de la sensibilité en fonction de la distance à la caméra.

Dans notre approche, nous avons introduit une version modifiée de cette fonction de coût, dans laquelle l'erreur est pondérée par l'inverse du carré de la profondeur stéréo. Cette pondération s'appuie sur une propriété fondamentale de la géométrie stéréo. En effet, l'erreur sur la profondeur estimée à partir d'une carte de disparité croît quadratiquement avec la distance à la caméra, selon la relation :

$$\Delta D \approx \frac{Bf}{d^2}\Delta d$$

où D est la profondeur, B la baseline entre les caméras, f la focale, d la disparité, et Δd l'erreur de correspondance. Ainsi, à disparité constante, une petite erreur de correspondance entraîne une erreur en profondeur d'autant plus grande que l'objet est éloigné. Pour compenser cette variation naturelle de précision, il est pertinent de renforcer la supervision sur les régions proches, là où la géométrie est mieux contrainte.

La fonction de coût pondérée adoptée est donc la suivante :

$$\mathcal{L}_{\text{depth}} = \sum_{i \in \mathcal{P}} \frac{1}{D_{\text{gt}}(i)^2} \left| D_{\text{r}}(i) - D_{\text{gt}}(i) \right|$$

où $D_{\text{r}}(i)$ désigne la profondeur estimée par rendu des splats, $D_{\text{gt}}(i)$ la profondeur mesurée par stéréo, et $\mathcal{P}$ l'ensemble des pixels visibles utilisés pour la supervision.

Cette pondération permet donc de corriger le biais inhérent à la précision variable des cartes de profondeur, en insistant plus fortement sur les objets proches, où une erreur serait plus visible et géométriquement significative. Elle contribue à limiter la dérive des splats dans ces zones sensibles et à améliorer la fidélité structurelle de la reconstruction.

Cette méthode a ainsi permis d'améliorer la stabilité et la cohérence géométrique globale de la scène, notamment dans les régions difficiles à superviser par la seule information photométrique.

6.5 Analyse de l'effet de la fonction de coût de distorsion

La fonction de coût de distorsion est une régularisation déjà implémentée dans le pipeline 2D Gaussian Splatting, bien qu'elle soit désactivée par défaut dans l'implémentation originale. Son rôle est de limiter la dispersion des splats le long d'un même rayon, pour améliorer la fidélité géométrique locale des reconstructions.

La fonction de coût de distorsion vise à concentrer la distribution de poids le long du rayon en minimisant les distances entre les splats actifs sur un même rayon. Formellement, cette régularisation s'écrit :

$$\mathcal{L}_d = \sum_{i,j} \omega_i \omega_j \, |z_i - z_j|$$

où z_i et z_j sont les profondeurs des splats i et j sur un même rayon, et ω_i, ω_j leurs poids d'accumulation respectifs.Cette formulation favorise la compacité géométrique des splats autour des surfaces visibles, améliorant ainsi la netteté des reconstructions.

Dans le cadre de notre travail, nous avons mené une analyse approfondie de l'effet de cette fonction de coût sur la qualité des reconstructions, notamment lors de l'ajout de vues éloignées et de scènes à grande échelle.

Dans notre pipeline, ce terme a été intégré au coût total d'optimisation, pondéré par un facteur λ_d. À la différence de l'implémentation d'origine, nous avons systématiquement activé cette régularisation pour toutes les scènes, en conservant un facteur constant. Cette généralisation est rendue possible par l'utilisation systématique d'un repère métrique dans nos reconstructions, garantissant une échelle cohérente entre les scènes.

Lorsque cette fonction de coût est activée avec une pondération adéquate, elle permet de préserver efficacement les détails géométriques sans introduire de dégradation dans les cas où elle n'est pas nécessaire. Elle constitue donc un mécanisme robuste, applicable par défaut dans des contextes multi-échelles.

L'analyse expérimentale de l'impact de cette régularisation est détaillée dans la section 8.5.

7 Données utilisées

Pour évaluer et entraîner les méthodes développées durant ce stage, plusieurs jeux de données ont été constitués à partir de prises d'images réalisées en interne.

7.1 Jeux de données

7.1.1 Séquence ZED stéréo

Cette séquence a été spécialement enregistrée avec une caméra stéréo ZED, en tournant autour d'objets présentant des détails fins et complexes à reconstruire. Elle comprend 213 images en paires stereo, dont 157 prises à proximité des objets. La résolution des images est de 672 × 376 pixels, la baseline de la caméra est de 12 cm, et les paramètres de calibration ont été

obtenus à l'aide d'une mire. L'objectif était d'évaluer l'impact de la variation d'échelle des cartes de profondeur sur la qualité de la reconstruction.

7.1.2 Séquence *Atelier électrique*

Cette séquence comprend 286 images en paires stéréo, avec une résolution de 1920 × 1200 pixels. La scène filmée est plus vaste que celle de la séquence ZED et présente des conditions complexes pour la reconstruction : le sol et un des murs de la pièce présentent des spécularités importantes, rendant la réflexion de la lumière variable selon la position de la caméra. De plus, le sol et le mur sont peu texturés le mur étant blanc avec une couleur dépendant de l'angle de prise de vue. La scène est très dense en objets de toutes tailles. Les images ont été prises à différentes distances, proches et éloignées, pour capturer la globalité de la scène. Les paramètres intrinsèques des caméras ainsi que leurs poses relatives doivent être estimés durant le pipeline, ce qui reflète des conditions réalistes et complexes d'acquisition.

7.2 Conditions de prise de vue

Les séquences ont été acquises dans des conditions proches de celles rencontrées en situation réelle :

- Les paramètres intrinsèques et extrinsèques des caméras ne sont pas toujours fournis et doivent être estimés.
- Les trajectoires sont non linéaires, avec des mouvements complexes autour ou à l'intérieur des scènes.
- Les conditions d'éclairage, les occlusions et la diversité des textures rendent la reconstruction plus difficile.

7.3 Types de données manipulées

Le pipeline manipule différents types de données, depuis les images brutes jusqu'à la forme finale exploitée pour des applications 3D :

- **Images stéréo** : paires d'images capturées par des caméras avec ou sans leurs calibrations, base de toute reconstruction.
- **Cartes de disparité** : matrices où chaque pixel indique le déplacement horizontal entre vues stéréoscopiques, utilisées pour estimer la géométrie de la scène.
- **Cartes de profondeur** : transformation métrique des disparités en distances réelles, exprimées en mètre.
- **Nuages de points denses** : ensembles de points 3D obtenus par reprojection des cartes de profondeur, fournissant une approximation spatiale dense.
- **Poses de caméra** : informations de position et d'orientation des caméras dans un repère commun, indispensables pour la cohérence spatiale.

- **Reconstruction 2D Gaussian Splatting** : représentation ide la scène sous forme de splats gaussiens 3D, utilisée pour le rendu différentiable et la visualisation.
- **Maillage polygonal** : représentation de la surface 3D extraite en post-traitement, souvent via une méthode volumétrique comme la fonction de distance signée tronquée (TSDF). Ce maillage est la forme finale exploitée pour de nombreuses applications et constitue un objectif majeur d'amélioration dans ce travail.

8 Résultats

Cette section détaille les résultats obtenus avec notre pipeline de reconstruction, en mettant l'accent sur les améliorations apportées par les différentes étapes du processus. Nous analysons notamment l'impact de la rectification stéréo, du processus de masquage d'objet,de l'initialisation par nuage dense,de la supervision par cartes de profondeur et de la fonction de coût de distorsion sur la qualité finale des reconstructions.

8.1 Impact de la rectification stéréo sur la validité des cartes de profondeur

Une analyse comparative a été menée pour évaluer l'effet de la rectification stéréo dans le pipeline de reconstruction. Pour cela, nous avons comparé la qualité des cartes de profondeur obtenues :

- sans rectification (projection directe des paires stéréo non rectifiées),
- avec rectification stéréo (incluant le recentrage du point principal).

Le critère évalué est le taux de pixels invalidés par le cross-check (`NaN`) dans les cartes de profondeur générées par correspondance stéréo. Les résultats sont les suivants :

Séquence	**Sans rectification**	**Avec rectification**
ZED	55.21%	**16.39%**
Atelier électrique	78.28%	**38.66%**

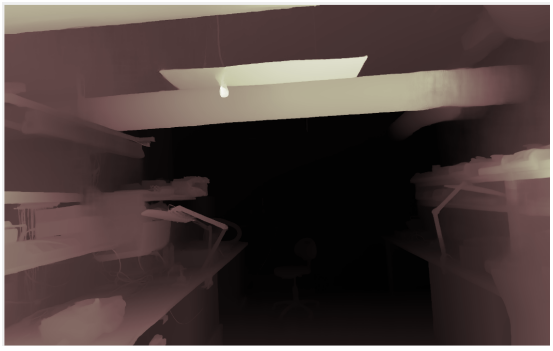
TABLE 1 – Taux de pixels invalidés (`NaN`) par le cross-check dans les cartes de profondeur avec et sans rectification.

Ces résultats montrent que la rectification est essentielle pour garantir des correspondances stéréo fiables, en réduisant fortement la proportion de pixels non reconstruits. Elle assure notamment une meilleure compatibilité géométrique entre les paires d'images utilisées par le pipeline de profondeur.

Pour illustrer ces observations, la figure 6 présente les cartes de profondeur obtenues pour la séquence *Atelier électrique* dans quatre configurations différentes, combinant l'effet de la rectification et celui du *cross-check* :

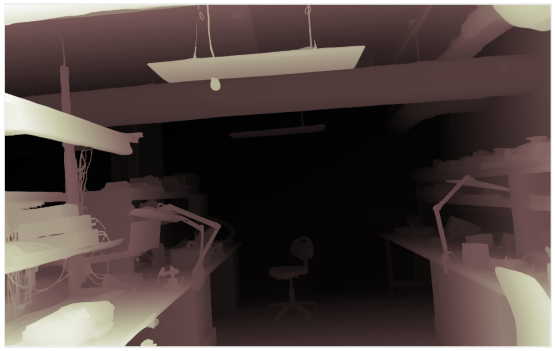
- Sans rectification, sans cross-check,
- Sans rectification, avec cross-check,

- Avec rectification, sans cross-check,
- Avec rectification, avec cross-check.

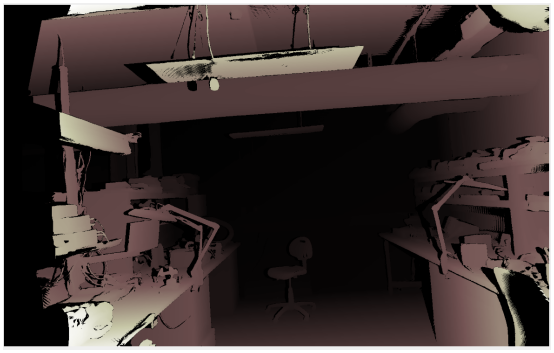


Sans rectification, sans cross-check

Sans rectification, avec cross-check



Avec rectification, sans cross-check



Avec rectification, avec cross-check

FIGURE 6 – Comparaison visuelle de l'effet du cross-check et de la rectification sur la carte de profondeur (séquence Atelier électrique).

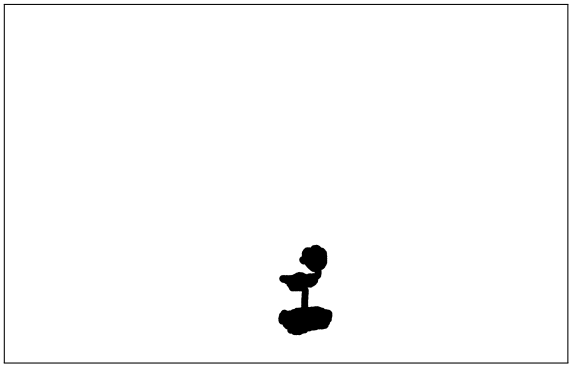
8.2 Résultat de la suppression d'objets par masquage : cas de la chaise

Dans la scène *Atelier électrique*, une chaise visible dans plusieurs vues a été ciblée pour suppression à l'aide de masques manuels. Ces masques ont été dessinés individuellement sur chaque image stéréo afin d'exclure les pixels correspondants de la supervision photométrique.

Deux configurations d'entraînement ont été comparées :

- **Sans masquage** : entraînement sur toutes les régions de l'image, y compris celles contenant la chaise.
- **Avec masquage** : désactivation de la supervision RGB sur les pixels de la chaise à l'aide des masques manuels.

Image originale



Masque manuel (noir : zones ignorées)

FIGURE 7 – Exemple de masque manuel utilisé pour exclure la chaise de la supervision photométrique.

Reconstruction sans masquage : chaise visible

Reconstruction avec masquage manuel : chaise supprimée

FIGURE 8 – Effet du masquage photométrique sur la suppression de la chaise. La reconstruction finale montre une disparition quasi complète de l'objet ciblé, démontrant l'efficacité de cette approche pour l'édition locale de la scène.

L'objet masqué a été complètement éliminé du rendu final, sans laisser de trace notable dans le maillage reconstruit. Cette expérience confirme que le masquage photométrique constitue une solution simple et efficace pour supprimer sélectivement des objets dans la scène. Il ouvre la voie à des applications comme l'anonymisation ou le nettoyage de scènes 3D.

À noter : le masquage n'affecte pour l'instant que la supervision RGB. Une amélioration future consisterait à appliquer également ces masques à la supervision par cartes de profondeur, afin de renforcer le contrôle sur la reconstruction dans les zones ciblées.

8.3 Apport de l'initialisation par nuage de points dense

Afin d'évaluer qualitativement l'impact de l'utilisation d'un nuage de points dense, issu des cartes de profondeur stéréo, pour l'initialisation des splats dans le pipeline 2D Gaussian Splatting, nous présentons plusieurs comparaisons visuelles réalisées sur la séquence *Atelier électrique*.

Les images ci-dessous illustrent les résultats obtenus selon deux configurations principales :

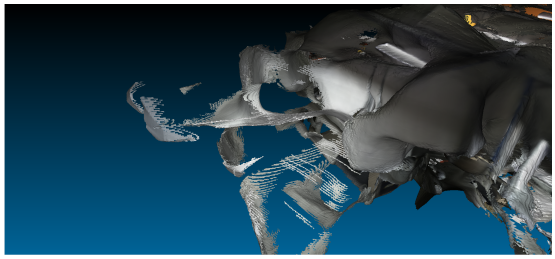
- **Initialisation sans nuage dense** : les splats sont initialisés à partir d'un nuage sparse provenant de la reconstruction SfM.
- **Initialisation avec nuage dense** : les splats sont initialisés à partir du nuage dense reprojeté des cartes de profondeur stéréo.

Cependant, on observe toujours une certaine dérive au cours de l'optimisation, qui empêche une reconstruction parfaitement stable du sol et du mur, comme le montrent les légères incohérences et déformations visibles, même avec l'initialisation dense (voir Figure 9) .

Bien que cette évaluation reste qualitative, les améliorations visuelles observées confirment l'intérêt de l'utilisation des cartes de profondeur pour générer un nuage dense en amont de l'optimisation des splats.

(a) Initialisation sans nuage dense, vue globale du maillage

(b) Initialisation avec nuage dense, vue globale du maillage



(c) Initialisation sans nuage dense, vu extérieur du maillage

(d) Initialisation avec nuage dense, vu extérieur du maillage

FIGURE 9 – Comparaison visuelle des maillages extraits après reconstruction 2D Gaussian Splatting sur la séquence *Atelier électrique*, avec et sans initialisation par nuage dense. (a) et (b) montrent la scène globale où l'on observe une meilleure cohérence géométrique avec l'initialisation dense. (c) et (d) correspondent à l'extérieur de la scène au niveau du mur blanc. Sans initialisation dense, le maillage présente une forte dispersion des splats, traduisant une dérive importante due à une mauvaise position initiale. Ces images confirment que l'initialisation par nuage dense améliore la stabilité géométrique et la qualité finale de la reconstruction.

8.4 Apport de la supervision par cartes de profondeur dans l'entraînement

Pour limiter la dérive des splats causée par les variations d'éclairage, les changements de couleur selon l'angle de vue et les phénomènes de spécularité, nous avons intégré une supervision explicite basée sur les cartes de profondeur dans la fonction de coût lors de l'entraînement.

Cette approche permet de corriger les erreurs géométriques dues à une supervision photométrique seule, en stabilisant l'optimisation et en améliorant la cohérence globale du maillage reconstruit.

Les images ci-dessous illustrent les résultats obtenus selon deux configurations principales :

- **Sans supervision par cartes de profondeur** : reconstruction initialisée par nuage dense, entraînée sans utiliser les cartes de profondeur.
- **Avec supervision par cartes de profondeur** : reconstruction initialisée par nuage dense, entraînée avec intégration des cartes de profondeur dans la fonction de coût.

On constate que l'intégration des cartes de profondeur dans la fonction de coût d'entraînement stabilise notablement la reconstruction, notamment sur les zones difficiles telles que le mur blanc et le sol, en supprimant une grande partie des incohérences géométriques liées aux perturbations photométriques.

Cette amélioration qualitative confirme l'intérêt d'une supervision géométrique supplémentaire pour renforcer la robustesse et la précision des reconstructions par 2D Gaussian Splatting.

8.5 Effet de la fonction de coût de distorsion sur la reconstruction

Cette section présente une analyse expérimentale de l'effet de la régularisation par fonction de coût de distorsion introduite dans la section 6.5. L'objectif est de valider son impact sur la qualité géométrique des reconstructions, en particulier lors de l'ajout de vues éloignées.

Pour évaluer l'apport de la fonction de coût de distorsion, nous avons utilisé la séquence stéréo ZED, capturée autour d'objets présentant des détails fins (cf. section 6). Chaque séquence comporte deux images représentatives, correspondant aux deux objets principaux observés, permettant d'analyser finement la qualité de reconstruction locale.

Trois configurations ont été comparées :

- **Séquence courte** : cette configuration utilise la même reconstruction SfM que la séquence longue, mais les images n'appartenant pas à la séquence courte sont simplement ignorées lors de l'apprentissage.
- **Séquence longue sans fonction de coût de distorsion** : l'ajout d'images éloignées provoque une perte notable de précision sur les objets.
- **Séquence longue avec fonction de coût de distorsion** : l'activation de la fonction de coût permet d'atteindre un niveau de détail supérieur à celui de la séquence courte.

Dans tous les cas, les cartes de profondeur issues de la stéréovision ont été intégrées dans le pipeline d'apprentissage, contribuant à la supervision géométrique.

Apprentissage sans supervision par cartes de profondeurs et avec initialisation dense, vue globale

Apprentissage avec supervision par cartes de profondeurs et avec initialisation dense, vue globale

Apprentissage sans supervision par cartes de profondeurs et avec initialisation dense, vu extérieur du maillage

Apprentissage avec supervision par cartes de profondeurs et initialisation dense, vu extérieur du maillage

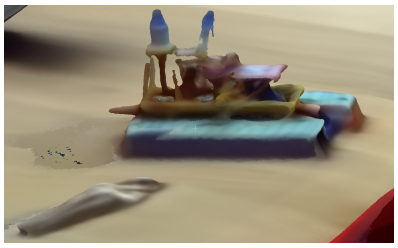
FIGURE 10 – Comparaison visuelle des reconstructions 2D Gaussian Splatting initialisées par nuage dense, sans puis avec intégration des cartes de profondeur dans la fonction de coût lors de l'entraînement. La supervision par profondeur réduit significativement les dérives causées par les conditions d'éclairage variables et les réflexions, améliorant la stabilité et la fidélité de la reconstruction.

Tous ces tests ont été réalisés avec les mêmes valeurs de poids pour la fonction de coût de distorsion. Il est important de noter que nos scènes étaient toutes exprimées dans un repère métrique cohérent, facilitant ainsi l'effet bénéfique de cette régularisation. En revanche, l'impact de cette fonction de coût dans un repère arbitraire, comme celui fourni directement par COLMAP sans transformation métrique préalable, n'a pas encore été exploré et pourrait nécessiter des adaptations spécifiques.

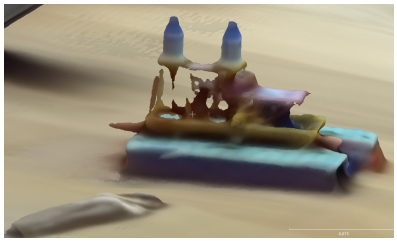
Objet 1, séquence courte

Objet 1, séquence longue sans distortion

Objet 1, séquence longue avec distortion



Objet 2, séquence courte



Objet 2, séquence longue sans distortion

Objet 2, séquence longue avec distortion

FIGURE 11 – Effet de la fonction de coût de distorsion sur la reconstruction des deux objets principaux de la séquence ZED.

Les résultats de la figure 11 illustrent clairement l'impact de la fonction de coût de distorsion :

- Dans la **séquence courte**, les détails fins des objets sont correctement reconstruits. La proximité constante des vues avec la scène évite les ambiguïtés sur la localisation des splats.
- Dans la **séquence longue sans fonction de coût de distorsion**, on observe une dégradation notable : certains éléments fins disparaissent ou sont mal définis.
- Dans la **séquence longue avec fonction de coût de distorsion**, les détails sont rétablis. La fonction de coût agit comme un terme de concentration : elle pousse les splats à se regrouper étroitement le long du rayon, ce qui empêche les configurations géométriques diluées. On a donc une qualité de reconstruction supérieur à celle obtenue dans la séquence courte.

Cette régularisation est donc précieuse pour préserver la fidélité géométrique lorsque des vues plus éloignées — ou simplement plus variées — sont utilisées. Cependant, nous avons également constaté que dans les cas où elle n'est pas nécessaire (comme la séquence courte), son activation n'introduit pas d'artefacts ni de dégradations. Elle se comporte donc comme une régularisation stable.

Enfin, cette analyse repose sur des scènes exprimées dans un repère métrique, ce qui garantit la cohérence des distances et des profondeurs. Dans un contexte où les reconstructions SfM ne sont pas transformées dans un repère physique, il resterait à évaluer l'effet de cette fonction de coût et, si nécessaire, à l'adapter en conséquence comme conseillé dans le 2DGS.

9 Conclusion

9.1 Bilan

Dans ce travail, nous avons exploré une approche innovante pour la reconstruction 3D à partir d'images stéréo, fondée sur la méthode de *2D Gaussian Splatting* avec une supervision explicite par cartes de profondeur. Cette supervision a permis d'améliorer significativement la cohérence géométrique et la précision de la reconstruction, en particulier dans les scènes complexes et les zones peu texturées.
L'initialisation des splats à partir d'un nuage de points dense, obtenu par reprojection des cartes de profondeur, a constitué une étape clé, réduisant la dérive et facilitant l'optimisation. La rectification stéréo adaptée a également contribué à une meilleure correspondance des images. Par ailleurs, l'utilisation des masques, initialement pensée pour exclure certaines zones problématiques, n'a pas apporté les bénéfices escomptés en termes de robustesse à l'optimisation. En revanche, leur réutilisation pour la suppression d'objets indésirables s'est révélée prometteuse pour l'édition de scènes. Une intégration plus poussée des masques, notamment dans la supervision par cartes de profondeur, pourrait améliorer la cohérence et l'efficacité de cette approche.
Pour limiter la dérive des splats causée par les variations d'éclairage, les changements de couleur selon l'angle de vue et les phénomènes de spécularité, nous avons intégré une supervision explicite basée sur les cartes de profondeur dans la fonction de coût lors de l'entraînement. Cette approche permet de corriger les erreurs géométriques dues à une supervision photométrique seule, en stabilisant l'optimisation et en améliorant la cohérence globale du maillage reconstruit.
Nous avons en particulier introduit une pondération de l'erreur en profondeur par l'inverse du carré de la profondeur stéréo, afin de compenser la perte naturelle de précision dans les zones éloignées. Cette stratégie permet de renforcer la supervision là où la géométrie est mieux contrainte, en insistant sur les zones proches où les erreurs seraient visuellement et structurellement plus importantes. Cette modification a conduit à une amélioration notable de la stabilité de l'optimisation et de la fidélité des reconstructions.
Une analyse spécifique a été menée concernant l'effet de la fonction de coût de distorsion, un terme de régularisation souvent négligé dans les configurations standard. Cette fonction de coût vise à concentrer les splats le long du rayon de vision, afin de compenser le fait que le rendu ne prend pas en compte leur distance relative dans l'espace. Nos résultats montrent qu'elle permet de préserver un haut niveau de détail même lorsque des vues éloignées sont intégrées à l'apprentissage. Elle s'est révélée particulièrement utile dans des configurations multi-échelles, tout en restant sans effet négatif notable lorsqu'elle n'était pas nécessaire. Tous les tests ont été réalisés dans un repère métrique bien défini, et des travaux futurs seront nécessaires pour confirmer la robustesse de cette régularisation dans des repères arbitraires issus directement de SfM.
Il est important de souligner que la comparaison avec les méthodes classiques n'a pas été réalisée de manière exhaustive. Seuls quelques tests exploratoires ont été menés, insuffisants pour fournir une évaluation quantitative robuste. De même, une comparaison directe avec les méthodes à l'état de l'art reste impossible pour l'instant, en raison de l'absence de code source disponible.

9.2 Perspectives

Plusieurs axes d'amélioration et de recherche s'offrent à nous pour prolonger ce travail.

D'abord, la compatibilité entre l'utilisation des masques et des cartes de profondeur pourrait être approfondie, permettant une suppression plus précise et géométriquement cohérente des éléments non désirés dans la scène.
Ensuite, il serait crucial de réaliser une évaluation plus complète, incluant une comparaison rigoureuse avec les méthodes classiques ainsi que les techniques à l'état de l'art, dès que celles-ci seront accessibles. Cette étape est essentielle pour valider pleinement l'apport de notre méthode.
Enfin, le comportement de certaines régularisations (comme la fonction de coût de distorsion) dans des repères non métriques ou arbitraires reste un point ouvert. Leur efficacité pourrait dépendre de la normalisation des profondeurs ou d'un recalage spatial préalable, ce qui constitue un axe d'expérimentation important.
Un autre prolongement intéressant serait d'étendre le pipeline au cas monoculaire. L'idée serait de générer des paires pseudo-stéréo en rectifiant entre elles des vues successives issues d'une seule vidéo. Ce scénario a été partiellement testé, mais a rencontré des limitations dues à la rectification stéréo d'OpenCV, qui impose certaines contraintes géométriques difficiles à satisfaire en monoculaire. Malgré ces obstacles, cette piste reste prometteuse, notamment pour l'adaptation du pipeline à des données plus accessibles et moins contraintes.
De plus, cette même logique de rectification pourrait être exploitée dans un cadre stéréo classique pour augmenter la couverture multi-vues : en créant des paires supplémentaires entre images non adjacentes mais spatialement compatibles, on pourrait enrichir la supervision géométrique et renforcer la cohérence des reconstructions.

Références

[1] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[2] Antoine Guédon, Diego Gomez, Nissim Maruani, Bingchen Gong, George Drettakis, and Maks Ovsjanikov. Milo : Mesh-in-the-loop gaussian splatting for detailed and efficient surface reconstruction, 2025.

[3] Binbin Huang, Zehao Yu, Anpei Chen, Andreas Geiger, and Shenghua Gao. 2d gaussian splatting for geometrically accurate radiance fields. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, SIGGRAPH '24, page 1–11. ACM, July 2024.

[4] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (SIGGRAPH Conference Proceedings)*, 42(4), July 2023.

[5] Jungeon Kim, Geonsoo Park, and Seungyong Lee. Multiview geometric regularization of gaussian splatting for accurate radiance fields, 2025.

[6] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything, 2023.

[7] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf : Representing scenes as neural radiance fields for view synthesis, 2020.

[8] Sadra Safadoust, Fabio Tosi, Fatma Güney, and Matteo Poggi. Self-evolving depth-supervised 3d gaussian splatting from rendered stereo pairs, 2024.

[9] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[10] Johannes Lutz Schönberger, Enliang Zheng, Marc Pollefeys, and Jan-Michael Frahm. Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)*, 2016.

[11] Zachary Teed and Jia Deng. Raft : Recurrent all-pairs field transforms for optical flow. In *European conference on computer vision (ECCV)*, pages 402–419. Springer, 2020.

[12] Bowen Wen, Matthew Trepte, Joseph Aribido, Jan Kautz, Orazio Gallo, and Stan Birchfield. Foundationstereo : Zero-shot stereo matching, 2025.

[13] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3d : A modern library for 3d data processing. In *Proceedings of the 2018 Eurographics Workshop on 3D Object Retrieval*, 2018.