



HAL
open science

Augmentation de la résolution latérale d'imageurs laser 3D par compressive sensing

Viala Erwan

► **To cite this version:**

Viala Erwan. Augmentation de la résolution latérale d'imageurs laser 3D par compressive sensing. Physique [physics]. ISAE - Institut Supérieur de l'Aéronautique et de l'Espace, 2023. Français. NNT : 2023ESAE0049 . tel-04661379

HAL Id: tel-04661379

<https://hal.science/tel-04661379v1>

Submitted on 24 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Institut Supérieur de l'Aéronautique et de l'Espace

Présentée et soutenue par :

Erwan VIALA

le lundi 9 octobre 2023

Titre :

Augmentation de la résolution latérale d'imagers laser 3D par compressive sensing

Increasing lateral resolution of 3D-LiDAR using compressive sensing

École doctorale et discipline ou spécialité :

ED AA : Physique

Unité de recherche :

Équipe d'accueil ISAE-ONERA OLIMPES

Directeur(s) de Thèse :

M. Laurent RISSER (directeur de thèse)
M. Nicolas RIVIÈRE (co-directeur de thèse)

Jury :

M. Jean-François AUJOL Université de Bordeaux - Examineur, Président
M. Adrian BASARAB INSA Lyon - Rapporteur
M. Paul-Édouard DUPOUY ONERA - Co-encadrant
M. Bertrand LE SAUX ESA/ESRIN - Rapporteur
M. Laurent RISSER Université Toulouse III Paul Sabatier - Directeur de thèse
M. Nicolas RIVIÈRE ONERA - Co-directeur de thèse

Remerciements

Tout d'abord, j'aimerais remercier le jury, en commençant par Adrian Basarab et Bertrand le Saux, qui ont accepté de rapporter ma thèse. Puis, Jean-François Aujol qui, malgré des conditions complexes liées à une activité incertaine de la SNCF, a réussi à m'écouter pendant plus de 45 minutes. De même, je remercie vivement Olivier Meyer pour avoir accepté de participer à mon jury.

Ensuite, j'aimerais remercier mon encadrement de thèse, Laurent, Nicolas, Paul-Édouard qui m'ont soutenu, tout au long de cette thèse. J'ai grandement apprécié de pouvoir échanger avec vous durant ces trois ans, que ce soit autour d'un café à l'IMT, le soir en partant du DOTA ou encore au labo entre deux suggestions musicales.

Pour tous les autres, comme c'est un exercice pour lequel je n'ai aucune appétence et je m'en excuse, je me contenterai d'un – peut-être trop sobre - : Merci à tous !

Table des matières

Introduction/Contexte	3
1 Notions	8
1.1 Systèmes imageurs LiDAR 3D	10
1.1.1 Introduction à l'imagerie laser	10
Fonctions de détection, reconnaissance et identification	11
Systèmes LiDAR 3D	13
Imagerie laser 3D embarquée sur plateformes mobiles	15
Concepts de l'imagerie LiDAR 3D à comptage de photons	20
1.1.2 L'équation lidar	22
1.2 La caméra single-pixel	26
1.2.1 Principe de fonctionnement	26
1.2.2 Élément de modulation spatiale du signal détecté	27
1.3 Méthodes de <i>Compressive Sensing</i>	29
1.3.1 Algorithmes classiques de résolution	33
Comparaison des algorithmes	38
1.3.2 Le <i>Compressive Sensing</i> et l'équation lidar	39
1.4 Logiciel MATLIS : simulations de signaux issus de LiDAR 3D	42
1.4.1 Utilisation et propriétés	42
Construction de l'impulsions laser	42
Propagation atmosphérique (aller et retour)	42
Interactions avec la scène	43
Modèle capteur	43
1.4.2 Définition des cas tests	43
1.5 Synthèse	45
2 CRISPS-3D : une chaîne de traitement générale pour le <i>Compressive Sensing</i> appliquée aux matrices de photodiodes à avalanches en mode Geiger	46
2.1 Présentation et motivations de CRISPS-3D	48
2.2 Algorithmes de pré-traitement	51
2.2.1 Reconstruction de l'onde complète	51
État-de-l'art sur le DTE	52
Correction du DTE dans CRISPS-3D	52
2.2.2 Estimation du support de la fonction	54
2.3 Algorithme de <i>Compressive Sensing</i>	57
2.3.1 Version optimisée : <i>Orthogonal Matching Pursuit</i> (OMP) utilisant l'approximation de Cholesky	58
2.4 Validation de la méthode sur des signaux simulés	59

2.4.1	Reconstruction de l'onde complète	59
2.4.2	Estimation de la fonction du support	62
2.4.3	Chaîne de traitement complète	63
2.5	Synthèse	67
3	Prise en compte dans CRISPS-3D d'impulsions dont la longueur temporelle est supérieure à la taille d'un pas de temps	68
3.1	Utilisation d'impulsions larges avec déconvolution	70
3.1.1	Motivation et contexte	70
3.1.2	Définition du problème de déconvolution avec <i>Compressive Sensing</i>	71
3.1.3	Solutions existantes au problème de déconvolution parcimonieuse dans le cadre d'acquisitions LiDAR 3D	72
3.1.4	Limitation des durées d'impulsions possibles due au DTE	74
3.2	Définition d'un algorithme de déconvolution pour CRISPS-3D	76
3.2.1	Déconvolution <i>via</i> une approximation parcimonieuse	76
	Application de l'estimation de profondeur dans CRISPS-3D	76
3.2.2	Amélioration de la méthode de déconvolution	79
3.2.3	Validation de la méthode sur des signaux simulés	80
3.3	Ajout d'hypothèses sur la régularité spatiale pour améliorer les résultats .	85
3.4	Synthèse	86
4	<i>Compressive Sensing</i> et mouvement	87
4.1	Estimation du mouvement <i>a posteriori</i>	88
4.1.1	Méthodes de recalage utilisant le transport optimal pour l'estimation de mouvement 3D	88
	Étape du calcul du plan de transport avec transport optimal	90
	Explication plus détaillée des paramètres de la méthode	92
4.1.2	Application de méthode de FLOT avec intensités sur des acquisitions LiDAR 3D simulées	92
4.2	<i>Compressive sensing</i> et mouvement entre l'acquisition des motifs	96
4.2.1	Modélisation du problème de <i>Compressive Sensing</i> simplifié prenant en compte le mouvement	96
4.2.2	Définition d'une stratégie de reconstruction d'image en utilisant l'approche <i>Large Deformation Diffeomorphic Metric Mapping</i>	98
	Définition de LDDMM	99
	Stratégie expérimentale de reconstruction de mouvement avec <i>Compressive Sensing</i>	101
4.2.3	Résultats dans le cas 2D	103
4.3	Synthèse	105
	Conclusion	107
	A Équivalence des corrections du DTE	110

Introduction/Contexte

L'observation à très longue portée (au-delà de 10 kilomètres) est nécessaire pour anticiper des menaces potentielles lors de la surveillance de sites [48, 106] (détection de drones), la planification de l'espace aérien ou maritime [60, 73] (identification de débris flottants, auto-protection de navires). La fonction d'identification, capacité d'identifier un objet, requiert une bonne résolution télémétrique et latérale des imageurs. Les besoins opérationnels imposent un maintien de la fonction de détection et/ou d'identification par tout temps et quelles que soient les conditions d'éclairement (jour, nuit, éblouissement solaire). Les techniques d'imageries dites « passives » c'est-à-dire qui n'ont pas de source d'illumination propre, dans le visible ou l'infrarouge répondent partiellement à la fonction d'identification [15]. Les opérateurs font états de contraintes lors du déploiement de ces imageurs, notamment par faible visibilité ou forte variation de l'éclairement solaire dans le temps. L'imagerie laser 3D y est moins sensible par l'emploi et la maîtrise de sa propre source d'illumination [82, 105, 147], comme illustré dans la Figure 1. La télémétrie, littéralement mesure de la distance, dans l'image apporte une dimension supplémentaire qui facilite l'identification. En effet, en imagerie laser 3D la distance des objets est directement mesurée, là où elle ne peut-être qu'estimée en imagerie passive (photogrammétrie [34, 80, 87]). De plus, bien que la photogrammétrie permette de reconstruire une scène en 3D à partir de quelques images [80] et parfois une seule [34, 87]. Elle nécessite un éclairage constant et suffisant entre chaque image [80].

L'imageur laser à plan focal 3D, qui utilise une matrice de pixels au plan focal, est le meilleur candidat pour l'identification à très longue portée [157]. Il met en œuvre des détecteurs sensibles (comptage de photons) et une excellente résolution temporelle, c'est-à-dire une forte résolution en profondeur de l'image (centimétrique). Les plans focaux 3D disponibles sont tous limités à quelques centaines de pixels de grandes tailles (20–50 μm) par rapport au standard en imagerie passive (4 μm). Cette limitation est essentiellement aux difficultés techniques liées à répartition de la puissance entre les pixels ainsi que la largeur de bande de lecture des données. Dans l'état de l'art, les plans focaux 3D à comptage de photon sont limités à 256×256 pixels pour le visible et 128×128 pixels pour l'infrarouge [9]. La résolution latérale ou le champ de vue en sont fortement contraints. Malgré leur faible résolution latérale, les caméras à comptage de photons sont de plus en plus utilisées dans les systèmes d'acquisitions LiDAR 3D, en particulier lorsqu'il est important d'avoir un système embarqué avec un SWaP-C (*Size, Weight and Power-Cost*) faible.

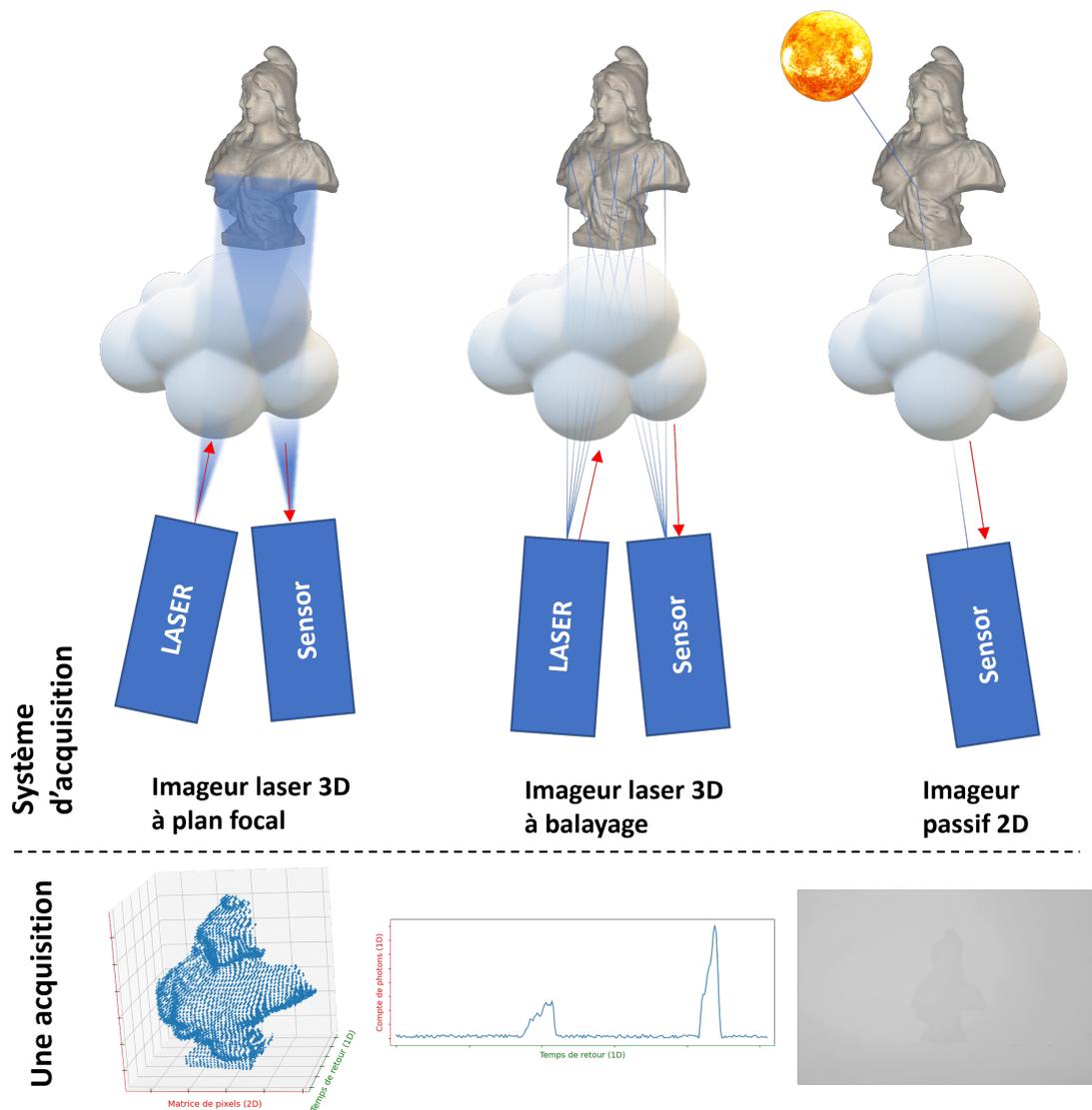


FIGURE 1. Comparaison entre imageur laser 3D (ou LiDAR 3D) à plan focal, à balayage et imageur passif 2D.

Bien que les plans focaux 3D aient une meilleure résolution latérale dans le visible, nous avons choisi de nous concentrer sur les systèmes utilisant des plans focaux dans l'infrarouge, et plus spécifiquement sur les systèmes utilisant des capteurs dit GmAPD (*Geiger-mode Avalanche Photo-Diode*). Les avantages liés à l'utilisation de caméras infrarouges, sont notamment la sécurité oculaire à la longueur d'onde d'intérêt ($1,5\ \mu\text{m}$) [167] ainsi que la capacité à traverser certains obscurants comme par exemple le brouillard [147]. L'utilisation de cette technologie de plans focaux 3D limite fortement la résolution latérale et/ou le champ de vue. Pour permettre de répondre à la fonction d'identification, il est primordial de les augmenter. Actuellement, une solution pour augmenter le champ de regard est la mise en place de techniques de balayage (voir Figure 1), sans modification de la résolution latérale intrinsèque au capteur.

En imagerie, il est admis que les données déjà acquises sont structurées et donc compressibles : cette approche est à la base des formats JPEG par exemple. L'intégration directe de la compression dans le procédé d'acquisition des données permet d'éviter l'acquisition de données « redondantes ». Le *Compressive Sensing* (acquisition comprimée en français) adresse cette problématique [29, 30, 32]. C'est un cadre théorique en traitement du signal qui permet l'acquisition d'un signal avec un nombre réduit de mesure de combinaison linéaire d'un signal compressible. Nous généralisons dans cette thèse l'utilisation du *Compressive Sensing* à l'imagerie laser 3D, afin d'augmenter la résolution des images tout en limitant l'augmentation du nombre d'acquisitions en comparaison avec des techniques de balayage. Cette approche est novatrice et très peu développée dans la littérature car elle associe les techniques de *Compressive Sensing* à de l'imagerie laser à base de plans focaux 3D.

Pour l'imagerie LiDAR 3D, le *Compressive Sensing* consiste à trouver une collection de modulations spatiales minimisant le nombre de mesures à effectuer pour reconstruire une image. Cette collection définit un processus de mesure correspondant à un système linéaire sous-déterminé. La reconstruction est alors obtenue en résolvant algorithmiquement un problème d'optimisation inversant le processus de mesure en ajoutant une hypothèse de régularité sur la parcimonie des données dans une certaine base. Le *Compressive Sensing* permet ainsi d'obtenir une meilleure résolution latérale tout en limitant l'impact de la méthode sur la fréquence d'acquisition. Cette technique permet de maintenir une bonne fidélité du signal [32] et est applicable à l'imagerie active pour augmenter la résolution latérale [55].

Le *Compressive Sensing* a déjà été appliqué à des caméras à comptage de photons dans le cas des caméras monopixel [55, 58, 59, 67, 74, 77, 78, 130, 142]. De même, d'autres méthodes d'augmentation de résolution tel que l'interpolation la super-résolution ou le balayage ont déjà été appliquées avec succès à longue ou moyenne portée [90, 92, 113, 144]. Cependant, les différentes méthodes de l'état de l'art ne permettent pas actuellement des acquisitions à très longue portée (au-delà de 10 km par exemple) utilisant une matrice de capteur GmAPD en temps d'acquisitions limité (< 1 s, temps alloué pour collecter des photons sur le capteur) avec un faible SWaP-C et un niveau de signal très faible (inférieur à un photon par pixel en moyenne). De plus, à ce jour, l'application du *Compressive Sensing* à l'observation de cibles dynamiques et son intégration sur des plateformes mobiles restent un défi en termes de traitement du signal et d'implémentation.

L'objectif de cette thèse est donc d'**augmenter la résolution latérale de plan focaux 3D pour l'acquisition LiDAR 3D en utilisant le *Compressive Sensing*** sous ces contraintes avec une latence aussi faible que possible (1 ms – 1 s). Pour ce faire, les travaux réalisés dans le cadre de cette thèse sont déclinés selon 2 axes de recherche :

1. Développement d'une chaîne de traitement nommée CRISPS-3D (*Compressive Reconstruction of Images for Single Photon System in 3D*), permettant d'augmenter la résolution de plans focaux 3D dans le cas statique.

Ce premier axe se concentre sur le développement d'une chaîne de traitement augmentant la résolution latérale de plans focaux 3D et plus spécifiquement des GmAPD en utilisant des méthodes de *Compressive Sensing*, adaptables aux spécificités du système optronique. Dans ce premier axe, les mesures seront supposées statiques ou quasi-statiques durant l'acquisition.

2. Développement d'une chaîne complète pour des applications en mouvement.

Ce deuxième axe vise à généraliser la chaîne de traitement précédente pour prendre en compte le mouvement des cibles et/ou la dynamique du porteur (ex. avion, drone, véhicule terrestre).

Tout au long de ces travaux nous supposerons que les turbulences, des perturbations du trajet optique liées à l'inhomogénéité de l'atmosphère, sont négligeables. Cependant, cela limite les applications possibles, notamment lorsqu'on considère des acquisitions à longue portée. L'intégration et la prise en compte des effets liées à la turbulence dans la chaîne de traitement est un véritable challenge en soit. Cela ne sera pas développé ici. Toutefois, il existe des contextes pratiques où les turbulences peuvent être négligées comme les acquisitions réalisées depuis un point haut (quelques dizaines de mètres [140], ex. cas de l'aéroporté ou de la prise de vue depuis un bâtiment).

Le premier chapitre de la thèse est dédié à l'introduction des différentes notions nécessaires à la compréhension des travaux. Le principe de fonctionnement des LiDAR et des caméras GmAPD est explicité, ainsi que le *Compressive Sensing*. Pour chacune de ces notions, un état de l'art est dressé. Puis, les outils déjà disponibles à l'ONERA sont détaillés, notamment le logiciel MATLIS permettant de simuler des acquisitions LiDAR 3D.

Le second chapitre présente la chaîne de traitement CRISPS-3D qui permet, à partir de données brutes, d'obtenir un nuage de points 3D à plus forte résolution latérale que celle de la caméra. Notre chaîne de traitement utilise en entrée une collection d'acquisitions LiDAR 3D modulées spatialement et produit un nuage de points en trois dimensions. Dans

ce chapitre, il est montré qu'un certain nombre d'algorithmes de traitement du signal est nécessaire, afin de maximiser le gain lié au *Compressive Sensing*. Ces algorithmes, adaptés au contexte d'acquisitions LiDAR 3D à comptage de photons, sont détaillés dans cette partie. Telle que décrite dans ce chapitre, CRISPS-3D permet de reconstruire convenablement la géométrie des objets si l'impulsion est courte temporellement par rapport à la résolution temporelle du capteur. Le traitement de cette limitation sur la durée des impulsions est développé au cours du troisième chapitre.

Le troisième chapitre est consacré à la prise en compte dans CRISPS-3D d'impulsions longues temporellement par rapport à la résolution temporelle du capteur c'est-à-dire supérieure à la durée d'un pas de temps. En effet, il est nécessaire de déconvoluer temporellement le signal entrant afin de reconstruire au mieux la géométrie des objets. Cela est démontré dans ce chapitre avant de préciser un algorithme permettant de corriger cet effet. L'introduction d'un algorithme de déconvolution dans CRISPS-3D implique des interactions non-triviales et spécifiques au *Compressive Sensing* et aux algorithmes de traitement du signal. Ces interactions sont développées dans ce chapitre, pour définir l'algorithme de déconvolution et donner un domaine d'application. Une fois CRISPS-3D améliorée par l'introduction de la déconvolution temporelle, les nuages de points 3D obtenus en fonction des méthodes utilisées sont comparés.

Dans le quatrième et dernier chapitre, les acquisitions dynamiques avec *Compressive Sensing* sont abordées. Tout d'abord, comme un problème de recalage de nuage de points, lorsque la vitesse des cibles ou du système n'est pas suffisamment importante pour impacter une acquisition complète. Ensuite, on suppose une vitesse relative des cibles et du système suffisante pour qu'une acquisition soit impactées. Les méthodes de recalages ne permettent alors plus de corriger le mouvement. Dans ce cadre, une méthode pour estimer et prendre en compte le mouvement dans une acquisition LiDAR 3D simplifiée avec *Compressive Sensing* est définie.

Chapitre 1

Notions

Points clés

- Introduction de différentes notions clés telles que le *Compressive Sensing*, les caméras *single-pixel* ou encore les différents principes physiques liés aux détecteurs ou à l'acquisition de signal et d'images plus largement.
- Définition des notations utilisées pour caractériser les grandeurs physiques et les matrices utilisées dans les chapitres suivants.
- Développement de l'équation lidar pour le *Compressive Sensing*, pour lier la physique de l'acquisition avec le *Compressive Sensing* et les différentes méthodes développées par la suite.
- Description des approches existantes dans le cadre d'acquisition d'image 3D infrarouge, d'images infrarouges longue portée et des caméras *single-pixel*.

1.1 Systèmes imageurs LiDAR 3D

1.1.1 Introduction à l'imagerie laser

L'objectif de la thèse est d'augmenter la résolution spatiale de systèmes imageurs laser pour de l'observation à très longue portée en appliquant une méthode d'acquisition spécifique et un traitement du signal adapté. Cette méthode doit être compatible d'une utilisation dans un contexte opérationnel c'est-à-dire répondre au déploiement du système complet sur un théâtre d'opération et correspondre aux besoins exprimés par les utilisateurs finaux. Ces termes font référence aux besoins de l'opérateur au sens usuel du terme : l'utilisateur final est potentiellement formé à l'utilisation du système imageur mais est non spécialiste des algorithmes de traitement du signal et de l'information abordés dans ce manuscrit. Les besoins opérationnels sont définis à partir des retours opérateurs effectués dans le cadre de projets ONERA en amont des travaux de cette thèse.

Dans notre démarche, nous développons et appliquons des méthodes de télédétection pour collecter des informations physiques d'une cible à distance sans aucun contact [99]. La géométrie précise des objets et l'intensité qui dépend d'un certain nombre de paramètres physiques des surfaces observées sont directement accessibles (voir Section 1.1.2).

Il est possible de distinguer la télédétection dite « passive » de la télédétection « active ». La première approche nécessite un éclairage naturel ou considère l'émission propre des objets [43]. Une information radiométrique est collectée à distance sur le capteur puis, après étalonnage de l'instrument, l'information physique de la surface observée est déduite. Dans le cadre des applications visées, les techniques d'imagerie passive souffrent de quelques inconvénients liés notamment à la non-maîtrise de l'environnement et à son influence directe sur le signal utile [14]. La seconde approche dite télédétection « active » est complémentaire. Elle considère un éclairage laser de la scène à partir d'une source colocalisée avec le détecteur [24]. Elle s'affranchit des problématiques de maîtrise de l'éclairage naturel et autorise des acquisitions par faible visibilité, de jour comme de nuit [105]. La télémétrie par temps de vol [156] ou par d'autres méthodes d'acquisition [157] est native et n'implique pas de calculs complémentaires comme en imagerie passive (cas de la stéréovision ou de la photogrammétrie passive).

Il existe plusieurs types de technologies de télédétection active pour mesurer une distance et imager une scène en trois dimensions tout en donnant accès à des propriétés physiques des objets (ex. radiométrie). Nous pouvons citer :

- Les systèmes SONAR (*Sound Navigation And Ranging*) qui utilisent des ondes sonores [4]. Ces instruments sont limités en distance et dépendent fortement de l'environnement. Ainsi, leur efficacité a tendance à diminuer en milieux urbains. Ils sont privilégiés pour des applications sous-marines par exemples.
- Les systèmes RADAR (*RAdio Detection And Ranging*) qui utilisent des micro-ondes ou des ondes radio [122]. Ces instruments permettent des propagations à

longue distance à travers une atmosphère dégradée. Cependant, les longueurs d'onde radio sont peu favorable à des images en haute résolution spatiale sans calculs complémentaires coûteux en consommation mémoire et temps de calcul.

- Les systèmes LiDAR (*Light Detection And Ranging*) qui utilisent des lasers (*Light Amplification by Stimulated Emission of Radiation*) émettant principalement dans le visible ou l'infrarouge (plus rarement dans l'ultraviolet) [99].

Ces trois technologies ont des similitudes. Elles sont toutes basées sur la détection d'une onde se propageant entre un émetteur, une cible d'intérêt et un détecteur. Dans le cas des systèmes SONAR, les ondes sonores ou acoustiques sont des ondes mécaniques nécessitant un support pour se propager de proche en proche. Dans les autres cas, nous pouvons considérer la propagation d'ondes électromagnétiques voire lumineuses dans le domaine optique dont les différentes longueurs d'ondes permettent d'exploiter les propriétés physiques d'une cible en couvrant un large spectre (voir Figure 1.1). Chacune de ces technologies a des avantages et des inconvénients. Il est intéressant d'identifier leurs limites et domaines d'utilisation pour les rendre complémentaire lorsqu'elles sont associées sur le même théâtre d'opération [125]. Ce manuscrit se concentre sur l'exploitation de systèmes LiDAR 3D.

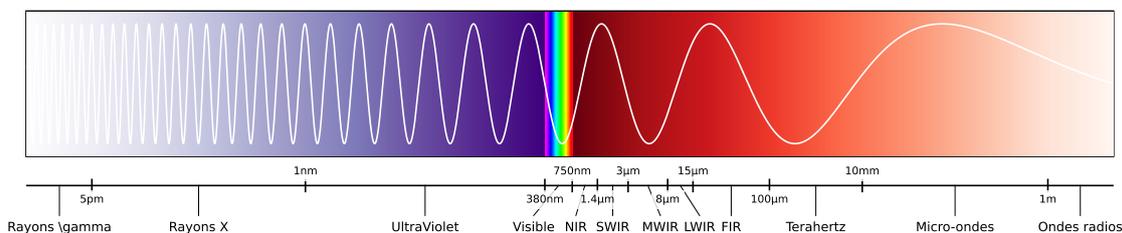


FIGURE 1.1. Spectre du rayonnement électromagnétique pour situer les différents domaine en fonction de leur longueur d'onde.

Fonctions de détection, reconnaissance et identification

Les applications de sécurité et de surveillance mettent souvent en œuvre des imageurs actifs pour permettre une observation par tout temps et par faible visibilité [14]. Des critères de Détection, Reconnaissance et Identification ou DRI sont alors considérés pour évaluer les performances de ces imageurs.

La fonction d'identification est définie comme étant la capacité à identifier un objet. Cette capacité peut être évaluée avec un critère dit d'identification. Le critère de Johnson [85] est un critère d'identification standard couramment utilisé dans l'industrie de la défense pour évaluer les performances d'un système d'imagerie en termes de probabilité d'interprétation par un opérateur. Il est défini sur les trois niveaux d'interprétation : détection, reconnaissance et identification (voir Figure 1.2). La détection fait référence à la capacité du système à détecter la présence d'une cible, la reconnaissance consiste à

classer un objet dans une catégorie générale et l'identification consiste à pouvoir spécifier l'appartenance à une classe plus précise.

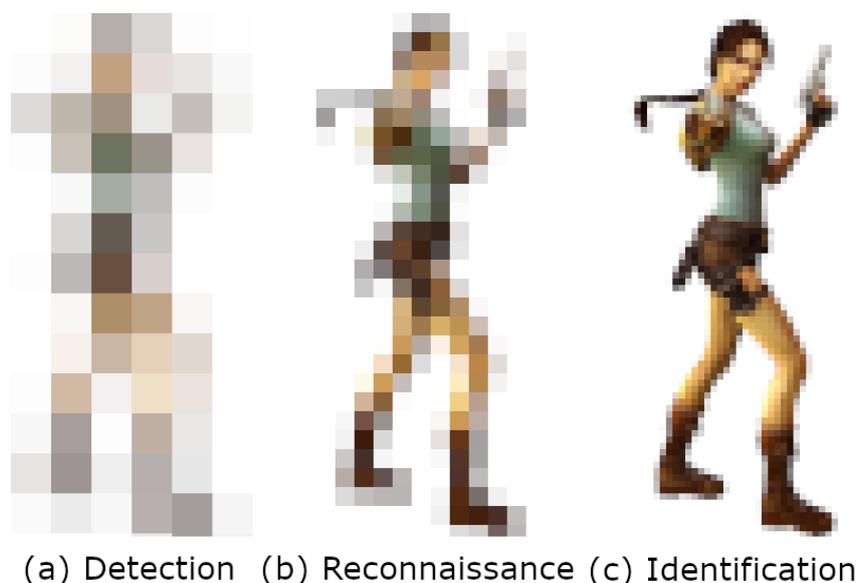


FIGURE 1.2. Exemple d'application du critère de Johnson ; (a) détection : « il y a quelque chose » ; (b) Reconnaissance : « c'est un humain » ; (c) identification : « c'est une archéologue célèbre ».

Le critère de Johnson est un outil mathématique relié à la perception humaine afin d'évaluer la performance d'un système de formation d'images pour les trois niveaux d'interprétation précédemment cités. Le critère de Johnson n'a pas d'équivalent unanime pour des informations tridimensionnelles comme les nuages de points issus d'un LiDAR 3D. Les nuages de points sont des ensembles de données où chaque élément a une coordonnée dans l'espace et éventuellement d'autres informations, comme l'intensité dans notre cas. Notons que quelques méthodes existent dans la littérature mais qu'elles ne font pas consensus dans la communauté et qu'aucune norme n'est adoptée et validée. [51]

Bien qu'il n'y ait pas de consensus sur la métrique à utiliser avec des données 3D pour quantifier la fonction de DRI, il reste important d'évaluer les performances des imageurs par tout temps, quelles que soient les conditions d'éclairage et d'environnement (ex. jour, nuit, éblouissement solaire). Les utilisateurs font état de contraintes lors du déploiement d'imageurs passifs en présence de telles conditions expérimentales. C'est particulièrement prégnant lorsque la visibilité est faible ou par forte variation de l'éclairage solaire. L'usage d'imageurs optroniques actifs leur est parfois préféré. Ils sont souvent déployés en complément pour augmenter l'opérabilité des systèmes. Les LiDAR 3D permettent de limiter l'impact des variations environnementales extérieures au système d'acquisition par l'emploi et la maîtrise de sa propre source d'illumination [82, 105, 147]. Dans cette thèse, l'impact de ces variations sur la qualité de la reconstruction ne sera pas évalué.

Systèmes LiDAR 3D

Les systèmes LiDAR 3D sont utilisés dans de nombreuses applications, pour créer des cartes topographiques à haute résolution de la surface de la Terre, ainsi que pour mesurer des distances précises ou des volumes (ex. exploitations minières et arpentages). Il est possible de retrouver ces instruments embarqués sur des plateformes mobiles telles que des véhicules autonomes ou des systèmes navigants dans un environnement inconnu initialement (cas de la robotique). Ils sont également déployés sur drones ou avions pour des applications d'aide à la navigation voire à de la navigation autonome, de collecte d'information en milieux difficilement accessibles (ex. après un tremblement de terre) ou en archéologie pour créer des modèles 3D de structures anciennes recouvertes par une végétation dense.

Il est possible de décomposer un système LiDAR en deux parties instrumentales essentielles pour réaliser la mesure. La première correspond au bloc « émission » généralement constituée par une source laser et la mise en forme du faisceau. La seconde est associée au bloc « réception » qui permet la collecte des photons sur un capteur ainsi que le prétraitement du signal utile. Les choix de la source laser et du capteur dépendent de l'application finale du système. Pour la partie émission, la grande variété de sources laser permet d'évaluer un grand nombre de paramètres instrumentaux comme la masse globale du système et sa compacité, la forme et la puissance crête de l'impulsion, le domaine spectral d'émission. . . Pour la partie détection, les familles de capteurs sont également nombreuses et dépendent de leur résolution, de leur sensibilité mais aussi de leur encombrement qu'il faut associer aux optiques de collection (ex. lentilles, télescopes). La multitude de paramètres rend impossible de tester l'ensemble des possibilités sur le terrain. L'ONERA a développé un logiciel de simulation des performances de LiDAR 3D. Nous présentons cet outil numérique dans la Section 1.4. Nous limitons notre étude en considérant un nombre de systèmes restreints et à un nombre de paramètres variables raisonnables. Une première restriction est de considérer dans cette thèse des systèmes embarquables à terme.

Notons qu'il existe deux catégories de systèmes LiDAR 3D : les systèmes continus FMCW ou AMCW (respectivement *Frequency-Modulated Continuous-Wave* ou *Amplitude-Modulated Continuous-Wave*) et les systèmes impulsionsnels qui mesurent le temps de vol des photons.

Les systèmes FMCW fonctionnent en émettant un faisceau laser continu à une fréquence modulée. Le faisceau laser est divisé en deux parties : le premier faisceau est dirigé vers la cible, tandis que l'autre est considéré comme étant la référence. Le faisceau réfléchi par la cible est ensuite détecté par l'instrument avant d'être comparé au signal de référence. La différence de fréquences entre les deux signaux permet de retrouver la distance à la cible [157].

Les systèmes AMCW fonctionnent en modulant l'amplitude du faisceau plutôt que sa fréquence. Ils émettent périodiquement des impulsions d'amplitude variable. Les impulsions réfléchies par la cible sont détectées puis analysées pour déterminer la distance en fonction de la modulation d'amplitude.

Bien que les systèmes LiDAR continus semblent être intéressants pour réaliser des mesures précises de distance, nous avons choisi de ne pas utiliser les LiDAR 3D FMCW ou AMCW car ces techniques sont limitées à des portées sub-kilométriques (voir Figure 1.3). Nous privilégions les systèmes impulsions dans la suite de nos travaux.

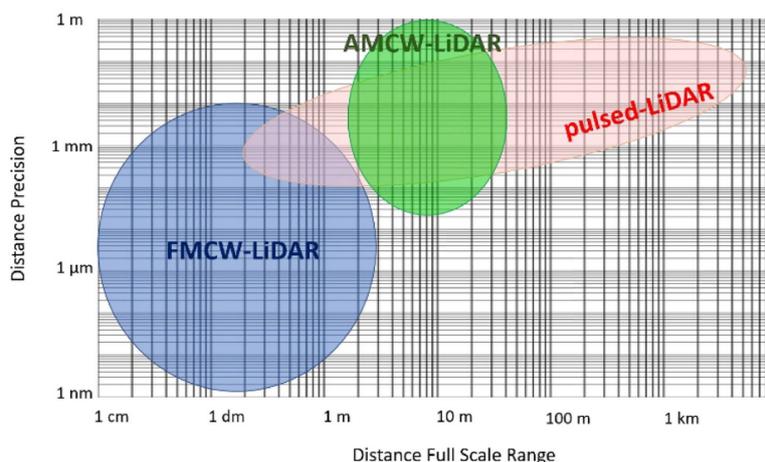


FIGURE 1.3. Résolution en profondeur de systèmes LiDAR-3D par rapport à la portée totale. Comparaison entre les systèmes impulsions, AMCW et FMCW pour des systèmes commercialement disponibles en 2021 [157].

À l'inverse des systèmes continus, les systèmes impulsions fonctionnent en émettant une impulsion laser puis en mesurant le temps de vol des photons, c'est-à-dire le temps mis par l'impulsion entre son émission, sa réflexion sur la cible et sa détection. Le temps de parcours est ensuite relié à la distance de la cible. Chaque maximum local de l'intensité reçue sur le capteur pour une même impulsion émise est associé à un écho temporel et à un point de l'espace dont les coordonnées spatiales (x,y,z) sont relatives à la position du LiDAR 3D. Une image 3D précise et un nuage de points 3D associés à la géométrie de la cible est finalement obtenu (voir Figure 1.4).

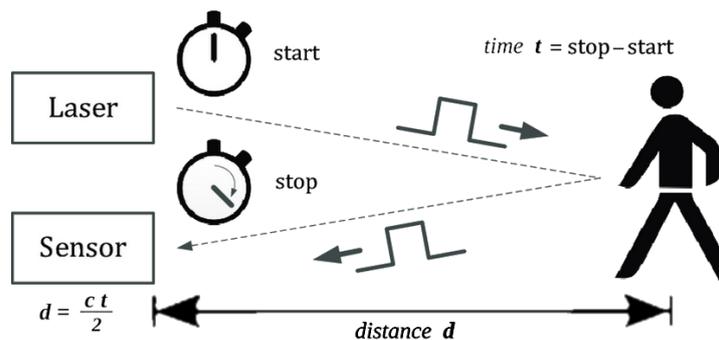


FIGURE 1.4. Une détection pour un pixel obtenu avec le temps de vol.

Imagerie laser 3D embarquée sur plateformes mobiles

Un « système embarqué » est un système qui est intégré dans un dispositif plus grand et mobile pour contrôler, surveiller ou assister le fonctionnement du dispositif tout en maintenant la mission qui lui est confiée. Les systèmes LiDAR sont souvent embarqués sur des véhicules roulants (ex. aide à la conduite [119], conduite autonome [63]), sur des avions (ex. cartographie tactique [127], archéologie [38]) et plus récemment sur des drones (ex. gestion de crises comme dans le projet européen INACHUS [125]). Certains systèmes LiDAR ont aussi été embarqués sur des satellites (ex. ICESat [134], SRTM [149]), bien que pour l’instant la résolution latérale de tels système reste très limitée (>30 m pour SRTM). L’une des principales contraintes de conception des systèmes embarqués est la nécessité de minimiser leur taille, leur masse et la puissance consommée du système tout en garantissant sa fonction de manière fiable et efficace [86, 128]. On parle alors de SWaP-C qui est l’acronyme de *Size, Weight, Power-Cost*. Maintenir un faible SWaP-C est primordial pour la conception et le déploiement de systèmes LiDAR-3D embarqués [27, 133] vis-à-vis de nos applications.

La **source laser** déployée dans le bloc « émission » influence le SWaP-C des systèmes embarqués. Le type de source et la puissance du laser sont des paramètres essentiels pour maintenir la fonction de DRI. Son choix est un consensus avec la sélection d’un détecteur plus ou moins sensible. Seuls les systèmes utilisant des sources laser impulsionnelles sont considérés dans notre étude car ils permettent de mesurer facilement le temps de vol des photons et d’accéder à une puissance crête plus importante que les lasers dits continus.

Les caractéristiques d’une impulsion laser se résument à trois paramètres : la longueur d’onde d’émission ou le domaine spectral d’émission, la forme spatiale de l’impulsion et sa forme temporelle (voir Figure 1.5). Chaque paramètre peut être associé à la façon dont l’énergie est répartie selon trois domaines :

- une distribution spectrale le plus souvent symétrique et centrée autour d’une longueur d’onde avec potentiellement des modes supplémentaires proches de cette longueur d’onde. Ces derniers sont négligés (voir Figure 1.5a en échelle logarithmique).
- une distribution spatiale le plus souvent gaussienne centrée par rapport au centre de l’impulsion. D’autres distributions sont possibles comme le profil « *top-hat* » (voir Figure 1.5b) qui permet un éclairage carré de la cible. Modifier la distribution spatiale d’une impulsion laser peut se révéler complexe, d’autant que cette forme évolue lors de sa propagation dans l’atmosphère.
- une distribution temporelle, le plus souvent gaussienne (voir la Section 1.1.2) résumée par sa puissance crête et sa largeur à mi-hauteur. La génération réelle du profil de l’impulsion peut s’écarter d’une représentation idéale gaussienne comme illustré sur la Figure 1.5c.

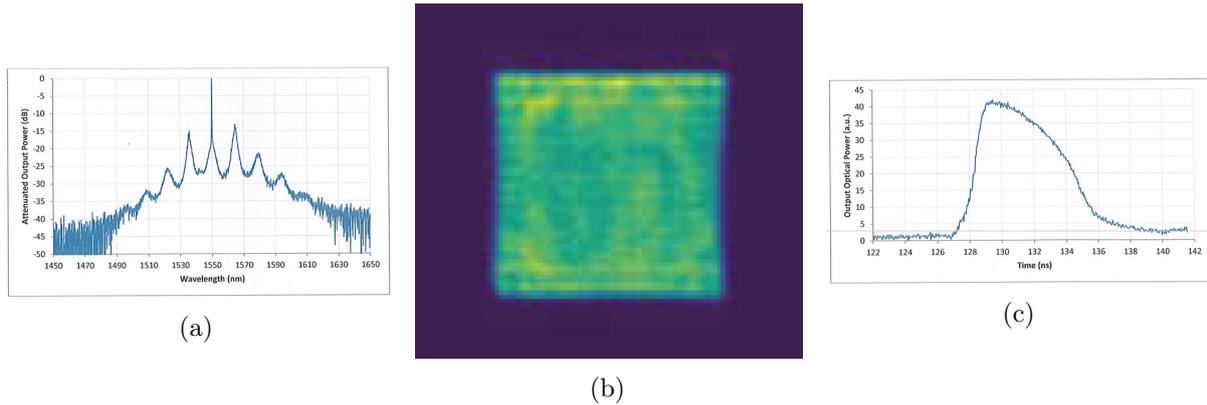


FIGURE 1.5. Exemples de distributions pour une source laser : (a) distribution spectrale centrée à $1,55\ \mu\text{m}$, distribution spatiale dite « *top-hat* » [112] et (c) distribution temporelle réelle mesurée.

L’observation à longue portée (au-delà du kilomètre) constitue l’un des objectifs de nos travaux tout en maintenant un faible SWaP-C. Cette condition influence le choix des trois paramètres décrits précédemment ainsi que de l’énergie totale par impulsion. Ils sont supposés variables indépendamment les uns des autres lors de l’analyse de performances. En pratique, ses trois paramètres sont dépendants de la technologie laser utilisée.

La puissance crête est maximisée afin d’obtenir un rapport signal sur bruit ou SNR (*Signal-Noise Ratio*) le plus important possible tout en considérant une énergie totale par impulsion fixe (limitée par le SWaP-C). Ainsi, l’utilisation d’une impulsion temporelle la plus courte possible est favorisée. Un gradient important sur les premiers pas temporel est privilégié. Cela permet de mieux différencier les échos de deux objets projetés donc d’augmenter la résolution temporelle du système. Nous reviendrons sur cette spécificité dans la suite de nos travaux.

Le choix de la longueur d’onde principale (et de sa distribution) influence la réponse globale du système. En effet, les propriétés de diffusion et d’absorption des matériaux de la cible ou des particules en présence dans l’atmosphère sont dépendantes de la longueur d’onde d’éclairement. En particulier le domaine spectral SWIR (*Short-Wave InfraRed*) compris entre 1 et $2,5\ \mu\text{m}$ permet de garantir une certaine sécurité oculaire (voir Figure 1.6) du système tout en limitant l’impact de la contribution solaire (bruit d’éclairement) comme l’illustre la Figure 1.7. De plus, ce domaine limite aussi l’impact des mauvaises conditions météorologiques (bruit de l’environnement lié aux multidiffusions par les aérosols et hydrométéores, bien que cela ne soit pas évalué dans la thèse [15]). Nous privilégions le choix de LiDAR 3D fonctionnant dans cette gamme spectrale car il existe des capteurs associés et optimisés autour de $1,5\ \mu\text{m}$. Ces capteurs photodiodes à avalanche dits « APD » (*Avalanche PhotoDiode*) ou « InP/InGaAs GmAPD » (*Geiger Mode APD*) sont principalement des matrices de pixels carrées, rectangulaires ou ronds.

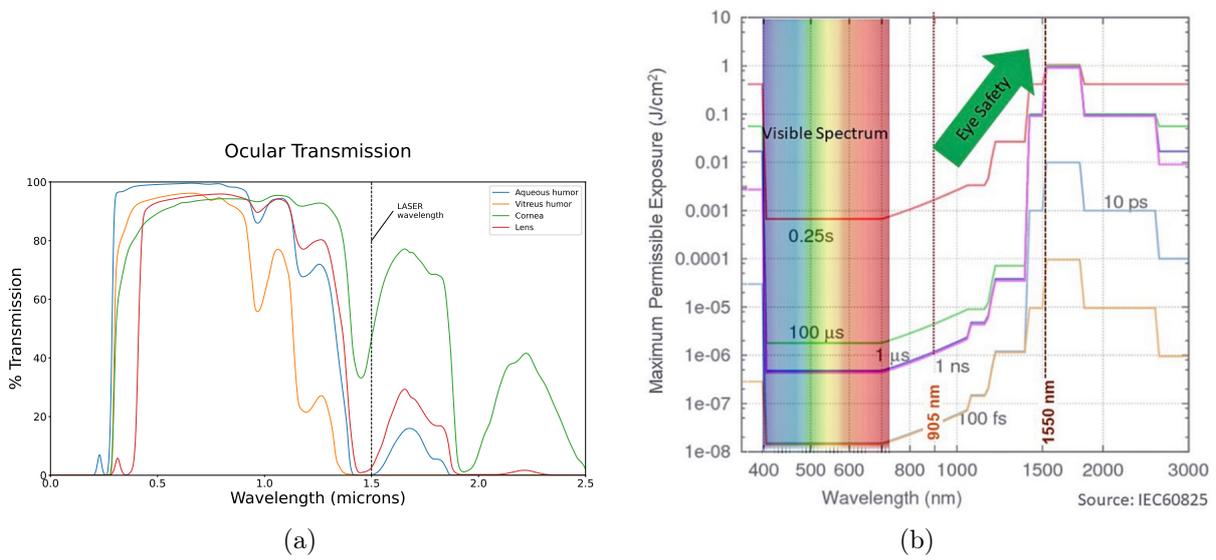


FIGURE 1.6. (a) Spectres de transmission des composants de l'œil [167] corrélés positivement avec le risque de dégâts (qui peut varier en fonction de la densité d'énergie, la longueur d'onde et la durée d'exposition ; (b) Exposition maximum permise à 1,55 μm pour plusieurs durées d'impulsions [2] d'après la norme IEC60825.

L'énergie reçue doit uniformément être répartie sur le plan image pour favoriser une augmentation de la résolution spatiale latérale d'un imageur laser 3D, comme nous le verrons dans la Section 1.1.2. Il est préférable de mettre en œuvre un éclairage laser ayant un profil identique à celui du plan focal du capteur [151]. Nous restreindrons par la suite nos travaux à l'utilisation de sources laser dont le faisceau est mis en forme de type « *top-hat* » carré (section carrée où l'énergie est uniformément répartie).

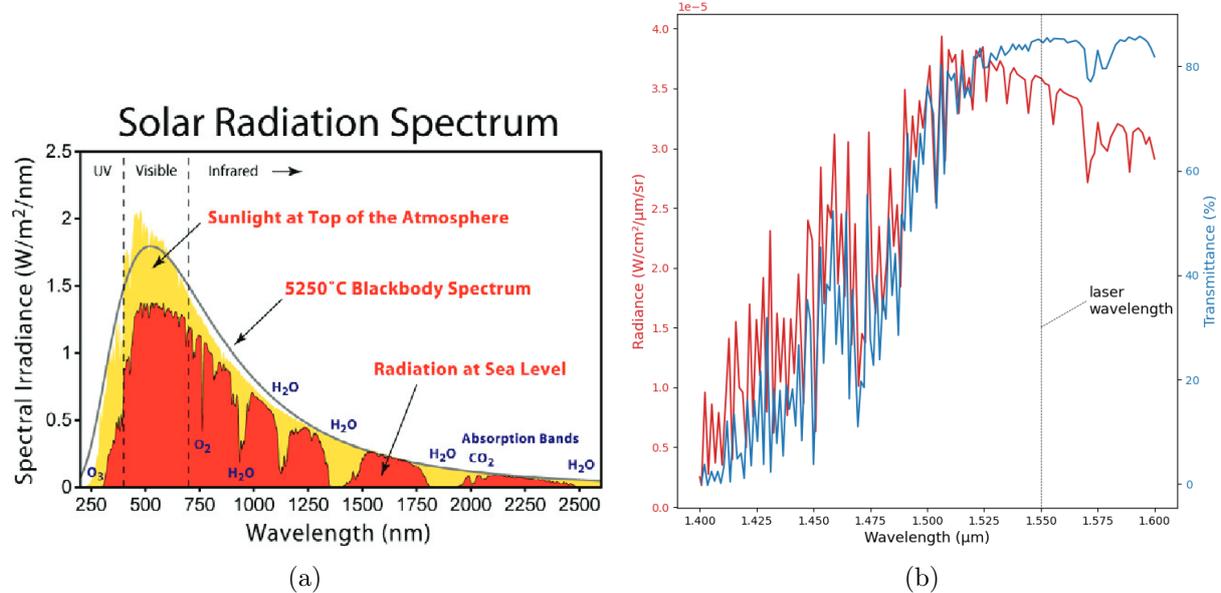


FIGURE 1.7. (a) Éclairage après avoir traversé l'atmosphère qui est corrélé positivement avec le bruit reçu par le système [62] ; (b) Zoom sur la zone 1,4 μm à 1,6 μm avec l'éclairage et la transmission dans l'atmosphère [62].

Le choix du **capteur** et des optiques de collection dans le bloc « réception » est à réaliser en parallèle de la sélection de la source laser. Les plans focaux 3D basés sur des éléments APD sont les meilleurs candidats pour des missions de DRI à longue portée. La sensibilité et la résolution temporelle [81] de ces capteurs sont tous les deux compatibles avec des systèmes LiDAR 3D à faible SWaP-C [27, 133]. Les imageurs à plans focaux 3D sont des « caméras » composées par une matrice de pixels située au plan focal de l’optique d’imagerie. Ils sont parfois opposés aux caméras dites « *single-pixel* » que nous décrivons dans la Section 1.2. On parle de trois dimensions car on ajoute à la résolution latérale 2D du plan focal, la profondeur de détection ou télémétrie (dimension temporelle qui est décrite dans la Section 1.1.1).

Le fonctionnement des détecteurs APD peut être résumé de la manière suivante [81]. Un photon incident sur le détecteur est absorbé puis converti en photoélectron dans les couches supérieures du détecteur (voir Figure 1.8). Le photoélectron est ensuite accéléré et son énergie déclenche un phénomène dit « d’avalanche » dans le semi-conducteur. Il s’agit d’un effet multiplicateur du courant électrique à l’intérieur de matériaux qui sont de bons isolants jusqu’au déclenchement du phénomène. Le nombre de photoélectrons est multiplié exponentiellement et crée un courant électrique qui va déclencher une détection.

Pour maximiser la conversion d’un photon incident en photoélectron et le déclenchement d’un effet d’avalanche, les premières couches du capteur sont portées à une tension électrique proche du seuil à laquelle ce phénomène se produit [81]. Lorsque la tension appliquée est en-dessous du seuil, le composant est dit en régime linéaire (voir Figure 1.9). Les APD en régime linéaire permettent d’obtenir un gain de 10^2 au plus. L’électronique de détection associée à cette catégorie de capteurs permet d’évaluer le flux de lumière à chaque pas de temps en continu. Il est possible d’appliquer une tension légèrement au-dessus du seuil d’avalanche : on parle alors de **mode Geiger**. Dans ce cas la détection n’est plus linéaire. Cette technique particulière, induit deux effets collatéraux au détecteur. Tout d’abord un bruit supplémentaire qui est appelé DCR pour *Dark Count Rate* et représente le taux de déclenchements du détecteur sans photon incident. Il est lié à l’énergie des couches supérieures qui est trop proche du seuil [82]. Le temps nécessaire pour réarmer le détecteur en vue d’une nouvelle acquisition limite l’acquisition d’une seule détection et ce, sans pouvoir discerner le nombre de photons ayant déclenché l’effet d’avalanche [81]. On perd ainsi la notion de linéarité et d’étalonnage de la conversion photon-photoélectron-courant : la radiométrie est nativement inaccessible. Ces détecteurs sont optimisés autour d’une longueur d’onde donnée. Traditionnellement, ils sont nommés différemment en fonction de la valeur de cette longueur d’onde : UV-APD dans l’ultraviolet [139], SPAD (*Single-Photon Avalanche Diodes*) dans le visible [46] et GmAPD (*Geiger-mode Avalanche PhotoDiodes*) dans l’infrarouge [81]. Pour résumer, les APD sont des détecteurs très sensibles permettant de détecter statistiquement jusqu’à un seul photon. Cependant, ils ne peuvent enregistrer que la première détection lorsque plusieurs événements interviennent dans des temps très

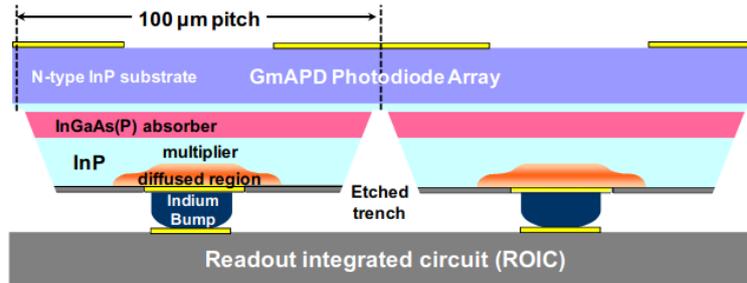


FIGURE 1.8. Coupe transversale schématique d'un détecteur GmAPD, d'après [81].

courts. A l'inverse, les systèmes linéaires dits « classiques » sont beaucoup moins sensibles mais donnent une information radiométriquement étalonnée sur le signal.

L'image obtenue *in fine* correspond à un nuage de points en trois dimensions. Le proces-

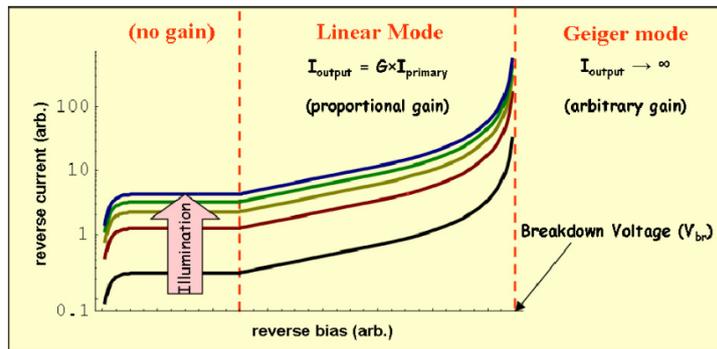


FIGURE 1.9. Gain en fonction de la tension dans les APD en mode Geiger ou linéaire, extrait des travaux de Williams et *al.* [163].

sus d'obtention des trois dimensions spatiales diffère en fonction du cas considéré : nous distinguons le traitement des deux dimensions spatiales latérales (x, y) de la profondeur (z) obtenue par télémétrie temporelle. Le temps de vol aller-retour des photons correspond à un nombre de pas de temps et est corrélé à la distance parcourue. En détection, la résolution temporelle de l'imageur est liée à la taille d'un pas de temps élémentaire que l'on appelle *bin* et pas de temps en français. Sa valeur est définie par la fréquence de l'horloge interne de la caméra synchronisée avec l'émission de l'impulsion laser. La résolution latérale projetée à une certaine distance, est quant à elle liée au nombre de pixels du plan focal 3D. Cependant, cette résolution dépend aussi du champ de vue ou FOV (*Field Of View*). En effet, le FOV est une mesure de l'étendue du champ observable par une caméra lors de la capture d'une seule image. Il est le plus souvent donné sous la forme d'une mesure angulaire exprimée en radian. En fonction de la distance d'observation, il est souvent converti en une distance exprimée en mètres. Ainsi, lorsque le FOV est fixé, augmenter le nombre de pixels de la caméra revient à augmenter la résolution latérale et inversement. Même s'il est nécessaire de traiter différemment les résolutions latérales et temporelles, leurs interprétations restent identiques : elles expriment la capacité du système à discerner deux objets proches dans la scène. Elles sont d'ailleurs souvent exprimées en

termes de distance minimale entre deux objets qui peuvent être résolus indépendamment l'un de l'autre.

Actuellement, les plans focaux à comptage de photons sont tous limités à quelques centaines de pixels de grande taille individuelle par rapport au standard disponible en imagerie passive. Dans l'état-de-l'art, les plans focaux 3D à comptage de photons sont limités à 256×256 pixels dans le visible et l'ultraviolet et à 128×128 pixels dans l'infrarouge [9]. Ainsi, la résolution latérale ou le FOV sont fortement contraints en imagerie laser 3D. Comme la fonction de DRI requiert une bonne résolution télémétrique et latérale des imageurs, il est nécessaire d'identifier des outils numériques ou des astuces instrumentales qui permettent d'augmenter virtuellement la résolution latérale et maintenir une fonction de DRI longue portée avec un LiDAR 3D.

Dans notre étude, nous faisons le choix d'un système LiDAR 3D entièrement paramétrable d'un point de vue capteur et source laser (distributions spatiale, temporelle et spectrale du laser) tout en restant compatible de systèmes embarquables. Nous privilégions un système LiDAR 3D composé d'un faisceau « *top-hat* » carré, une émission à $1,5 \mu\text{m}$ et une impulsion piquée comme celle représentée sur la Figure 1.5 Un plan focal à comptage de photons de type InP/InGaAs GmAPD est utilisé. Au-delà de l'instrument, nous devons également considérer l'environnement de travail. Une hypothèse forte est faite puisque nous négligeons les effets de la turbulence sur la propagation du faisceau laser et la formation des images tridimensionnelles. La turbulence atmosphérique est due à un mouvement chaotique des particules dans l'air et à la non-homogénéité de l'indice de réfraction sur l'ensemble du parcours optique. Les masses d'air en mouvement vont induire des variations locales de l'indice de réfraction de l'air, introduire des surfaces d'onde locales non planes et perturber la propagation des ondes électromagnétiques de la lumière en modifiant l'orientation locale du vecteur d'onde. La caractérisation et la mesure des phénomènes de turbulence ne sont pas triviales. Il en résulte un « flou » et une certaine « distorsion » des images au niveau du plan focal si elles ne sont pas modifiées par post-traitement.

Concepts de l'imagerie LiDAR 3D à comptage de photons

Les caméras à comptage de photons sont de plus en plus utilisées dans les systèmes de détection de type LiDAR 3D, notamment dans les systèmes à faible SWaP-C. Nous nous concentrons dans notre étude sur les systèmes utilisant des capteurs GmAPD [66] car leurs performances intrinsèques sont intéressantes pour nos applications. Comme nous l'avons déjà évoqué, les capteurs GmAPD sont limités à quelques pixels de grande taille (jusqu'à 128×128 pixels) [9, 21], ce qui peut entraîner une résolution insuffisante pour des fonctions de DRI à longue portée. L'augmentation du nombre de pixels de la matrice d'acquisition impose ses propres contraintes telles que la taille du plan image et les capacités de l'électronique proche capteur à collecter et traiter le signal de tous les pixels d'une même matrice.

À ce jour, il n'existe pas de méthode dans l'état-de-l'art permettant d'augmenter la résolution spatiale d'un LiDAR 3D en prenant en compte les contraintes suivantes :

- Acquisitions à longue portée (supérieure au kilomètre) en haute résolution ;
- Utilisation de matrices de capteur de type GmAPD ;
- Court temps d'acquisition, de l'ordre de la milliseconde pour la navigation ou de la seconde dans le cadre d'acquisition avec post-traitement ;
- Compatible de systèmes embarqués avec un faible SWaP-C ;
- Faible niveau de signal (statistiquement inférieur à un photon en moyenne).

Dans nos travaux, nous proposons une méthode originale d'acquisition comprimée inspirée de l'application du *Compressive Sensing* (CS) aux caméras *single-pixel*. Cette approche répond à la plupart des contraintes citées précédemment. Par exemple, Gordon et al. [69] ont réalisé des images à longue portée (9 km) en utilisant un plan focal muni de 32×32 pixels de type GmAPD avec une résolution latérale constante ne permettant pas d'améliorer la fonction de DRI lorsque la distance d'observation croît. D'autres approches augmentent la résolution latérale en présence de contraintes différentes en imagerie LiDAR 3D : en 2016, Shin et al. [138] ont démontré la reconstruction efficace d'images à l'aide de SPAD en balayant des scènes statiques avec un système motorisé deux axes.

Des acquisitions ont été réalisées avec une portée de 10 km [113] en 2017 et de 45 km en 2020 [90,92] en utilisant un système à balayage. En 2019, Tachella et al. [144] ont démontré une application en temps réel à courte portée (de l'ordre de 300 m) avec une résolution latérale améliorée par super-résolution. Dans ce cas, le principe de la méthode utilisé est de répartir l'information localement *a posteriori*, une fois l'image 3D acquise chaque est divisée en $3 \text{ pixels} \times 3 \text{ pixels}$. Ensuite, l'algorithme qu'ils ont développé est utilisé sur ces données augmentées pour affiner la reconstruction (débruitage, estimation de la profondeur et de l'intensité). En utilisant des systèmes à balayage, la période pour une trame est compatible avec des applications statiques mais inadaptée pour des objets en mouvement dans un contexte opérationnel. Les matrices de plans focaux offrent des fréquences d'acquisition plus élevées que les systèmes à balayage [157]. En outre, le *Compressive Sensing* a permis d'obtenir des résultats équivalents en utilisant seulement 10% du nombre d'acquisitions requises par les systèmes à balayage classique [55]. Les méthodes de super-résolution sont inadaptées pour résoudre des objets sub-pixels. Sun et al. [143] ont montré que la super-résolution tend à être plus floue que la reconstruction par *Compressive Sensing*. Les méthodes basées sur le *Compressive Sensing* sont moins vulnérables à ces problématiques et c'est essentiellement pour cette raison que notre approche intègre cette technique [29,30,32].

Depuis 2008, le *Compressive Sensing* est utilisé par la communauté LiDAR 3D pour augmenter la résolution latérale de caméras *single-pixel* [55, 58, 59, 67, 74, 77, 78, 130, 142]. Dans la plupart des cas, ces travaux utilisent les algorithmes d'*Orthogonal Matching Pursuit* (OMP) [148] ou de *Total Variation* (TV) [89]. En 2008, Duarte et al. [55]

introduisent le concept de caméras single-pixel en utilisant le Compressive Sensing pour reconstruire une image à partir d'un unique pixel en modulant le signal. Dans ces travaux ils démontrent aussi la faisabilité de la méthode en construisant un système utilisant une caméra single-pixel. En 2013, Howland et *al.* [78] ont démontré un système LiDAR à comptage de photon utilisant une caméra single-pixel permettant de reconstruire une carte de profondeur. En 2017, Edgar et *al.* [58] ont réalisé une acquisition en temps réel en utilisant une approche de Compressive Sensing avec une portée métrique et un capteur mono-pixel. Pour les caméras single-pixel, en plus des différents systèmes, les bases de mesure utilisées dans le Compressive Sensing pour le single-pixel a fait l'objet d'un certain nombre d'étude afin de diminuer le nombre d'acquisition nécessaire en particulier sur les bases de mesures orthogonales, tel que : les bases d'Hadarmard [142], d'Ondelettes [131] ou de Fourier [166]. Notre apport par rapport à l'état de l'art actuel porte sur l'utilisation du *Compressive Sensing* pour les caméras infrarouges et sur le traitement du signal. Les traitements en amont du *Compressive Sensing* pour des capteurs de type GmAPD ainsi que sa généralisation à des plan-focaux plutôt que des caméras *single-pixel* sont compatibles d'une implémentation à très faible latence pour des applications embarquées. La théorie et les outils de *Compressive Sensing* seront détaillés dans la Section 1.4.

1.1.2 L'équation lidar

Le début de ce chapitre présente comment construire une image en trois dimensions avec un système LiDAR 3D embarqué en évaluant l'intensité voire la puissance reçue au cours du temps par rapport à celle émise (cas des détecteurs linéaires) ou à évaluer le temps de vol des photons à l'aide de détecteurs à comptage de photons. Il est possible de modéliser théoriquement le courant généré au niveau des capteurs en tenant compte des phénomènes physiques mis en jeu lors de la propagation des photons dans l'atmosphère, leur interaction avec une cible et leur détection *via* un modèle physique du capteur. Dans notre étude, nous supposons les hypothèses suivantes pour obtenir l'équation lidar adaptée :

- La longueur d'onde laser est unique. Le faisceau laser est émis sur une bande spectrale très étroite de l'ordre de quelques nanomètres. Le bloc « détection » inclut un filtre spectral étroit qui limite efficacement le spectre reçu autour de cette même longueur d'onde.
- La forme de l'impulsion est connue avec une durée courte de l'ordre de la nanoseconde pour privilégier une forte puissance crête.
- Le profil du faisceau laser est connu. On choisit une distribution spatiale uniforme et dont le profil est carré pour correspondre à la forme des capteurs de type GmAPD. Notons que ce choix est motivé par la simplification de l'utilisation de modulations spatiales en *Compressive Sensing* et la minimisation du nombre d'acquisitions à réaliser.

- Les premières centaines de mètre ne contiennent pas d'objet, ce sont les acquisitions longue portée (>10 km) qui sont considérées. Cette hypothèse permet la simplification du terme de transmission atmosphérique de l'Équation (1.1) et n'est pas strictement nécessaire. Afin de prendre en compte ces premières centaines de mètre il suffirait de garder la dépendance à la distance de ce paramètre, tout au long des développements.
- Les turbulences atmosphériques sont négligeables et le milieu dans lequel le faisceau laser se propage est homogène. Ces conditions sont vérifiées lorsqu'on considère des scénarii où l'imageur n'est pas proche du sol. L'ONERA a déjà démontré l'influence des échanges thermiques entre le sol et la basse atmosphère pour des applications en imagerie active [140, 150]. Ici, nous considérons que le système est embarqué sur une plateforme à plusieurs mètres du sol. En négligeant cette contribution, nous souhaitons consacrer principalement notre étude à la compréhension des algorithmes de formation d'images 3D hautement résolues par une approche de *Compressive Sensing*.
- Les effets de Speckle de cible ne sont pas considérés dans notre étude. L'ONERA a mis en évidence lors de précédents travaux [126] que cette contribution est négligeable pour l'observation de la plupart des matériaux présents sur une scène urbaine ou naturelle. Nous distinguerons uniquement les matériaux dits « Lambertiens » des matériaux spéculaires.

L'équation générale dite « équation lidar » qui régit la puissance collectée $P(\lambda, t)$ sur un capteur [100] est définie par :

$$P(\lambda, t) = A_0 \xi(\lambda) \int_0^{\frac{ct}{2}} \frac{1}{R^2} \beta(\lambda_L, \lambda, R) C_a(\lambda, R) \int_{\text{supp}(A)} \mathcal{O}(R, r) I(R, r) dA(R, r) dR \quad (1.1)$$

Avec :

- A_0 Aire de la lentille (m^2)
- λ_L Longueur d'onde émise par le laser (m)
- $\xi(\lambda)$ Transmission spectrale du système (sans dimension)
- c Célérité de la lumière ($\text{m} \cdot \text{s}^{-1}$)
- R Distance entre un élément de la scène et l'instrument optique (m)
- $\beta(\lambda_L, \lambda, R)$ Coefficient de rétrodiffusion du milieu ($\text{m}^{-1} \cdot \text{sr}^{-1}$)
- $C_a(\lambda, R)$ Facteur d'atténuation à la transmission atmosphérique (sans dimension)
- $\text{supp}(A)$ Support de la section du laser à la distance R (m^2)

- r Position dans $\text{supp}(A)$ dans un plan parallèle au plan de la caméra à la distance R
- $\mathcal{O}(R, r)$ Facteur lié à la géométrie et à la radiométrie des objets de la scène (sans dimension)
- $I(R, r)$ Éclairement ($\text{W} \cdot \text{m}^{-2}$)
- $dA(R, r)$ Mesure liée à la distribution latérale et temporelle du laser (m^{-2})

Le support d'une fonction est défini au sens mathématique comme l'ensemble des points où la fonction ne s'annule pas. Ainsi le support de $A(R, \cdot)$ est défini comme suit :

$$\text{supp}(A(R, \cdot)) = \{r \in \mathbb{R}^2, \text{ tel que } A(R, r) \neq 0\} \quad (1.2)$$

Dans le cas des imageurs LiDAR 3D, on s'intéresse au problème inverse qui consiste à retrouver le facteur lié à la géométrie et à la radiométrie des objets $\mathcal{O}(R, r)$ à partir de la puissance mesurée $P(\lambda, t)$. Le facteur $\mathcal{O}(R, r)$ représente la probabilité qu'une radiation provenant du point r à la distance R atteint le détecteur en ne prenant en compte que les propriétés géométrique et radiométrique de la scène (champ de regard, géométrie et radiométrie des objets). Expérimentalement, nous faisons l'hypothèse que le filtre spectral placé en réception est suffisamment étroit pour qu'une seule longueur d'onde correspondante à l'émission soit détectée. Cela limite également l'influence d'un bruit qui serait généré par l'environnement ou une source de lumière extérieure. La dispersion spectrale est négligeable dans l'équation lidar de telle sorte que la longueur d'onde d'acquisition λ peut être remplacée par celle du laser λ_L . Comme on suppose que P ne dépend plus que du temps (λ_L étant une constante), nous modifions les notations pour plus de clarté. De plus, la transmission atmosphérique est approchée par un coefficient C_a , en effet comme $C_a = e^{-Ra}$ [33] (avec a un coefficient d'extinction) après quelques centaines de mètres dans un milieu homogène C_a est quasiment constant. On obtient ainsi l'équation suivante :

$$P(t) = A_0 C_a \xi_{\lambda_L} \int_0^{\frac{ct}{2}} \frac{1}{R^2} \beta_{\lambda_L}(R) \int_{\text{supp}(A)} \mathcal{O}(R, r) I(R, r) dA(R, r) dR \quad (1.3)$$

En supposant que le milieu est homogène, on peut approcher $\beta_{\lambda_L}(R)$ par 1.

$$P(t) = A_0 C_a \xi_{\lambda_L} \int_0^{\frac{ct}{2}} \frac{1}{R^2} \int_{\text{supp}(A)} \mathcal{O}(R, r) I(R, r) dA(R, r) dR \quad (1.4)$$

En supposant que le profil du faisceau laser a une distribution spatiale uniforme et de forme carrée, l'éclairement ne dépend que du temps.

$$P(t) = A_0 C_a \xi_{\lambda_l} \int_0^{\frac{ct}{2}} \frac{I(R)}{R^2} \int_{\text{supp}(A)} \mathcal{O}(R, r) dA(R, r) dR \quad (1.5)$$

On note $\Omega(R)$ la surface représentant la tâche carrée du laser à une distance R . Cette surface est équivalente à la surface projetée du plan focal à la même distance. Ainsi, on note $\Omega_{i,j}$ les sous-éléments surfaciques uniformes de $\Omega(R)$ tels que $\Omega_{i,j}$, $\{i, j\} \in [0, N_p] \times [0, N_p]$ correspondent à la partie du signal reçu par le pixel i, j (N_p^2 étant le nombre de pixels de la caméra). En utilisant la relation de Chasles sur l'équation précédente, on obtient :

$$P(t) = A_0 C_a \xi_{\lambda_l} \int_0^{\frac{ct}{2}} \frac{I(R)}{R^2} \sum_{i,j=0}^{N_p} \int_{\Omega_{i,j}(R)} \mathcal{O}(R, \omega) d\omega dR \quad (1.6)$$

On explicite maintenant l'éclairement $I(R) = \frac{P_E(t - \frac{2R}{c})}{A_L(R)}$ d'après [100], où $P_E(t)$ correspond à l'énergie émise au temps t et $A_L(R)$ correspond à l'aire de $\Omega(R)$. Dans le cas où la distribution temporelle de l'impulsion laser n'est pas disponible, on modélise la puissance émise par la source laser à un instant donné en suivant les travaux de Carlsson [33].

$$P_E(t) = \left(\frac{3.5}{l_{fhw}} t \right)^2 e^{-\frac{3.5}{l_{fhw}} t} \quad (1.7)$$

où $t \geq 0$ et l_{fhw} est la largeur à mi-hauteur du LASER.

On a donc :

$$P(t) = A_0 C_a \xi_{\lambda_l} \int_0^{\frac{ct}{2}} \frac{P_E(t - \frac{2R}{c})}{A_L(R) R^2} \sum_{i,j=0}^{N_p} O_{i,j}(R) dR \quad (1.8)$$

On pose le changement de variable suivant, $T = \frac{2R}{c}$:

$$P(t) = \frac{2A_0 C_a \xi_{\lambda_l}}{c} \int_0^t P_E(t - T) \sum_{i,j=0}^{N_p} \frac{O_{i,j}(\frac{cT}{2})}{A_L(\frac{cT}{2}) T^2} dT \quad (1.9)$$

De plus, la fonction sous le signe de l'intégrale est mesurable car les fonctions sous le signe de l'intégrale sont toutes continues ou continues par morceaux $\forall t > 0$ (en supposant que $O_{i,j}$ est au moins continu par morceaux, condition généralement vraie dans un cadre opérationnel). On peut ainsi appliquer le théorème de Fubini-Tonelli et intervertir les signes somme et intégrale tel que :

$$P(t) = \sum_{i,j=0}^{N_p} \frac{2A_0 C_a \xi_{\lambda_l}}{c} \int_0^t P_E(t - T) \frac{O_{i,j}(\frac{cT}{2})}{A_L(\frac{cT}{2}) T^2} dT \quad (1.10)$$

On pose alors, $P_{i,j}(t)$ la puissance reçue au temps t et correspondante à un pixel (i, j) de la caméra :

$$P_{i,j}(t) = \frac{2A_0 C_a \xi_{\lambda_l}}{c} \int_0^t P_E(t-T) \frac{O_{i,j}(\frac{cT}{2})}{A_L(\frac{cT}{2})T^2} dT \quad (1.11)$$

Ce que l'on peut réécrire comme :

$$P_{i,j}(t) = \frac{2A_0 C_a \xi_{\lambda_l}}{c} \times (P_E * f)(t) \quad (1.12)$$

avec $f(t) = \frac{O_{i,j}(\frac{ct}{2})}{A_L(\frac{ct}{2})t^2}$

Pour résumer, la puissance équivalente et reçue en chaque pixel du plan focal est considérée comme étant indépendante de la puissance reçue par les autres pixels du détecteur matriciel. La puissance en chaque pixel est donc calculée à partir de l'équation lidar précédente pour chaque pixel du plan focal sans tenir compte d'une quelconque influence des pixels voisins. En mesurant $P_{i,j}(t)$, on peut retrouver des informations sur la scène aux constantes et à une convolution près. On s'intéressera dans un premier temps (voir Chapitre 2) au cas où $X \approx O^A$, c'est-à-dire lorsque l'impulsion est temporellement plus courte que δ_t . Dans un second temps, nous aborderons la déconvolution du signal reçu (voir Chapitre 3), c'est-à-dire lorsque l'impulsion est temporellement plus longue que δ_t .

1.2 La caméra single-pixel

1.2.1 Principe de fonctionnement

À partir de 2008, le développement des caméras dites *single-pixel* a suivi celui du *Compressive Sensing* [55]. Afin d'imager une scène avec un capteur monopixel, on intègre au système optronique un élément permettant de moduler spatialement le signal. Nous décrirons dans cette partie les différentes possibilités de génération de « schémas de modulation » c'est-à-dire comment générer une modulation spatiale à l'aide de motifs définis et permettant de reconstruire numériquement une image en trois dimensions. La Figure 1.10 présente les éléments essentiels qui constituent l'instrument LiDAR 3D avec modulation spatiale du signal détecté et quelques motifs de modulation. Les motifs utilisés pour la modulation sont généralement obtenus à partir d'une transformée discrète comme la transformée de Fourier discrète ou la transformée en cosinus discrète, où une ligne de la matrice correspond à un motif. Dans un premier temps, un signal est acquis pour chaque motif du schéma de modulation. Dans un second temps, il est alors nécessaire d'inverser numériquement le processus. L'inversion peut être simple, dans le cas des transformées orthogonales il suffit d'appliquer numériquement cette même transformée.

L'intérêt du *Compressive Sensing* dans ce cadre est de réduire le nombre de motifs

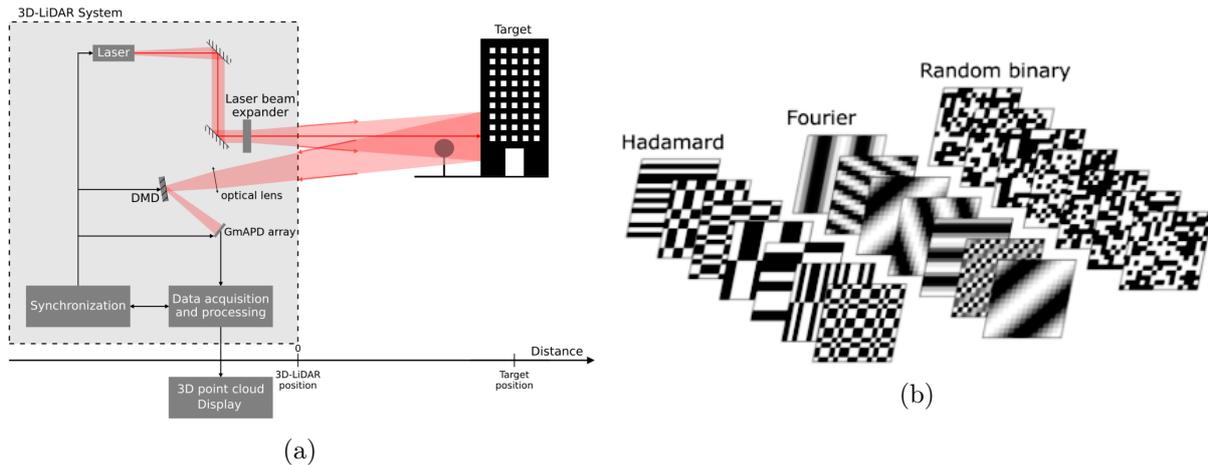


FIGURE 1.10. (a) Schéma simplifié d'un système LiDAR 3D incluant une matrice de micro-miroirs (DMD) pour moduler spatialement le signal reçu, (b) Exemples de motifs de modulation spatiale.

nécessaires pour reconstruire une image à partir des N acquisitions. Notons que cette reconstruction est considérée comme étant exacte lorsqu'il n'y a pas de bruit associé à la mesure. La motivation majeure des travaux actuels qui mettent en œuvre des approches de *Compressive Sensing* et des caméras monocapteur réside dans l'utilisation d'éléments optroniques à prix modéré et avec des propriétés potentiellement indisponibles pour des caméras composées de nombreux pixels. La résolution latérale équivalente reste conservée. En revanche, le principe d'acquisition est modifié sur deux points : il faut ajouter un élément de modulation spatiale dans le montage optique et utiliser un algorithme de reconstruction spécifique.

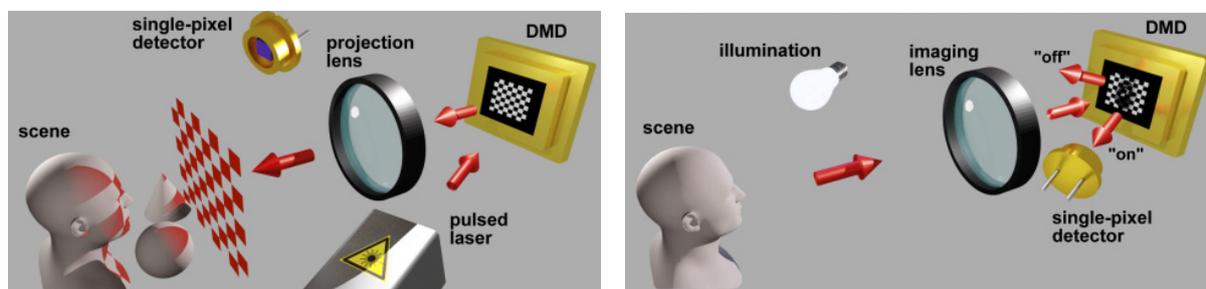
1.2.2 Élément de modulation spatiale du signal détecté

On s'intéresse maintenant à l'ajout au système LiDAR 3D d'un élément de modulation spatiale soit dans le bloc « émission », soit dans le bloc « détection » :

- Illumination structurée (Figure 1.11a) : La modulation se fait en émission en sortie de la source laser où le faisceau sortant est modulé spatialement. La scène est éclairée directement avec un profil de faisceau structuré. Les méthodes sont nombreuses et correspondent souvent au *Computational Ghost Imaging*. Cette approche est le plus souvent adaptée pour imager à travers des milieux obscurants [67] et est appliquée par exemple pour de la vision sous-marine dans de l'eau turbide.
- Détection structurée (Figure 1.11b) : La modulation se fait en réception en amont du capteur où le signal optique est modulé spatialement. La scène est éclairée par un faisceau uniforme et les photons réfléchis sont collectés *via* le bloc « réception » du LiDAR 3D. Le signal est ensuite structuré spatialement juste avant la détection. Les méthodes utilisant ce type de structuration sont communément nommées *single-pixel imaging* [67].

En comparaison avec la détection structurée, l'illumination structurée permet de générer plus de schémas de modulation différents mais est limitée à l'imagerie active par nature. A l'inverse, la détection structurée permet moins de schémas de modulation mais est compatible de l'imagerie active et de l'imagerie passive [143]. Il existe de nombreux éléments pour réaliser la modulation spatiale. Dans le cas des systèmes utilisant des capteurs mono-pixel on trouve essentiellement [67] :

- Les sources de lumière pseudo-thermiques qui permettent de moduler aléatoirement le faisceau laser en passant à travers une plaque de verre dépolie en rotation.
- Les *Liquid Crystal Spatial Light Modulators* ou LC-SLM. Dans ce cas, un cristal liquide permet de modifier la phase et/ou l'amplitude d'un faisceau laser en fonction d'un courant électrique variable.
- Les matrices de type LED programmables qui permettent par nature uniquement des modulations binaires en illumination structurée.
- Les *Digital Micro-mirror Device* ou DMD constitués par une matrice de miroirs orientables et qui permettent uniquement des modulations binaires.



(a) Illumination structurée

(b) Détection structurée

FIGURE 1.11. Deux façons d'ajouter un DMD à un système imageur mono-pixel, schémas extraits de [67]

Dans notre cas, nous cherchons à imager des cibles dures en réflexion. Le choix le plus commun pour ce type d'applications est le DMD en détection structurée [67]. Le rôle du DMD est de « compresser » le signal pendant l'acquisition et d'augmenter la résolution latérale d'un imageur lorsque le nombre de pixels d'un plan focal est limité. Cette approche s'inspire du principe du *single-pixel imaging*. Augmenter la résolution latérale par cette technique augmente le nombre d'acquisitions nécessaires. Cependant, le temps d'acquisition peut être limité grâce à des méthodes de *Compressive Sensing* que nous décrivons dans la section suivante. Les DMD possèdent de nombreuses propriétés intéressantes [79, 143]. En particulier, le taux de rafraîchissement est adapté aux applications envisagées (ex. DRI et objets en mouvement) par rapport à d'autres éléments de modulation spatiale. D'autre part, les micro-miroirs constituant la surface d'un DMD ont un coefficient de réflexion élevé et proche de 1 pour des longueurs d'onde visibles ou proche-infrarouges. Le niveau de signal est considéré comme non modifié après réflexion. La détection structurée est

programmable donc adaptable en fonction de la scène observée. Nous verrons un peu plus loin dans le manuscrit qu'il est possible de limiter le nombre d'acquisitions nécessaires à la reconstitution d'une image car la base de motifs est fortement corrélée à la structure des objets observés.

En détail, un DMD est une matrice de micro-miroirs orientables individuellement à haute cadence [83] tels que :

- Configuration dite « on » avec un angle de renvoi du signal égal à $+\theta$. Le miroir local est orienté de manière à renvoyer le signal vers le capteur.
- Configuration dite « off » avec un angle de renvoi du signal égal à $-\theta$. Le miroir local est orienté de telle sorte que le signal soit renvoyé dans une direction opposée à celle du capteur.

Notons que l'angle θ dépend du type de DMD mais qu'il varie généralement entre 10° et 12° en fonction des modèles.

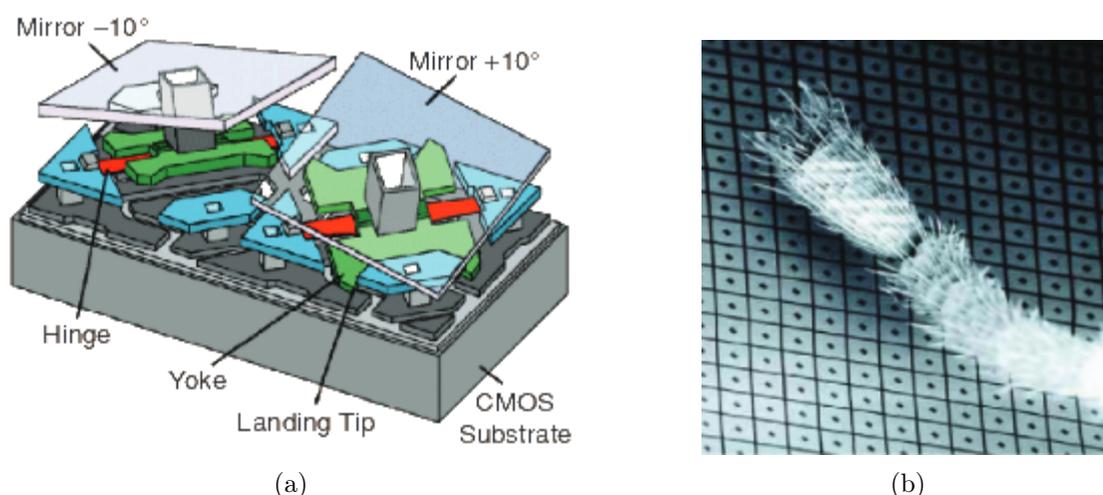


FIGURE 1.12. (a) Schéma d'un DMD avec 2 pixels orientés dans des directions opposées (configurations « on » et « off ») (b) Photo d'un DMD avec une patte de fourmi comme échelle. [54]

1.3 Méthodes de *Compressive Sensing*

Depuis le milieu du siècle précédent, le théorème de Shanon-Nyquist [110,135] est couramment utilisé pour l'acquisition de signaux et d'images. Ce théorème précise que si la fréquence d'acquisition est au minimum deux fois supérieure à la fréquence maximale de l'image, elle est retrouvée exactement. Dans le cas d'une image discrétisée, le critère de Nyquist revient à dire qu'il faut acquérir chaque pixel d'une image pour retrouver l'intégralité de ladite image, ce qui équivaut à une acquisition en scannant chaque pixel l'un après l'autre. Pour améliorer la résolution latérale des images, nous mettons en œuvre une approche par *Compressive Sensing* [96] qui permet de retrouver exactement une image

avec un nombre de mesures inférieur au critère de Nyquist à condition que le signal soit suffisamment parcimonieux dans une certaine base (structurée).

On sait qu'en utilisant des hypothèses de régularité spatiale [115], une part importante des données acquises n'apporte pas ou peu d'informations : cette approche est à la base des formats de compression JPEG par exemple. Ainsi, certaines données peuvent être supprimées sans perte significative sur la qualité de l'image c'est-à-dire que l'on peut approcher efficacement un signal en prenant ses plus grandes composantes dans une base correctement choisie.

Soient $N^2 > 0$, $x_1 \in \mathbb{R}^{N^2}$ et $\varepsilon_K > 0$, l'erreur de compression. On dit que x_1 est « compressible » si on peut trouver une base Ψ telle que :

$$x_1 = s\Psi \text{ et } \|x_1 - x_2\|_2 \leq \varepsilon_K \quad (1.13)$$

avec $x_2 \in \mathbb{R}^{N^2}$ les K plus grandes composantes de s et $\varepsilon_K \ll \|x_1\|_2$, pour des images naturelles ε_K est le plus souvent faible [115].

Le *Compressive Sensing* tire notamment parti d'une notion connexe, la **parcimonie**. En science, la parcimonie peut-être utilisée tel le rasoir d'Ockham [57] comme un principe consistant à n'utiliser que le minimum de causes élémentaires pour expliquer un phénomène [17]. Pour récupérer de l'information, on définit la parcimonie d'un vecteur de la façon suivante :

$$X \in \mathbb{R}^{N^2} \text{ } K\text{-sparse} \Leftrightarrow \|X\|_{l_0} = \#\{X_i \neq 0, i \in \mathbb{R}^{N^2}\} \leq K \quad (1.14)$$

L'idée est d'intégrer une modulation linéaire (combinaison linéaire) et non-adaptative à l'acquisition (qui ne change pas avec les données). On note donc, $N^2 \gg M > 0$ respectivement le nombre de mesures et la dimension de l'image, $Y \in \mathbb{R}^M$ le vecteur des mesures (modulées), $X \in \mathbb{R}^{N^2}$ l'image vectorisée que l'on cherche à obtenir et $\Phi \in \mathbb{R}^{M \times N^2}$ la matrice de mesure (ou la base de mesure parfois dite de « compression ») qui est connue. On a ainsi :

$$Y = \Phi X, \text{ } X \text{ } K\text{-sparse} \quad (1.15)$$

Une fois Y acquis, il faut décompresser le signal. Cela revient à résoudre le problème suivant :

$$\begin{cases} \min_{X \in \mathbb{R}^{N^2}} \|X\|_{l_0} \\ s.c. \|Y - \Phi X\|_2^2 \leq \varepsilon, \varepsilon \geq 0 \end{cases} \quad (1.16)$$

On sait que trouver une solution exacte à ce problème avec un algorithme itératif qui donne un minimum global est *NP-complexe* [107]. Certains algorithmes itératifs dits « gloutons »

(ex. *Matching Pursuit*, OMP...) arrivent à une solution approchée qui correspond souvent à un minimum local. De plus, l_0 n'est ni convexe, ni coercif, ni différentiable, ce qui rend complexe l'utilisation d'algorithmes d'optimisation [36]. Il est donc utile de définir un problème plus général qui permettrait de s'affranchir de cette difficulté. On pose le problème approché suivant [96] :

$$\begin{cases} \min_{X \in \mathbb{R}^{N^2}} g(X) \\ s.c. \|Y - \Phi X\|_2^2 \leq \varepsilon, \varepsilon \geq 0 \end{cases} \quad (1.17)$$

avec g une application mesurant la parcimonie (ex. la norme l_0 , la norme l_1 ou bien la variation totale). Le problème relaxé suivant peut aussi être considéré [96] :

$$\min_{X \in \mathbb{R}^{N^2}} g(X) + \frac{\lambda}{2} \|Y - \Phi X\|_2^2 \quad (1.18)$$

avec $\lambda > 0$.

Dans certains cas, X n'est pas parcimonieux dans la base canonique mais l'est dans une autre base Ψ . Dans ce cas précis, il est possible de minimiser la parcimonie de la représentation s de X dans la base Ψ . X est ensuite retrouvé en ramenant s dans la base canonique ($X = \Psi s$). Cela donne le problème d'optimisation suivant :

$$\begin{cases} \min_{s \in \mathbb{R}^{N^2}} g(s) \\ s.c. \|Y - \Phi \Psi s\|_2^2 \leq \varepsilon, \varepsilon \geq 0 \end{cases} \quad (1.19)$$

Avant de s'intéresser à sa résolution, on identifie que le problème n'est pas forcément bien posé en fonction de Φ . Par exemple, si Φ n'engendre pas \mathbb{R}^{N^2} , il peut ne pas exister de solution. On peut aussi remarquer que dans un cas général, si Φ engendre \mathbb{R}^{N^2} la solution ne sera pas unique ($N^2 \gg M$).

Certaines propriétés sont définies sur Φ pour que le problème soit bien posé [96]. La première propriété impose de ne pas nécessairement retrouver tout l'espace \mathbb{R}^{N^2} mais uniquement l'ensemble des vecteurs K -sparse de \mathbb{R}^{N^2} . C'est la condition d'espace nul NSP (*Null Space Property*) [96]. Dans le cas de mesures non bruitées, la NSP est nécessaire et suffisante pour garantir une erreur de l'ordre de :

$$\|\hat{X} - X\| \leq C_1 \frac{\varepsilon_K}{\sqrt{K}} \quad (1.20)$$

où \hat{X} est la reconstruction de X par un algorithme résolvant l'Équation (1.19) et g la norme l_1

Dans le cas où X est exactement parcimonieux et que la mesure n'est pas bruitée, cela permet d'obtenir une reconstitution exacte. Le problème de cette propriété (et de la norme l_0) est qu'elle est particulièrement sensible aux bruits.

T. Tao et E. Candès ont introduit une nouvelle propriété en 2006 [30] permettant d'obtenir un Φ plus robuste au bruit sous certaines conditions liées à la parcimonie des données. On dit que Φ suit la propriété d'isométrie restreinte de niveau K (K -RIP, *Restricted Isometry Property*) si pour $K < N$, $\exists \delta_K \in [0, 1[$, tel que :

$$(1 - \delta_K)\|X\|^2 \leq \|\Phi_K X\|^2 \leq (1 + \delta_K)\|X\|^2 \quad (1.21)$$

et ce pour toute matrice Φ_K composée de K colonnes extraites de Φ .

La K -RIP permet d'avoir Φ qui respecte la NSP en plus d'être robuste au bruit [31] de la façon suivante :

$$\|\hat{X} - X\| \leq C_1 \frac{\varepsilon_K}{\sqrt{K}} + C_2 \|b\| \quad (1.22)$$

où b correspond au bruit ajouté aux mesures.

Il existe d'autres propriétés comme la NSP ou la RIP appliquées au *Compressive Sensing*. Dans les travaux de Donoho [52, 53], nous identifions une approche basée sur des dictionnaires de bases, là où E. Candès, T. Tao et J. Romberg étudient plutôt des bases spécifiques (Fourier, Gaussienne, Bernoulli) avant de généraliser l'approche à d'autres bases. Dans ses travaux Donoho [52] présente la propriété dite de *spark* d'une base qui, sous certaines conditions, permet la résolution de problèmes de *Basis Pursuit*. Cette propriété est plus faible que la *Restricted Isometry Property*. En effet, le *spark* permet d'assurer l'existence et unicité d'une solution K -sparse alors que la $2K$ -RIP permet en plus d'avoir l'équivalent avec la norme l_0 et la norme l_1 [146]. Dans [53], Donoho revient sur une partie de ses travaux et de ceux développés par Candès, Tao et Romberg : il donne une approximation du nombre de mesures nécessaires pour reconstruire une image d'une certaine parcimonie, ainsi qu'une boule de faible rayon autour de la mesure dans laquelle la solution reste valide.

Dans leurs travaux, E. Candès, T. Tao et J. Romberg [29, 30, 32] présentent plusieurs bases de mesures aléatoires suivant la K -RIP, notamment les matrices Gaussienne ou de Bernoulli, où chacun des coefficients sont tirés aléatoirement en suivant la loi de probabilité associée. Dans ces mêmes travaux, ils présentent aussi des bases déterministes mais en échantillonnant aléatoirement leurs lignes comme par exemple la transformée de Fourier discrète ou les bases orthogonales comme les ondelettes. Davenport et *al.* [96] prouve que

toute matrice sub-gaussienne d'espérance nulle et de variance $\frac{1}{M}$ suit la K -RIP pour un certain δ_K , presque sûrement. Dans un premier temps, l'utilisation d'une base dont les composantes suivent une loi de Bernoulli est choisie. L'avantage de cette base, en plus de respecter la K -RIP presque sûrement est, qu'en étant définie sur deux valeurs, elle peut être encodée simplement sur un DMD. Nous reviendrons sur cette possibilité dans la Section 2.3.

1.3.1 Algorithmes classiques de résolution

Maintenant que l'on a posé le problème et le cadre du *Compressive Sensing*, on s'intéresse aux algorithmes permettant de résoudre une des formulations du problème d'optimisation (voir Équations (1.16), (1.18) et (1.19)). Pour cela, il existe de nombreux algorithmes avec différents avantages et inconvénients. On peut les regrouper en trois catégories [8, 97] :

- Les algorithmes dits « gloutons » [20]
- Les algorithmes d'optimisation non-convexe [10]
- Les algorithmes basés sur des algorithmes de relaxation convexe [28]

Bien qu'il existe différentes possibilités d'algorithmes dans la littérature, nous limitons notre étude à trois algorithmes pour chacune de ces catégories. Les algorithmes sélectionnés sont présentés sur la Figure 1.13 car ils répondent à nos attentes en termes de résolution du problème d'optimisation. Chaque algorithme est détaillé dans le paragraphe suivant et leurs performances sont intercomparées.

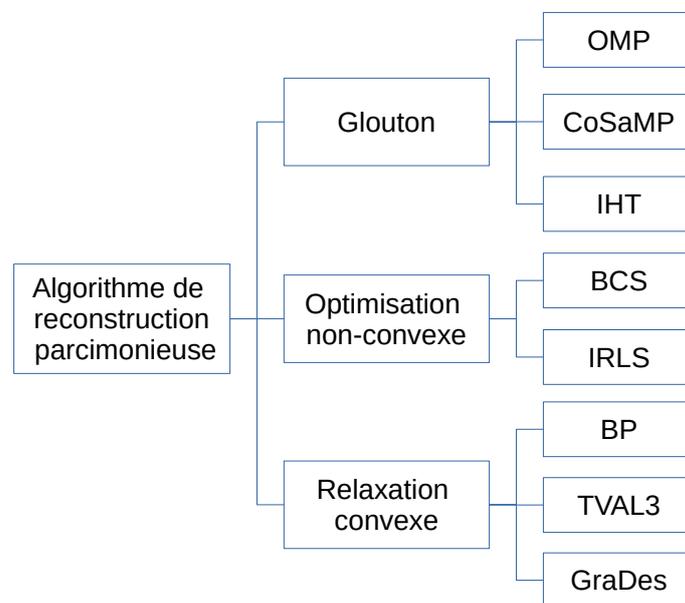


FIGURE 1.13. Classification des algorithmes de reconstruction parcimonieuse.

Les différents **algorithmes** « **gloutons** » existants résolvent les problèmes de *Compressive Sensing* en cherchant une solution de manière itérative. Ils optimisent la solution pour un minimum local à chaque pas d'itération en attendant trouver un minimum global en fin de processus [97]. Trois algorithmes ont été identifiés comme répondant à nos exigences :

Orthogonal Matching Pursuit ou OMP [148], *Compressive Sampling Matching Pursuit* ou CoSaMP [108] et *Iterative Hard Thresholding* ou IHT [18]. La majorité des algorithmes « gloutons » sont basés sur le diagramme de la Figure 1.14 pour résoudre le problème de *Compressive Sensing* de manière itérative.

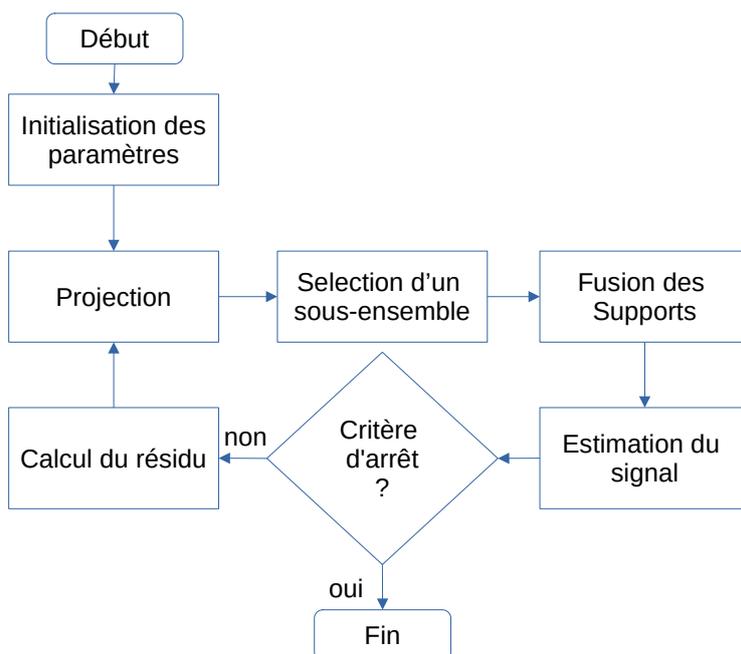


FIGURE 1.14. Diagramme des algorithmes « gloutons » [97].

L'**OMP** (*Orthogonal Matching Pursuit*) [148] est une amélioration de l'algorithme de *Matching Pursuit* qui utilise des projections orthogonales. L'idée est de partir d'un signal défini sur un support vide puis d'ajouter à chaque itération une composante du support du signal jusqu'à retrouver le signal en K itérations. Pour ce faire, on définit un résidu permettant d'évaluer la quantité restante du signal à retrouver. À chaque itération, on cherche la plus grande composante du produit scalaire entre le résidu et les colonnes de la base de mesure. Cette opération permet de trouver l'élément de la base le plus corrélé avec le résidu qui représente la partie restante du signal à reconstruire. Ensuite, une fois la composante sélectionnée, on obtient la nouvelle solution approchée en calculant la projection orthogonale sur l'espace formé par l'union du support à l'itération précédente et la nouvelle composante sélectionnée. Enfin, on met à jour le résidu. Outre ses bonnes performances qui sont détaillées dans la Section 1.3.1 de comparaison, l'OMP a l'avantage par rapport aux autres algorithmes d'avoir été beaucoup étudié [97]. Il est largement utilisé dans le cadre du traitement de signaux LiDAR mono-pixel en donnant

de bonnes reconstructions. Le cadre de son utilisation est donc clairement défini. Par exemple, certaines conditions portant sur la magnitude des composantes non-nulles nous permettent d'obtenir le support exact du signal en prenant en compte du bruit [161]. Notons enfin que l'OMP est particulièrement optimisé en temps de calcul.

Le **CoSaMP** (*Compressive Sampling Matching Pursuit*) [108] est un algorithme qui est approprié à notre problématique car il est conçu pour résoudre des problèmes de *Compressive Sampling* (autre nom du *Compressive Sensing*). L'algorithme est basé sur l'OMP décrit précédemment et s'inspire de la K -RIP pour améliorer ses performances dans le cadre du *Compressive Sensing*. Les différences majeures entre les deux algorithmes viennent de la façon de retrouver le support du signal. Dans CoSaMP, au lieu de chercher la composante la plus importante du produit scalaire entre les colonnes de la base de mesure et le résidu, on cherche les $2K$ plus grandes composantes du produit scalaire. Une fois le signal reconstruit, on ne garde que les K plus grandes composantes pour obtenir la nouvelle estimation du signal. Ainsi, on a une estimation du signal K -sparse pour chaque pas d'itérations de CoSaMP. Notons, qu'en plus de définir cet algorithme, Needle et Tropp [108] bornent l'erreur de la méthode pour différents critères d'arrêt liés au résidu dans le cadre d'acquisitions bruitées.

L'**IHT** (*Iterative Hard Thresholding*) est considéré comme un algorithme « glouton » [97] bien que l'on ne puisse pas le formuler sous la forme de la Figure 1.14. Le fonctionnement de l'IHT est assez proche de celui de CoSaMP. Ici, on part du signal puis on réduit son support en utilisant des troncatures successives de la façon suivante :

$$\hat{X}_{k+1} = H_K \left(\hat{X}_k + \Phi^T (Y - \Phi \hat{X}_k) \right) \quad (1.23)$$

où H_K est l'opérateur non-linéaire qui garde uniquement les K plus grandes magnitudes. Tel que défini à l'origine, l'IHT nécessite que $\|\Phi\|_2 < 1$ et $\forall i, \|\Phi_i\|_2 > 0$ pour assurer de bonnes propriétés de convergence [18]. Dans une autre de leurs publications, Blumensath et Davies [19] complètent les propriétés de l'IHT en ajoutant des conditions d'arrêt permettant de borner l'erreur en cas de mesures bruitées. Ils donnent aussi de nouvelles conditions sur Φ dans le cadre du *Compressive Sensing*. Ils montrent expérimentalement que les résultats de l'IHT sont proches de ceux obtenues par CoSaMP.

Les différentes approches d'**optimisation non-convexe** regroupent les algorithmes n'appartenant pas aux deux premières catégories. Deux algorithmes sont abordés : *Bayesian Compressive Sensing* (BCS) et *Iterative Reweighted Least Square* (IRLS). Ces algorithmes peuvent être classés dans sous-catégories : le BCS appartient à une catégorie d'algorithmes développées dans un cadre « bayésien », l'IRLS utilise une version approchée d'une fonction de coût non-convexe. Ces algorithmes pourraient être classés dans différentes catégories, le BCS appartient à une catégorie d'algorithmes développés dans le

cadre bayésien là où IRLS utilise une version approchée d'une fonction de coût non-convexe.

Dans le **BCS** (*Bayesian Compressive Sensing*), il est nécessaire de modifier le problème que l'on souhaite résoudre pour se placer dans un cadre bayésien. On introduit les lois conditionnelles suivantes [10] :

- (i) $p(X|\lambda) = \left(\frac{\lambda}{2}\right)^N e^{-\lambda \sum_{i=1}^N |X_i|}$, avec λ
- (ii) $p(Y|X, \sigma^2) = \mathcal{N}(\Phi X, \sigma^2)$

La loi conditionnelle (i) est appelée la densité de Laplace et permet de favoriser la parcimonie. En utilisant le maximum *a posteriori* (on ne garde que les K plus grandes valeurs), on peut ainsi retrouver un signal K -sparse. Ici, λ est un facteur lié à la parcimonie des données X . La loi conditionnelle (ii) est équivalente à la contrainte du problème d'optimisation représentée par l'équation Équation (1.16) et permet de retrouver $Y = \Phi X$. Notons que cette loi permet aussi de prendre en compte l'ajout d'un bruit qui est supposé gaussien centré de variance σ^2 [10]. Le principe des algorithmes de BCS est de partir d'une distribution *a priori* de X et Y puis de raffiner ces distributions en prenant peu à peu en compte les mesures [10]. Les contraintes et les hypothèses du problème d'optimisation équivalent sont formulées sous la forme de probabilités. Les hyper-paramètres λ et σ^2 sont aussi mis à jour dans l'algorithme (s'ils ne sont pas connus initialement). Certains algorithmes utilisent cette propriété pour estimer la probabilité d'erreur de chaque pic de X ou encore pour estimer le niveau de bruit [10, 84]. Des exemples d'algorithmes de reconstruction bayésiens, tels que *Fast Laplace* ou « Lap » et *Relevance Vector Machine* ou RVM (voire BCS) sont décrits dans la littérature et ne seront pas repris dans ce manuscrit [8].

Le principe de l'IRLS (*Iterative Reweighted Least Square*) [37] est de résoudre un problème non-convexe proche de la formulation classique du *Compressive Sensing* comme décrit dans l'équation Équation (1.16). La fonction de coût est approchée par la norme quadratique pondérée par des poids calculés à chaque itération. Le problème que l'on souhaite résoudre ici est le problème décrit dans l'équation Équation (1.19) avec g la norme l_p et $0 \geq p < 1$. Ici, on s'intéresse à une version approchée du problème telle que :

$$\left\{ \begin{array}{l} \min_{X \in \mathbb{R}^{N^2}} \sum_{i=1}^{N^2} w_i X_i^2 \\ s.c. Y = \Phi X \end{array} \right. \quad (1.24)$$

où les w_i sont des poids recalculés à chaque itération de l'algorithme. A la k -ième itération, on a :

$$w_i = \left((X_i^{(k-1)})^2 + \varepsilon_k \right)^{\frac{p}{2}-1} \quad (1.25)$$

avec $\varepsilon_k > 0$ un facteur d'amortissement permettant d'éviter les cas où w_i n'est pas défini lorsque $X_i^{(k-1)} = 0$ et $\varepsilon_k = 0$. Chartrand [37] démontre que si la suite $\varepsilon_k \xrightarrow[k \rightarrow \infty]{} 0$, que X est K -sparse et que Φ suit la $2K$ -RIP alors la solution du problème approché par l'équation Équation (1.24) tend vers le signal que l'on souhaite reconstruire.

Pour la famille des algorithmes d'**optimisation convexe**, différents algorithmes de relaxation convexe existants sont abordés pour résoudre des problèmes de *Compressive Sensing : Basic Pursuit* (BP) [40], *Gradient Descent with Sparsification* (GraDeS) [65], *Total Variation minimization by Augmented Lagrangian and ALternating direction ALgorithms* (TVAL3) [88].

Parler de **BP** (*Basis Pursuit*) comme d'un algorithme est un abus de langage courant dans la littérature concernant le *Compressive Sensing* [40]. En réalité, le BP est le nom donné au problème représenté par l'équation Équation (1.19) avec g la norme l_1 . On devrait plutôt parler de *BP-simplex* ou *BP-interior* (algorithme respectivement du simplex ou des points intérieurs pour le *Basis Pursuit*) [40]. Donoho et Chen ont aussi décrit une version adaptée à la prise en compte du bruit nommé *BPDN-simplex*, en utilisant une version pénalisée du problème [40]. Bien que théoriquement utilisable dans le cadre de l'imagerie par *Compressive Sensing*, le BP est limité lorsque les dimensions du problème d'optimisation deviennent grandes, ce qui est courant en imagerie. Pour pallier ce problème, il existe une solution classique en programmation linéaire comme la résolution du problème dual définit dans [101].

Le **GraDeS** (*Gradient Descent with Sparsification*) est une amélioration de l'IHT. En rajoutant un facteur de pénalisation γ à chaque itération, on obtient un algorithme moins exigeant sur Φ comme nous le verrons dans la Section 1.3.1 dédié à la comparaison des différentes méthodes. Il converge vers un minimum global [65] (pour un γ bien choisi). Ainsi la mise à jour de l'estimation devient :

$$\hat{X}_{k+1} = H_K \left(\hat{X}_k + \frac{1}{\gamma} \Phi^T (Y - \Phi \hat{X}_k) \right) \quad (1.26)$$

avec $\gamma > 1$. Il est possible de borner l'erreur pour le cas où $\gamma = 1 + \delta_{2K}$ (δ_{2K} est défini dans l'équation Équation (1.21)) [65].

Le **TVAL3** (*Total Variation minimization by Augmented Lagrangian and ALternating direction ALgorithms*) [88] est un algorithme de variation totale adapté au *Compressive Sensing*. Il est utilisé dans de nombreux algorithmes de *Compressive Sensing* qui minimisent la variation totale pour les LiDAR mono-pixel. La variation totale est intéressante comme norme dans le cadre du *Compressive Sensing* pour l'imagerie car, en plus d'être une bonne alternative à la norme l_0 , il permet de minimiser la variation totale en supprimer les artefacts et conserve les bords des différents objets de la scène. L'idée est

de minimiser le problème décrit par l'équation Équation (1.17) avec la variation totale en introduisant le Lagrangien. Au lieu d'optimiser directement ce Lagrangien, c'est la version pénalisée du Lagrangien qui est minimisée. La méthode de Barzilai et Borwein est ensuite utilisée pour choisir les paramètres du pas de plus profonde descente [11]. La méthode des directions alternées de Peaceman [114] permet enfin d'accélérer la minimisation du Lagrangien.

Comparaison des algorithmes

Pour comparer les performances théoriques et expérimentales des huit algorithmes introduits précédemment, nous appuyons notre étude sur trois articles [8, 64, 97] qui obtiennent des résultats similaires. La Figure 1.15 présente une comparaison des performances pour différents cas.

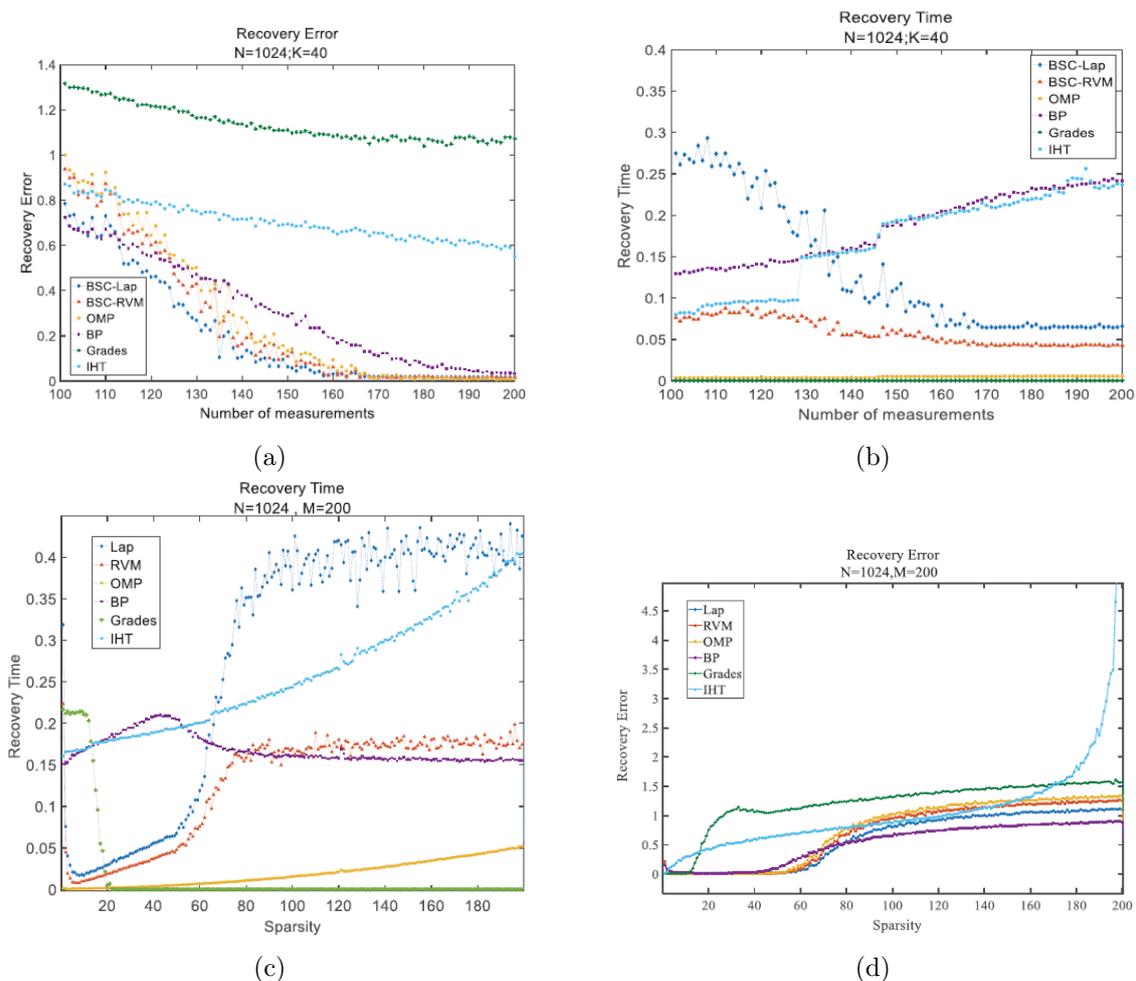


FIGURE 1.15. Comparaisons expérimentales des performances des différents algorithmes issues de [8], avec : (a) l'erreur par rapport à la dimension des mesures ; (b) le temps d'exécution par rapport à la dimension des mesures ; (c) le temps d'exécution par rapport à la parcimonie des données ; (d) Erreur par rapport à la parcimonie des données.

TVAL3 n'est pas présent dans la Figure 1.15 mais à titre de comparaison, on retrouve une comparaison numérique de quatre de ces algorithmes dans la littérature [64]. TVAL3 a un taux d'erreur inférieur aux autres algorithmes présentés mais des temps de reconstruction importants quand le taux de compression augmente. Notons que les temps de reconstruction sont faibles quand le taux de compression diminue mais qu'ils demeurent toujours supérieurs au temps de calcul de l'OMP. Les algorithmes de reconstruction les plus communs en imagerie LiDAR sont TVAL3 ainsi que l'OMP. De plus, comme déjà évoqué en introduction, on cherche à garder aussi faible que possible la latence. Pour ces deux raisons, l'utilisation de l'OMP sera préférée dans la suite de nos travaux. TVAL3 reste une très bonne alternative dans un cadre où la faible latence n'est pas une nécessité car cette approche limite l'apparition d'artefacts liés à la reconstruction. D'autres algorithmes moins conventionnels en imagerie LiDAR 3D pourraient aussi être étudiés : par exemple, dans le cas où l'on souhaite prendre en compte des *a priori* sur les formes à reconstruire, il pourrait être intéressant de considérer les algorithmes bayésiens.

1.3.2 Le *Compressive Sensing* et l'équation lidar

Le DMD nous permet d'encoder physiquement des masques binaires à partir de motifs pré-définis (base de mesure/compression) et donc de compresser le signal pendant l'acquisition des données (voir Section 1.2.2). De la même façon que dans l'équation Équation (1.5), la contribution de chaque pixel de la caméra et de chaque sous-pixel correspondant au DMD peut être explicitée pour mettre en évidence le problème de *Compressive Sensing*.

Nous distinguons le terme de « pixels » qui correspond aux éléments de surface du plan focal de la caméra de celui de « sous-pixels » qui correspond aux différents micro-miroirs constituant le DMD. On note $\Omega(R)$ la surface carrée éclairée par un faisceau laser à une distance R et $\Omega_{i,j,k}$ le sous-maillage uniforme de $\Omega(R)$ tel que $\Omega_{i,j,k}$, $\{i, j, k\} \in [0, N_p] \times [0, N_p] \times [0, N_{DMD}]$ qui correspond au signal reçu sur le pixel i, j et le sous-pixel k (N_p^2 étant le nombre de pixels de la caméra et N_{DMD} le nombre de micro-miroirs du DMD par pixel de la caméra). En utilisant la relation de Chasles dans l'équation précédente, on obtient :

$$P(t) = A_0 C_a \xi_{\lambda_l} \int_0^{\frac{ct}{2}} \frac{I(R)}{R^2} \sum_{i,j=0}^{N_p} \sum_{k=0}^{N_{DMD}} \int_{\Omega_{i,j,k}(R)} \mathcal{O}(R, \omega) d\omega dR \quad (1.27)$$

On ajoute le terme correspondant au DMD dans l'équation et pour simplifier son écriture, nous imposons $O_{i,i,k}(R) = \int_{\Omega_{i,j,k}(R)} \mathcal{O}(R, \omega) d\omega$:

$$P^m(t) = A_0 C_a \xi_{\lambda_l} \int_0^{\frac{ct}{2}} \frac{I(R)}{R^2} \sum_{i,j=0}^{N_p} \sum_{k=0}^{N_{DMD}} (\Phi_{i,j})_{m,k} O_{i,j,k}(R) dR \quad (1.28)$$

avec $(\Phi_{i,j})_{m,k} \begin{cases} 1 & \text{si le sous-pixel } i, j, k \text{ du DMD renvoi les photons vers le capteur} \\ & \text{pour le motif } m \\ 0 & \text{sinon} \end{cases}$

et $m \in [0, M]$ où M est le nombre total de mesures.

On explicite maintenant l'éclairement $I(R) = \frac{P_E(t - \frac{2R}{c})}{A_L(R)}$ d'après [100], où $P_E(t)$ correspond à l'énergie émise au temps t et $A_L(R)$ est l'aire de $\Omega(R)$. On a donc :

$$P^m(t) = A_0 C_a \xi_{\lambda_l} \int_0^{\frac{ct}{2}} \frac{P_E(t - \frac{2R}{c})}{A_L(R)R^2} \sum_{i,j=0}^{N_p} \sum_{k=0}^{N_{DMD}} (\Phi_{i,j})_{m,k} O_{i,j,k}(R) dR \quad (1.29)$$

On pose le changement de variables suivant, $T = \frac{2R}{c}$:

$$P^m(t) = \frac{2A_0 C_a \xi_{\lambda_l}}{c} \int_0^t P_E(t - T) \sum_{i,j=0}^{N_p} \sum_{k=0}^{N_{DMD}} (\Phi_{i,j})_{m,k} \frac{O_{i,j,k}(\frac{cT}{2})}{A_L(\frac{cT}{2})T^2} dT \quad (1.30)$$

Comme les fonctions sous le signe de l'intégrale sont toutes continues ou continues par morceaux $\forall t > 0$ (en supposant que $O_{i,j,k}$ est au moins continue par morceaux comme cela est le cas dans la pratique), la fonction sous le signe de l'intégrale est mesurable. On peut donc appliquer le théorème de Fubini-Tonelli et intervertir les signes somme et intégrale :

$$P^m(t) = \sum_{i,j=0}^{N_p} \frac{2A_0 C_a \xi_{\lambda_l}}{c} \int_0^t P_E(t - T) \sum_{k=0}^{N_{DMD}} (\Phi_{i,j})_{m,k} \frac{O_{i,j,k}(\frac{cT}{2})}{A_L(\frac{cT}{2})T^2} dT \quad (1.31)$$

En suivant la même logique pour la deuxième somme, on obtient :

$$P^m(t) = \sum_{i,j=0}^{N_p} \sum_{k=0}^{N_{DMD}} (\Phi_{i,j})_{m,k} \frac{2A_0 C_a \xi_{\lambda_l}}{c} \int_0^t P_E(t - T) \frac{O_{i,j,k}(\frac{cT}{2})}{A_L(\frac{cT}{2})T^2} dT \quad (1.32)$$

On pose alors, $P_{i,j}(t)$ la puissance reçue au temps t sur le pixel (i, j) du plan focal de la caméra :

$$P_{i,j}^m(t) = \sum_{k=0}^{N_{DMD}} (\Phi_{i,j})_{m,k} \frac{2A_0 C_a \xi_{\lambda_l}}{c} \int_0^t P_E(t - T) \frac{O_{i,j,k}(\frac{cT}{2})}{A_L(\frac{cT}{2})T^2} dT \quad (1.33)$$

On définit $P_{i,j,n}(t)$ la puissance cumulée reçue au n -ème bin (pas de temps de taille δ_t) sur le pixel (i, j) du plan focal de la caméra pour la modulation m :

$$P_{i,j,n}^m = \int_0^{\delta_t} P_{i,j}^m(n\delta_t + t) dt \quad (1.34)$$

$$P_{i,j,n}^m = \int_0^{\delta_t} \sum_{k=0}^{N_{DM D}} (\Phi_{i,j})_{m,k} \frac{2A_0 C_a \xi_{\lambda_l}}{c} \int_0^t P_E(n\delta_t + t - T) \frac{O_{i,j,k}(\frac{cT}{2})}{A_L(\frac{cT}{2})T^2} dT dt \quad (1.35)$$

Comme $P_E(t - T) = 0, \forall T > t$ on peut écrire :

$$P_{i,j,n}^m = \int_0^{\delta_t} \sum_{k=0}^{N_{DM D}} (\Phi_{i,j})_{m,k} \frac{2A_0 C_a \xi_{\lambda_l}}{c} \int_0^{\infty} P_E(n\delta_t + t - T) \frac{O_{i,j,k}(\frac{cT}{2})}{A_L(\frac{cT}{2})T^2} dT dt \quad (1.36)$$

Toujours en appliquant le théorème de Fubini-Tonelli, on a :

$$P_{i,j,n}^m = \sum_{k=0}^{N_{DM D}} (\Phi_{i,j})_{m,k} \frac{2A_0 C_a \xi_{\lambda_l}}{c} \int_0^{\infty} \int_0^{\delta_t} P_E(n\delta_t + t - T) dt \frac{O_{i,j,k}(\frac{cT}{2})}{A_L(\frac{cT}{2})T^2} dT \quad (1.37)$$

Pour faire ressortir la convolution, on pose :

$$f(T) = \frac{2A_0 C_a \xi_{\lambda_l}}{c} \int_{n*\delta_t}^{(n+1)*\delta_t} P_E(T + t) dt \text{ et } O_{i,j,k}^A(T) = \frac{O_{i,j,k}(\frac{cT}{2})}{A_L(\frac{cT}{2})T^2} \quad (1.38)$$

Ici O^A représente une information : la géométrie des objets de la scène discrétisée en suivant les pixels du DMD pour chaque pixel du plan focal de la caméra est atténué en fonction de la distance de l'objet à la caméra. En remplaçant f et O^A dans l'équation précédente, on obtient :

$$P_{i,j,n}^m = \sum_{k=0}^{N_{DM D}} (\Phi_{i,j})_{m,k} \int_0^{\infty} f(n\delta_t - T) \times O_{i,j,k}^A(T) dT \quad (1.39)$$

$$P_{i,j,n}^m = \sum_{k=0}^{N_{DM D}} (\Phi_{i,j})_{m,k} (f * O_{i,j,k}^A)(n\delta_t) \quad (1.40)$$

Ainsi, en considérant $(Y_{i,j,n})_m = P_{i,j,n}^m$, et $(X_{i,j,n})_k = (f * O_{i,j,k}^A)(n\delta_t)$, on retrouve un problème de *Compressive Sensing* :

$$Y_{i,j,n} = \Phi_{i,j} X_{i,j,n} \quad (1.41)$$

Donc, le problème de *Compressive Sensing* est résolu pour chaque pas de temps et en chaque pixel du plan focal de la caméra afin de retrouver l'information latérale souhaitée O^A , à une convolution près.

1.4 Logiciel MATLIS : simulations de signaux issus de LiDAR 3D

1.4.1 Utilisation et propriétés

Développé par l'ONERA depuis 2006, le logiciel MATLIS [7] permet de simuler des signaux issus de systèmes d'imagerie active. Sa raison d'être est triple. Tout d'abord la simulation à des fins de dimensionnement de systèmes optroniques en faisant varier *in-silico* les différents paramètres du système et environnementaux pour évaluer leurs impacts sur les performances. Ensuite, il est utilisé pour prototyper, évaluer et implémenter des algorithmes de traitement de signal LiDAR. Enfin une utilisation pour la fermeture de boucle de la co-conception, capteur-traitement vise à dimensionner et évaluer conjointement les performances du système (au sens matériel + logiciel de traitement de signal).

La simulation physique implémentée dans MATLIS modélise le processus d'imagerie active en utilisant des modèles physiquement réalistes développés et validés à partir de résultats de laboratoire et de mesures réalisées en environnement maîtrisé sur le terrain.

La boucle de calcul du logiciel se décompose suivant :

- Construction de l'impulsion laser, que ce soit dans la dimension temporelle (longueur et forme de l'impulsion) et dans la dimension latérale (divergence, forme du faisceau).
- Propagation aller de l'impulsion dans l'atmosphère. Cette propagation inclut notamment l'interaction avec les aérosols et l'impact de la turbulence atmosphérique sur la répartition latérale de l'énergie.
- Construction de la réponse temporelle de la cible, en fonction de la géométrie et des propriétés optiques des matériaux.
- Propagation retour de l'impulsion prenant en compte la turbulence et les aérosols.
- Modèle capteur.

Construction de l'impulsions laser

Dans le cadre de ces travaux, l'impulsion laser est définie dans la simulation par l'Équation (1.7).

La répartition spatiale du faisceau laser est considérée comme uniforme dans tout le champ de vue de l'imageur (hypothèse d'un faisceau mis en forme), la divergence est considérée comme égale au champ de vue du détecteur.

Propagation atmosphérique (aller et retour)

Durant la propagation atmosphérique, deux phénomènes co-existent. Les aérosols (poussières, gouttelettes de liquide), qui produisent un signal lié à la diffusion de la lumière

et l'absorption. La turbulence optique, qui est due aux variations d'indice optique sur la ligne de visée introduit une inhomogénéité latérale de l'intensité. Dans le cadre de ces travaux, l'effet des aérosols est réduit à de l'absorption le long de la ligne de visée. La turbulence est négligée (cas d'un détecteur positionné à quelques mètres de la surface du sol).

Interactions avec la scène

Dans le logiciel MATLIS, la scène est définie par la combinaison d'une géométrie 3D, de matériaux, et de mouvements. Lors de l'interaction de l'impulsion laser avec la scène, celle-ci est figée (les temps d'interaction avec la scène étant largement plus faible que l'échelle des mouvements des objets décrits).

La construction de la réponse de la scène avec l'impulsion laser est obtenue par convolution de l'impulsion laser décrite précédemment avec une somme d'impulsions de Dirac. Cette méthode permet, par échantillonnage latéral de reconstruire la réponse de surfaces pouvant être inclinées par rapport à la normale à la direction de propagation de l'impulsion laser sans avoir à recourir à un produit de convolution coûteux en temps de calcul.

Modèle capteur

Le logiciel MATLIS dispose de plusieurs modèles capteurs permettant de simuler une grande partie des détecteurs disponibles sur le marché. Dans le cadre de ces travaux, le modèle capteur utilisé est le modèle GmAPD. Ce modèle prend en compte les effets du bruit et de la non-linéarité du capteur qui seront présentés plus loin.

Dans le cadre de nos travaux, l'utilisation d'un logiciel de permet d'avoir un contrôle total sur les paramètres de simulation. Cette liberté de choix des paramètres permet d'évaluer rapidement de nombreux cas de figure. De plus l'utilisation de la simulation permet l'accès à des informations qui ne sont pas disponibles dans la réalité, comme par exemple le nombre de photons reçus par le capteur GmAPD ou la nature de l'évènement détecté (signal ou bruit).

La grande liberté offerte par la simulation dans le choix des paramètres implique une attention particulière dans la sélection des valeurs pour rester dans les contraintes de la physique et des possibilités technologiques actuelles.

1.4.2 Définition des cas tests

De par sa nature, la chaîne de traitement CRISPS-3D développée dans le Chapitre 2 est sensible à la géométrie de la scène observée. Pour tester la robustesse de la chaîne algorithmique, plusieurs scènes, illustrées ci-dessous, ont été évaluées.

La scène décrite sur la Figure 1.16a fournit plusieurs types de surfaces planes, y compris des pentes. A partir de celles-ci, nous évaluons la fréquence spatiale, les résolutions latérales et temporelles et les résultats de notre chaîne d'acquisition.

La scène Figure 1.16b fournit des détails plus fins, avec une distribution plus large du contenu en fréquence spatiale. La devise française et le buste de Marianne sont utilisés pour illustrer les qualités et les limites de notre processus.

Toutes les simulations effectuées dans cette thèse visent à améliorer la résolution d'un facteur soixante-quatre.

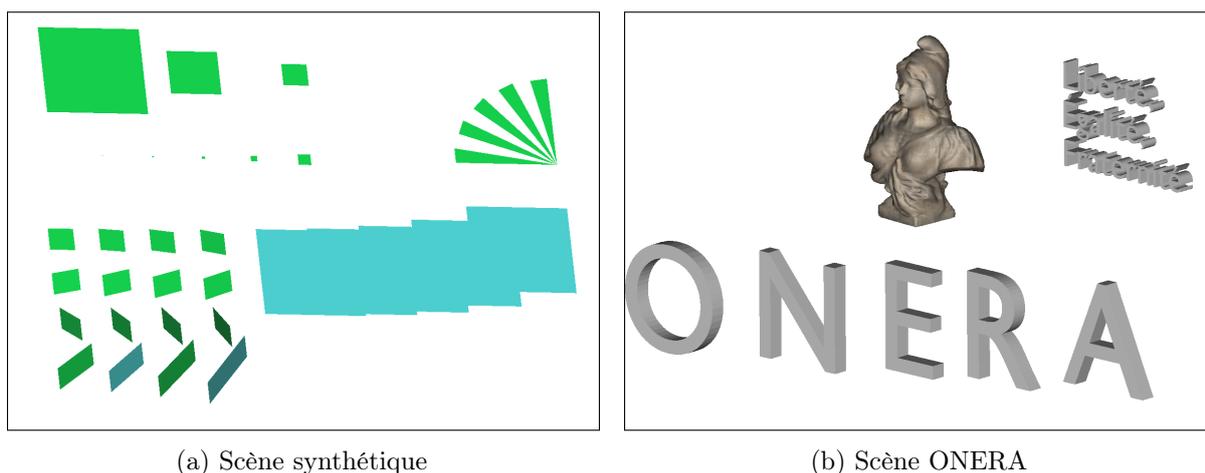


FIGURE 1.16. Scènes sélectionnées pour les tests de simulation. (a) Une scène synthétique à 13km du capteur avec un champ de vue de 0,8mrad dans les directions horizontale et verticale équivalent à une résolution de 4 cm à 13 km. (b) La devise française, un buste de la Marianne française et des lettres « ONERA » de 1 m de large à 13 km du système LiDAR 3D avec un champ de vue carré de 0,4 mrad équivalent à une résolution de 2 cm à 13 km.

Les objets utilisés sont définis comme ayant une réflectance hémisphérique lambertienne de 10% afin d'obtenir une réponse uniforme de tous les pixels du capteur. La caméra est simulée à une hauteur suffisante pour négliger l'effet de la turbulence atmosphérique. Nous présentons des spécifications dans la Tableau 1.2 compatibles avec un LiDAR 3D à comptage de photons utilisant des composants disponibles dans le commerce. En particulier, l'état-de-l'art des sources LASER fibrées robustes compatibles avec les systèmes à faible SWaP-C utilisant une énergie suffisante ($\approx 100 \mu\text{J}$) et une longueur d'impulsion de l'ordre de la nanoseconde est d'environ 20 kHz [26] à $1,55 \mu\text{m}$. Les caméras GmAPD disponibles dans le commerce peuvent atteindre des fréquences d'images de 186 kHz [81].

LASER	
Largeur d'impulsion	0,25 ns
Energie de l'impulsion	100 μ J à 150 μ J
Fréquence de répétition d'impulsion	20 kHz
Longueur d'onde	1550 nm
Capteur	
Fréquence d'acquisition du GmAPD	186 kHz
Fréquence de rafraichissement du DMD	5 kHz
Longueur de Bin	0,25 ns
DCR	1 MHz
Efficacité quantique	0,4

TABLE 1.2. Spécifications du système LiDAR 3D.

1.5 Synthèse

Dans ce chapitre nous avons introduit les différentes notions nécessaires à la compréhension des travaux sur le *Compressive Sensing* pour les LiDAR 3D. Nous avons aussi évoqué les différentes techniques et méthodes d'acquisition et de reconstruction. Les approches sélectionnées pour nos travaux ont été présentées en expliquant chacun de nos choix. Nous avons pu voir comment l'ajout de la modulation spatiale permet d'augmenter la résolution dans le cas des caméras mono-pixel puis comment le *Compressive Sensing* permet de réduire drastiquement le nombre d'acquisitions. Enfin avec le logiciel MATLIS de l'ONERA, nous avons introduits les moyens disponibles pour évaluer et valider nos différentes méthodes sur des scènes représentatives. Cela nous permettra dans le prochain chapitre d'expliquer et de valider le fonctionnement de notre chaîne de traitement nommée CRISPS-3D (*Compressive Reconstruction of Images for Single Photon System in 3D*) [154].

Chapitre 2

CRISPS-3D : une chaîne de traitement générale pour le *Compressive Sensing* appliquée aux matrices de photodiodes à avalanches en mode Geiger

Points clefs

- Il n'existe pas de systèmes d'imageurs 3D dans le SWIR à forte résolution latérale ($\geq 256 \text{ pixels} \times 256 \text{ pixels}$) permettant des acquisitions à longue portée ($> 10 \text{ km}$) en temps d'acquisitions limité ($< 0,05 \text{ s/motif}$) avec une faible latence pouvant être embarqués (faible SWaP-C).
- Il est nécessaire de bien appréhender la physique des systèmes et des scènes pour reconstruire efficacement une image plus résolue. Certaines altérations comme les effets liés aux temps morts sont limitants pour l'imagerie 3D.
- Il est possible de réaliser un tel système sous un certain nombre de contraintes, ce qui a été démontré numériquement. D'autres approches d'intérêt permettant l'amélioration de CRISPS-3D sont discutées.

Contributions principales

- Définition d'une nouvelle stratégie de débruitage exploitant les spécificités du système LiDAR : en particulier, la différence entre la fréquence d'impulsion de la source laser et la fréquence d'acquisition de la caméra. Cette différence de fréquences permet l'acquisition d'images passives (sans impulsion laser) entre deux images actives pour accéder au bruit. Notre stratégie est d'utiliser ces deux types d'images pour différencier statistiquement le bruit du signal.
- Extension des travaux de Liu et *al.* [91] sur l'application du *Compressive Sensing* à des données LiDAR dans un régime à haut flux, *i.e.* soumis aux effets de temps morts.
- Développement de CRISPS-3D [154], une chaîne de traitement optimisée pour augmenter la résolution latérale de matrices de GmAPD en utilisant du *Compressive Sensing*. Validation de CRISPS-3D avec des données simulées par le logiciel ONERA MATLIS. CRISPS-3D permet des acquisitions au moins 64 fois plus résolues latéralement que les caméras commercialement disponibles dans le cadre spécifique des caméras à comptage de photons en mode Geiger.

L'objectif de ce chapitre est de développer une méthode pour reconstruire numériquement des nuages de points en 3D avec une meilleure résolution latérale en utilisant du *Compressive Sensing* sous les contraintes imposées : (i) très longue portée utilisant une matrice de capteurs GmAPD en temps d'acquisitions limités ($<0,05$ s/motif) avec un faible SWaP-C et (ii) très faible niveau de signal (<1 photon en moyenne) avec une résolution temporelle suffisante.

2.1 Présentation et motivations de CRISPS-3D

Afin de reconstruire une image 3D, la chaîne de traitement CRISPS-3D utilise le *Compressive Sensing*. Un certain nombre d'algorithmes de traitements du signal sont définis. Chacun de ces algorithmes est représenté schématiquement sur la Figure 2.1 et inclut les étapes suivantes.

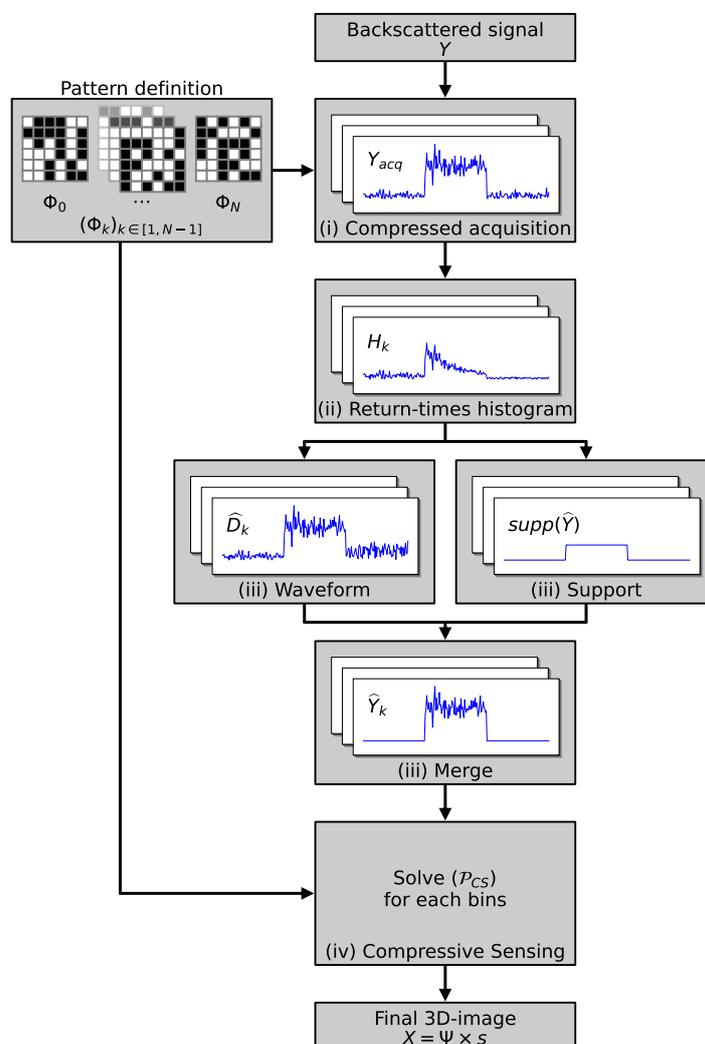


FIGURE 2.1. Illustration de la chaîne de traitement CRISPS-3D et des notations associées en considérant un exemple de reconstruction pour un pixel.

(i) **Acquisition compressée**

Afin d'appliquer le *Compressive Sensing* pour reconstruire une image, le signal supposé compressible est d'abord modulé physiquement par une base de mesures. A ce stade, le signal n'est toujours pas détecté.

(ii) **Histogramme des temps de retours**

La détection se fait en utilisant des capteurs à comptage de photons comme les GmAPD. Ces capteurs ont une dynamique de 1 (0 dB). Ceci rend complexe l'acquisition d'images. Du fait de l'utilisation de matrices GmAPD, on ne détecte que le premier photoélectron sans information sur le nombre de photons incidents (voir Section 1.1.1). Ici, nous considérons une méthode pour augmenter la dynamique de la caméra. Dans la plupart des cas, les signaux acquis à l'aide d'un capteur GmAPD sont reconstruits en utilisant une méthode dite de *Time-Correlated Single Photon Counting* (TCSPC) [12] décrite sur la Figure 2.2. Cette méthode utilise le fait que la probabilité de détecter un photon au cours d'une période suffisamment courte (ci-après dénommée « pas de temps temporel ») est très inférieure à l'unité lorsque l'intensité lumineuse est faible. Cette condition se manifeste par exemple lors d'acquisitions à longue portée avec des détecteurs de type GmAPD. Il suffit de considérer la détection d'un photon comme un évènement statistique à enregistrer. Ainsi, en mesurant le temps de chaque détection par rapport à une référence commune, un histogramme des temps de retour des photons peut être construit. L'onde complète est statistiquement reconstruite. La forme de l'onde complète permet de remonter à la géométrie de la scène.

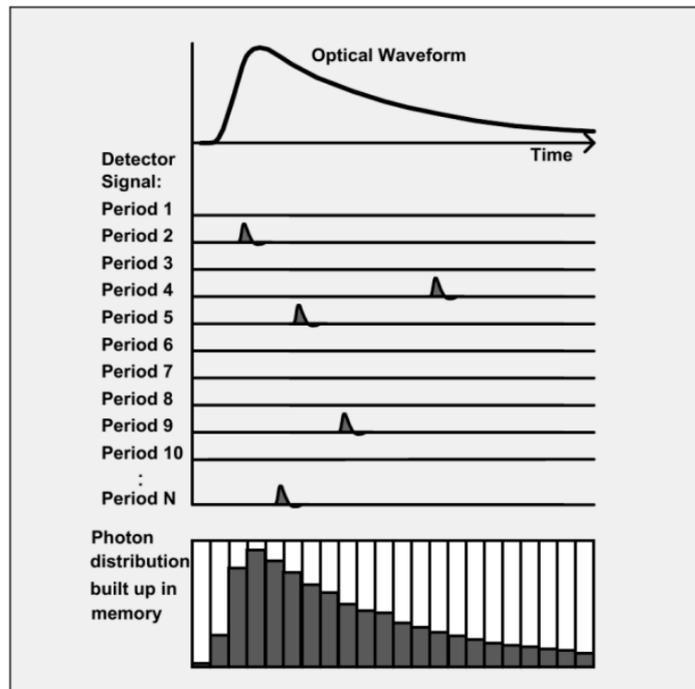


FIGURE 2.2. Principe général de la méthode TCSPC extrait des travaux de Becker et al. [12]

(iii) Reconstruction de l'onde complète (correction du *Dead-Time Effect*)

Lorsque le niveau de signal est suffisamment faible par rapport à l'unité, il n'est pas nécessaire d'avoir la possibilité de détecter plusieurs photons au cours d'une période de signal [159]. En effet, pour des niveaux de signal inférieurs à 0,1 photon, il est assez peu probable que de multiples échos soient détectés par le capteur pour une même impulsion. Cependant, pour maintenir la dynamique obtenue grâce à l'histogramme de l'onde complète reconstruite tout en limitant d'autant le niveau de signal, il est nécessaire d'augmenter proportionnellement le nombre d'acquisitions. De plus, bien qu'il soit possible de diminuer le signal reçu pour un contexte donné, le bruit ne diminue pas (notamment le DCR, voir le Chapitre 1) donc le SNR diminue nécessairement. Un certain nombre de travaux s'intéressent aux acquisitions au-delà de cette limite en corrigeant cet effet que nous nommons *Dead-Time Effect* (DTE). Dans la littérature, il est parfois appelé *pile-up* [71, 120].

Une diminution du SNR ou de la dynamique peut fortement limiter la capacité d'identification du système [5]. De plus, nous verrons que le DTE peut entraîner une plus forte dégradation de la reconstruction dans le cas de l'augmentation de la résolution latérale par une méthode de *Compressive Sensing* [91]. Ce phénomène est explicité dans la Section 2.2.1 puis illustré dans la Figure 2.10 et la Figure 2.11.

Pour conserver une fréquence d'acquisition et un SNR élevés, nous proposons l'ajout d'un algorithme de correction du DTE.

(iv) Estimation du support (pour débruiter le signal)

Pour la même raison, un algorithme permettant de supprimer le bruit est nécessaire pour garder un SNR suffisant. Pour cela, nous proposons une nouvelle stratégie exploitant la différence de fréquences « image » entre le détecteur et la source laser (c.f. Tableau 1.2). Cette différence permet l'acquisition d'images sans illumination laser pendant le temps mort de la source. Ces acquisitions supplémentaires sont exploitées pour débruiter le signal en estimant son support. L'idée est d'effectuer un test statistique sur chaque intervalle de temps, pour chaque pixel, afin de séparer le signal du bruit et d'estimer directement son support.

(v) Fusion des informations

Le signal estimé avant *Compressive Sensing* est obtenu grâce à une multiplication point à point entre le support estimé et le signal corrigé.

(vi) *Compressive Sensing*

Une fois le signal traité et pour chaque pixel de la matrice du capteur, on applique un solveur de *Compressive Sensing* classique à chaque pas de temps afin de reconstruire l'image 3D finale. L'ensemble du processus est regroupé dans la chaîne de traitement CRISPS-3D.

2.2 Algorithmes de pré-traitement

2.2.1 Reconstruction de l'onde complète

Le DTE se produit lorsque le capteur reçoit un signal avec de multiples échos et suffisamment d'intensité (niveau de signal $>0,1$ photon en moyenne). Le coefficient d'un motif est défini comme étant le signal acquis pour un pas de temps (discret, souvent appelé « *bin* » dans la littérature) et un motif donné. Considérons deux acquisitions avec deux motifs différents mais une section efficace identique. Dans ce cas, leurs coefficients de motif doivent être égaux indépendamment des détections aux instants précédents. En effet, le coefficient d'un motif correspond à une mesure notée Y . Dans le cadre du *Compressive Sensing*, des erreurs sur la mesure introduisent nécessairement des erreurs lors de la reconstruction de l'image en trois dimensions. Ces erreurs se manifestent principalement sous la forme d'artefacts sur le nuage de points reconstruit. La Figure 2.3 illustre les corrélations temporelles. Deux motifs sont sélectionnés : (i) le motif Φ_1 ne cache qu'un seul objet et (ii) les deux objets sont visibles en appliquant le motif Φ_2 . Dans le premier cas (Φ_1), le signal atteint une hauteur de pic deux fois plus élevée que dans le second cas. Le signal récupéré pour Φ_2 devrait présenter une hauteur de pic identique pour l'objet « triangle » car ils ont une section efficace identique. Cet exemple illustre bien l'impact de la combinaison entre le DTE et le motif sur le signal reconstruit. Le même effet se produit pour une pente (signal retour élargi) ou toute combinaison de cibles produisant des échos multiples.

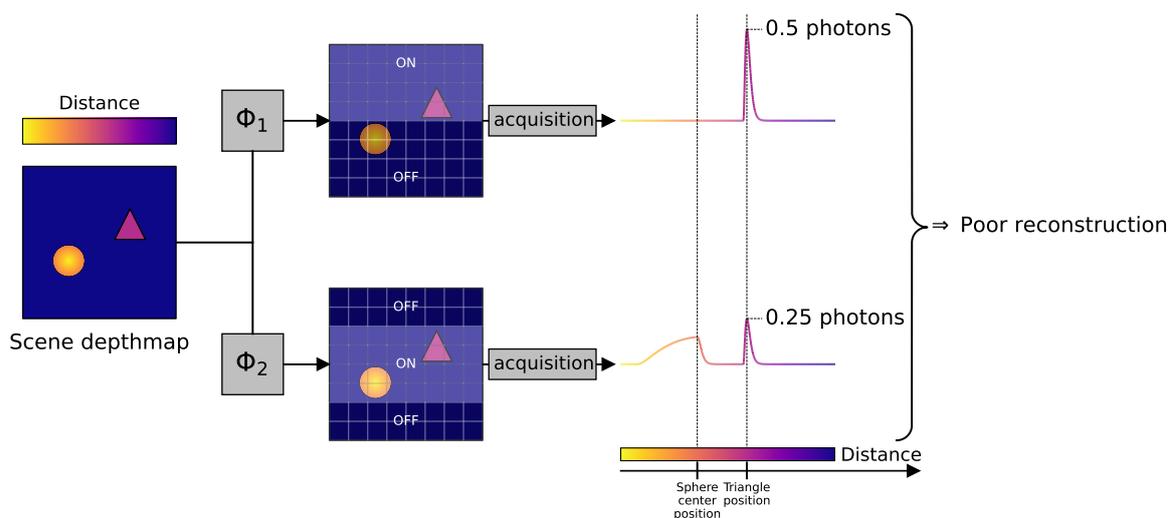


FIGURE 2.3. Illustration d'un cas où la corrélation temporelle est particulièrement forte, ce qui conduit à une image 3D imprécise.

État-de-l'art sur le DTE

Pour traiter le DTE, deux méthodes sont couramment utilisées et permettent de :

- Diminuer le niveau du signal pour atteindre le domaine où les GmAPD ont un comportement statistiquement linéaire, typiquement 0,01 photon en moyenne [159]. C'est l'approche la plus commune.
- Définir un processus pour corriger le DTE induisant la corrélation temporelle [44, 45, 71, 91, 121].

Dans les avancées récentes sur la correction du DTE, une méthode inverse basée sur la méthode bayésienne a été utilisée en démontrant une précision temporelle inférieure à la picoseconde pour des images 3D et une longueur d'impulsion de 50 ps [71]. Une autre approche récente proposée par Rapp et *al.* [121] définit un modèle de correction probabiliste utilisant des chaînes de Markov en considérant un flux de lumière élevé dans le cadre du comptage de photons avec des APD. Cette méthode a été pensée pour des capteurs APD spécifiques (non parallélisables et asynchrones) et permet des mesures avec une plage dynamique plus importante. En effet, Rapp et *al.* [121] montre qu'il faut 100 fois plus d'acquisitions en régime linéaire ($<0,01$ photon) pour obtenir des résultats équivalents à des acquisitions allant jusqu'à 5 photons avec leur méthode de correction sinon, seuls les flux de lumière les plus importants sont détectés. Ainsi, en considérant un temps d'acquisition limité ne permettant pas d'augmenter suffisamment le nombre d'acquisitions et une détection, cette méthode permet d'imager avec une plage dynamique plus importantes (dépendant de l'efficacité de la détection).

Nous définissons dans la Section 2.2.1 un estimateur de signal qui fonctionne lorsque le rapport entre la longueur de l'impulsion et le pas de temps est faible (généralement inférieur ou égal à l'unité). Pour utiliser des impulsions plus longues, il faut envisager une autre stratégie algorithmique [71, 121], plus précise dans ces conditions, au détriment de la vitesse de calcul et donc de la latence totale.

Correction du DTE dans CRISPS-3D

Les capteurs GmAPD sont conçus pour détecter des événements au niveau du photon unique. Ils ont une dynamique unitaire puisqu'ils ne peuvent détecter qu'un événement sans information sur le nombre de photons. Dans ce cadre, O'Brien et Fouche [111] ont défini et validé le modèle suivant :

$$\mathbb{P}(p; t_1; t_2) = \frac{1}{p!} [Y(t_1, t_2)]^p e^{-Y(t_1, t_2)} \quad (2.1)$$

où $\mathcal{P}(p; t_1; t_2)$ est la probabilité, pour un modèle donné, qu'une quantité p d'événements se produise entre les temps t_1 et t_2 et $Y(t_1, t_2)$ est la fonction de taux du processus [111] qui varie avec le bruit, la scène (y compris les propriétés optiques et géométriques) et la configuration optique.

Compte-tenu de l'état de l'art des caméras, nous ne considérons que des pas de temps de taille égale. Comme les capteurs GmAPD ne peuvent que mesurer que « au moins un photon a été détecté », nous définissons deux probabilités pour chaque pas de temps :

$$\begin{cases} \mathbb{P}(p_k = 0) = \mathbb{P}(0; t_k; t_{k+1}) = e^{-Y(t_k, t_{k+1})} \\ \mathbb{P}(p_k > 0) = 1 - \mathbb{P}(0; t_k; t_{k+1}) = 1 - e^{-Y(t_k, t_{k+1})} \end{cases} \quad (2.2)$$

où p_k est le nombre d'événements qui se produisent entre t_k et t_{k+1} . Cette probabilité suppose que toutes les acquisitions sont indépendantes à chaque pas de temps et que seul le premier retour est détecté. Ainsi, nous définissons la probabilité D_k qu'une détection soit faite au pas de temps k en Équations (2.3) et (2.4) en tenant compte de la dépendance temporelle :

$$D_k = \mathbb{P}(p_k > 0) \prod_{j=0}^{k-1} \mathbb{P}(p_j = 0) \quad (2.3)$$

$$\Rightarrow D_k = (1 - e^{-Y_k}) \prod_{j=0}^{k-1} e^{-Y_j} \quad (2.4)$$

où $Y_k = Y(t_k, t_{k+1})$ est la fonction de taux du processus entre les temps t_k et t_{k+1} . Ainsi, pour retrouver la forme d'onde correcte, nous devons approcher la fonction de taux Y_k pour chaque pas de temps k . Le signal est récupéré à partir de l'histogramme des temps de retour. Soit H_k l'histogramme normalisé au pas de temps k , tel que :

$$H_k = \frac{1}{N} \sum_{j=0}^{N-1} \mathbb{1}_{z_j=k} \quad (2.5)$$

où N est le nombre d'acquisitions par motif utilisé, et z_j correspond à la case où le signal est détecté pendant l'acquisition j . D'après la loi des grands nombres, H_k converge vers la probabilité D_k de détection à la case k .

$$H_k \xrightarrow[N \rightarrow \infty]{} \mathbb{E} [\mathbb{1}_{z=k}] = \mathbb{P}(z = k) = D_k \quad (2.6)$$

En ce sens, H_k est une approximation asymptotique de D_k . Nous utilisons l'Équation (2.6) pour en déduire une approximation de tous les $(Y_k)_{k \in [0, K-1]}$, ce qui est le but de cette étape. Si $D_k \in]0, 1]$, alors nous déduisons de l'Équation (2.4) que :

$$e^{-Y_k} = 1 - D_k \times \left(\prod_{j=0}^{k-1} Y_j \right)^{-1} \in]0, 1] \quad (2.7)$$

Nous avons $H_k \in]0, 1]$ par définition. En utilisant le logarithme de l'Équation (2.7) et en approximant D_k par H_k , on définit \hat{Y}_k un estimateur de Y_k par récurrence dans

Équation (2.8).

$$\hat{Y}_k = -\ln \left(1 - e^{\sum_{j=0}^{k-1} \hat{Y}_j} \times H_k \right) \quad (2.8)$$

En résumé, l'Équation (2.8) permet d'estimer le signal à l'aide d'histogrammes. Cette méthode est inspirée du processus de correction du DTE de Coates [44, 45] et de Liu et al. [91] qui est une version étendue pour des capteurs GmAPD asynchrones dans une configuration *Compressive Sensing*. On a pu montrer par la suite que cet estimateur, celui de Coates ou l'estimateur obtenu par minimisation de la vraisemblance sont tous égaux (c.f. Annexe A).

Ainsi, comme décrit dans la Section 2.1, une fois le signal corrigé, il doit être débruité. Pour cela, il est nécessaire d'estimer le support de la fonction.

2.2.2 Estimation du support de la fonction

Nous introduisons une nouvelle méthode pour débruiter un signal en utilisant les caractéristiques de la configuration définie précédemment (Section 1.4.2) à notre avantage. Le bruit est presque invariant dans le temps pour chaque pixel, mais variable entre les pixels de la matrice GmAPD [82]. Comme mentionné précédemment, il est possible d'acquérir quelques images passives (au moins une) qui correspond à une acquisition sans signal utile (identifiable au bruit) entre chaque image active (*i.e.* avec une illumination laser concomitante). Nous comparons statistiquement l'ensemble des coefficients de motifs disponibles pour un pas de temps k donné avec (H_k^1) et sans (H_k^0) illumination laser concomitante. Ainsi, nous déterminons le support signal directement à partir de la forme d'onde, avec une probabilité fixée α . Pour tous les modèles, nous calculons le support du signal et de la forme d'onde corrigée. Ensuite, nous multiplions les deux comme illustré sur la Figure 2.1, étape (iii).

Nous testons statistiquement chaque ensemble de pas de temps avec un signal potentiel H_k^1 contre ceux sans signal concomitant H_k^0 pour définir quelle population est la plus grande. Le support de la fonction est ensuite construit à partir des pas de temps où H_k^1 est stochastiquement plus grand.

Nous définissons l'hypothèse nulle comme suit : « les distributions des deux populations sont identiques ». Le plus souvent, un test de *Student* est utilisé pour tester cette hypothèse statistique [16]. Dans notre cas, les distributions de H_k^1 et H_k^0 ne sont pas normales. De fait, l'utilisation de tests non-paramétriques, sans hypothèse supplémentaire sur le bruit ou la distribution des données, est préférable. Nous choisissons parmi eux le test de rang de Mann-Whitney [95]. Il est défini par l'Équation (2.10) et illustré dans la Figure 2.4. Ici, nous testons la forme d'onde entière pour chaque pixel, en considérant chaque coefficient de motif par pas de temps comme un élément de la population.

En pratique, nous concaténons H_k^1 et H_k^0 pour former une population globale. Ensuite, la liste résultante est triée par nombre de détections au pas de temps. La statistique U est définie comme étant le nombre de fois où H_k^0 précède H_k^1 comme dans l'Équation (2.9). Enfin, U est utilisé pour comparer le rang des éléments des deux populations :

$$U = \sum_{m=1}^M \sum_{l_1=1}^{N_{pulse}} \sum_{l_2=1}^{N_{noise}} \mathbb{1}_{(H_k^1)_{m,l_1} > (H_k^0)_{m,l_2}} + 0.5 \mathbb{1}_{(H_k^1)_{m,l_1} = (H_k^0)_{m,l_2}} \quad (2.9)$$

où M est le nombre de motifs, $(\cdot)_{m,l}$ les acquisitions de coefficient l du motif m , N_{pulse} et N_{noise} sont respectivement le nombre d'acquisitions avec et sans illumination laser concomitante.

$$P(U \leq \bar{U}) \leq \alpha, \text{ sous l'hypothèse nulle} \quad (2.10)$$

\Rightarrow l'hypothèse de distribution identique est rejetée

Ainsi, nous considérons le test comme significatif avec le niveau α si la probabilité que U soit inférieure ou égale à \bar{U} sous l'hypothèse nulle, c'est-à-dire si les distributions étaient identiques. Sous l'hypothèse nulle, les deux distributions sont égales et les rangs sont donc uniformément aléatoires. Le paramètre α détermine le niveau de tolérance auquel nous pouvons affirmer qu'une case est dans le support en rejetant l'hypothèse nulle.

En pratique, la probabilité $P(U \leq \bar{U})$ n'est jamais calculée dans CRISPS-3D. En effet, il est possible en connaissant α , N_{noise} et N_{pulse} de calculer U_{crit} identifiée comme étant une valeur critique de U à partir de laquelle cette probabilité sera inférieure à α . Le calcul de U_{crit} est fait en amont de la chaîne de traitement. Lors de l'exécution, seule une comparaison entre U et U_{crit} est réalisée. U_{crit} est déterminée par récurrence en suivant la formule définie dans [25].

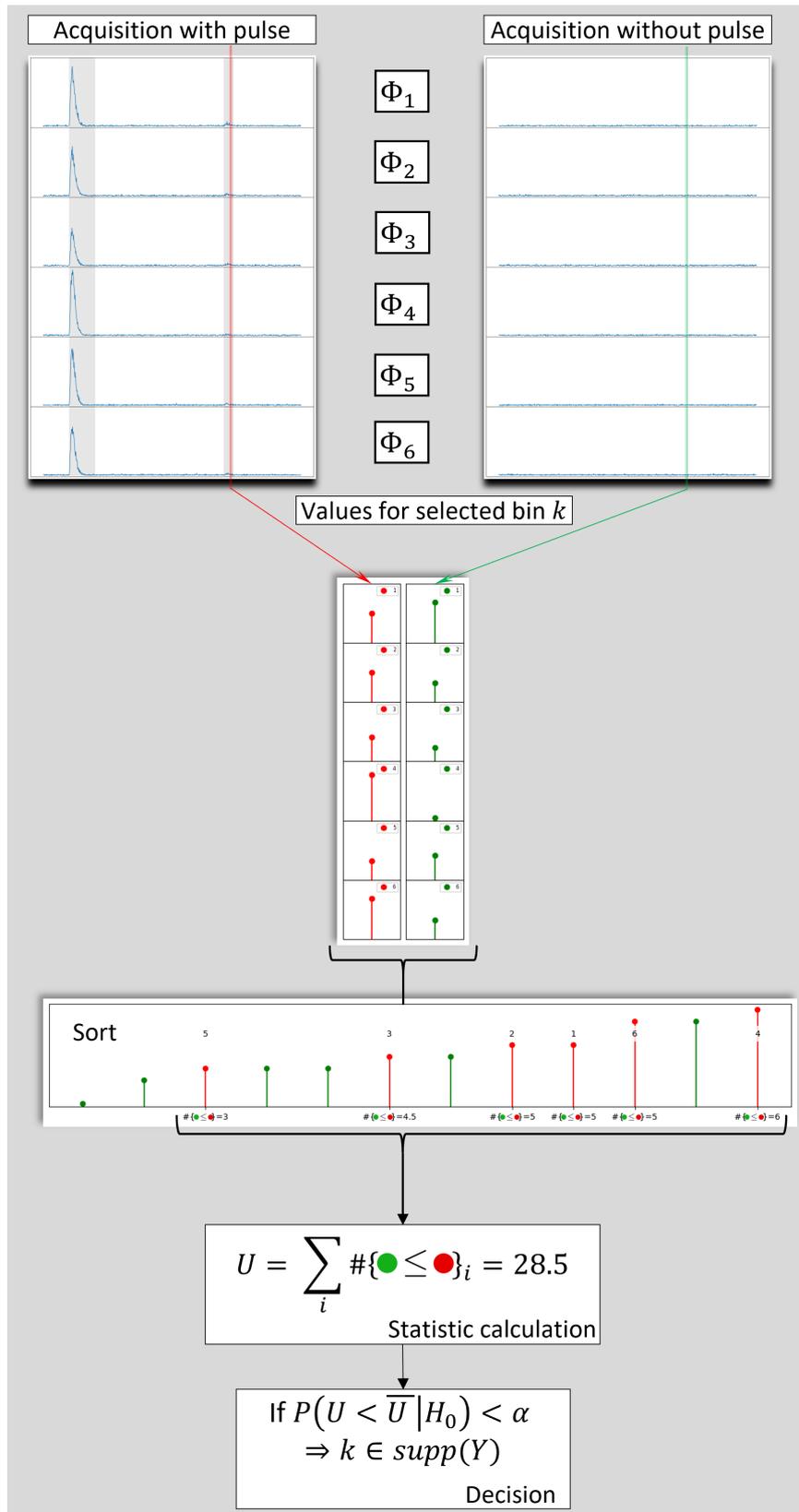


FIGURE 2.4. Illustration du test de rang de Mann-Whitney [95] dans CRISPS-3D

2.3 Algorithme de *Compressive Sensing*

Après la reconstruction par les deux premières étapes (Sections 2.2.1 et 2.2.2), chaque pas de temps est indépendant des autres et le bruit n'est présent que là où le signal est détecté. Comme expliqué dans la Section 1.3.2, on reconstruit la résolution latérale en résolvant un problème de *Compressive Sensing* à chaque pas de temps pour chaque pixel. En pratique, nous simulons un ensemble de pixels DMD imagés par un des pixels GmAPD du plan focal. La reconstruction est effectuée en considérant chaque pixel de la matrice GmAPD et le sous-ensemble de pixels DMD associé comme une configuration de caméra à un seul pixel. Pour chaque pixel de la caméra, nous résolvons un problème de *Compressive Sensing* pour chaque pas de temps temporel afin de reconstruire des images 2D locales que nous concaténons pour obtenir une image 3D globale.

Plus formellement, soit X_k l'image reconstruite vectorisée au pixel (i, j) et au pas de temps k , et Y_k la mesure correspondante. Soit Φ la base de mesure. En raison de la nature binaire du DMD, il existe une transformation linéaire permettant de convertir cette base en $\{0, 1\}$ binaire. Soit Ψ la base où X_k est K -parcimonieux (*i.e.* seules K composantes de X_k sont non nulles). Soit N le nombre de pixels DMD dans un groupe correspondant à la dimension de X_k . Nous pouvons exprimer le problème d'optimisation *Compressive Sensing* comme suit :

$$(\mathcal{P}_{CS}) \begin{cases} \min_{s \in \mathbb{R}^N} \|s\|_{l_0} \\ \text{s.t. } \|Y - \Phi\Psi s\|_2^2 \leq \varepsilon, \quad \varepsilon \geq 0 \end{cases} \quad (2.11)$$

où $s_k = \Psi X_k$ et ε est un terme d'erreur. Pour récupérer avec précision s_k à partir de Y_k puis de X_k , il est établi [29,32] que Φ doit suivre la $2K$ -RIP (*Restricted Isometry Property*). Cela signifie que s_k peut être déduit de Y_k , à condition que s_k soit K -sparse et Ψ soit hautement incohérent avec Φ . En pratique, la résolution exacte du problème avec la norme l_0 est connue pour être NP-difficile [107]. Par conséquent, l'utilisation d'une approche différente est nécessaire. Le problème d'optimisation peut être relaxé en utilisant la norme l_1 . Une autre approche consiste à approcher la solution avec un algorithme « glouton » [148]. Ici, nous choisissons cette dernière option avec l'OMP car elle fournit le meilleur rapport précision/performance (voir Section 1.3.1). Dans ce chapitre, le processus de mesure Φ est défini comme une transformée de Hadamard sous-échantillonnée et la base de parcimonie Ψ comme la transformée de Haar 2D. Ce choix est motivé par les nombreux exemples d'utilisation de matrices d'Hadamard avec succès dans la littérature single-pixel [55,67,78]. De même, le choix d'une base de Haar 2D est motivé par la forte parcimonie des images « naturelles » dans cette base [141] ainsi que sa forte incohérence asymptotique avec la base d'Hadamard [129]. Cependant, d'autres bases de mesure (*e.g.* Bernoulli($\frac{1}{2}$) [96]) ou de parcimonie (*e.g.* TV [89]) donnent de bons résultats. Notons qu'aucune comparaison rigoureuse de ces différentes métriques n'est volontairement présentée dans ce manuscrit.

2.3.1 Version optimisée : *Orthogonal Matching Pursuit* (OMP) utilisant l'approximation de Cholesky

L'Algorithme 1 suivant est une manière efficace d'implémenter l'OMP tel que défini dans [148].

Algorithm 1 Orthogonal Matching Pursuit

```

1: procedure OMP( $N \times d$  measurement matrix  $\Phi$ ,  $N - dim$  measurement vector
    $Y$ ,  $N \times N$  Sparsifying matrix  $\Psi$ )
2:    $r \leftarrow Y$ 
3:    $\Lambda \leftarrow \emptyset$ 
4:   while halting criterion not True do
5:      $\Lambda \leftarrow \Lambda \cup \left\{ \operatorname{argmax}_{j=1, \dots, d} |\langle r, (\Phi\Psi)_j \rangle| \right\}$ 
6:      $S \leftarrow \operatorname{argmin}_s \|Y - (\Phi\Psi)_{|\Lambda} s\|_2$ 
7:      $r \leftarrow Y - (\Phi\Psi)_{|\Lambda} S_{|\Lambda}$ 
8:   end while
9:   return  $\Psi S$ 
10: end procedure

```

Nous pouvons résoudre le problème de minimisation l_2 évalué à chaque étape de la boucle « *while* » de l'Algorithme 1 par la projection orthogonale $(\Phi_{|\Lambda}^T \Phi_{|\Lambda})^{-1} \Phi_{|\Lambda}^T$. Le calcul de la projection est accéléré par l'approximation de la matrice inverse en utilisant la décomposition QR ou de Cholesky. Cette stratégie a été proposée par Davis et *al.* ainsi que Rubinsten et *al.* [50, 132]. Cette implémentation est efficace lorsque la taille de la matrice est suffisamment faible ($< 10^2$, d'après Rubinsten et *al.* [132]). Dans le cas contraire des tailles de matrices plus importantes, une approximation de la projection doit être faite avec une méthode similaire comme l'algorithme Batch-OMP [132]. Nous avons obtenu une réduction du temps d'exécution de deux ordres de grandeur avec cette implémentation. Le problème se prête bien à une mise en œuvre parallèle, par exemple sur GPU, puisque nous exécutons notre algorithme indépendamment pour chaque pixel. Les capteurs disponibles dans le commerce ont une taille de matrices de 32 pixels \times 32 pixels voire 32 pixels \times 128 pixels, ce qui est compatible avec le nombre de blocs de la génération actuelle de GPU (Quadro RTX 6000 dans notre cas). Avec la version Cholesky de l'OMP, nous réduisons également la mémoire nécessaire par bloc par rapport à l'OMP classique. Ceci permet le calcul de la chaîne complète pour tous les pixels simultanément. On remarque que cette version de l'OMP est presque uniquement une multiplication vectorielle qui peut être optimisée pour travailler en parallèle en utilisant les différents threads dans chaque bloc. Lorsque l'on exécute notre chaîne de traitement CRISPS-3D uniquement en cuda/C++ sur GPU (Quadro RTX 6000 dans notre cas), on gagne encore deux ordres de grandeur en termes de temps d'exécution par rapport à la version Python sur CPU (Intel Xeon

W-2125). Un temps de traitement court permet une reconstruction à faible latence, ce qui est un avantage important pour nos applications futures, *e.g.* pour les applications de surveillance nécessitant d'être proche du temps réel pour le traitement de l'information et la détection d'anomalies.

2.4 Validation de la méthode sur des signaux simulés

La chaîne de détection complète est appliquée à plusieurs scènes simulées par le logiciel MATLIS disponible à l'ONERA [7, 47, 48] pour valider notre approche. Afin de couvrir la plupart des types d'objets réels imagés par un LiDAR 3D, nous sélectionnons quatre catégories représentatives : *(i)* les objets répartis sur les 2 dimensions latérales (surfaces planes à grandes sections transversales), *(ii)* les objets produisant des échos multiples dans un seul pixel, *(iii)* les objets répartis sur la dimension temporelle (surfaces inclinées) et *(iv)* les objets sous-pixel.

2.4.1 Reconstruction de l'onde complète

Les résultats obtenus après l'étape de repondération (voir Section 2.2.1) sont représentés sur les Figures 2.5 et 2.6. Les résultats de l'estimation du support et leurs comparaisons avec la vérité de base (voir Section 2.2.2) sont représentés sur la Figure 2.7 et résumés dans Tableau 2.1. Dans les Figures 2.5 à 2.7, Y_k et son support sont obtenus à partir de données artificiellement disponibles grâce à la simulation. Ces données ne sont pas accessibles à l'algorithme, car ces variables sont juste considérées pour les simulations directes dans le logiciel MATLIS. Néanmoins, nous stockons ces valeurs pour une exploitation future notamment pour comparer la vérité terrain avec les nuages de points reconstruits. \widehat{D}_k est la forme d'onde reconstruite, décorrélée dans le temps, telle que définie dans la Section 2.2.1. H_k est l'histogramme des temps de retour. Nous utilisons le rapport signal-bruit de pointe (PSNR défini par l'Équation (2.12)) pour évaluer la qualité de l'estimation de la forme d'onde (voir Figure 2.6).

$$PSNR_{Y_k}(\widehat{Y}_k) = 20 \log_{10} \left(\frac{\max_Y}{\sqrt{\frac{1}{N} \sum_{i=0}^N \|Y_{k,i} - \widehat{Y}_{k,i}\|_2^2}} \right) \quad (2.12)$$

Nous traitons 100 signaux pour mieux représenter les tendances de l'estimation. Nous calculons ensuite la moyenne du signal et un intervalle de confiance estimé à 95% pour chaque pas.

Nous sélectionnons quatre pixels d'intérêt pour illustrer les résultats sur différentes situations spécifiques :

- (A) Surface plane perpendiculaire à l'axe optique de la caméra, dont la taille est supérieure à celle d'un pixel de la caméra.
- (B) Surface plane perpendiculaire à l'axe optique de la caméra, d'une taille inférieure à un pixel DMD projeté.
- (C) Deux plans perpendiculaires à l'axe optique de la caméra, à deux distances différentes, chacun remplissant environ la moitié du champ de vision d'un pixel du capteur.
- (D) Une pente de 75° par rapport à l'axe optique de la caméra, remplissant tous les pixels.

Alors que le PSNR moyen de l'histogramme est de 64,6 avec une variance de 180,4, celui de l'estimation est de 71,3 avec une variance de 2,6. Ces deux PSNR sont calculés sur la scène synthétique (Section 1.4.2). L'erreur de forme d'onde estimée est moins sensible à la distribution spatiale des objets et au niveau du signal (voir Figure 2.5). Les erreurs d'estimation de la forme d'onde à haut (6×10^{-1}) et bas (6×10^{-3}) comptage de photons sont plus stables à l'erreur d'estimation à bas (6×10^{-3}) comptage de photons que pour l'histogramme. Cette méthode est sensible à la saturation : si le capteur sature (le nombre moyen de photons est proche de l'unité) alors l'estimation tend à être instable en raison du terme exponentiel de la formule de reconstruction. Si une mesure du nombre moyen de photons proche de l'unité est nécessaire, alors une méthode plus complexe et exigeante en calcul comme celle définie par Heide et *al.* [71] ou Rapp et *al.* [121] doit être utilisée pour remplacer cette étape.

La forme d'onde estimée est moins corrélée dans le temps, ce qui est l'objectif principal de cette étape. Cet objectif est atteint dans la plupart des cas, sauf dans le second où les niveaux de signal sont trop faibles pour observer l'effet du temps mort. L'erreur augmente avec le temps alors que le signal diminue. Cet effet est illustré dans le cas (D). Si le niveau du signal est suffisamment faible, les pas de temps suivants ne seront pas fortement affectés, comme dans le cas (B). La réponse temporelle est corrélée à la longueur de l'impulsion. Des impulsions longues par rapport à la longueur du pas de temps provoqueront une distorsion temporelle. Dans ce cas, l'ajout d'une étape de déconvolution utilisant les connaissances disponibles sur l'impulsion doit être envisagé. Ici, nous ne considérons que les impulsions courtes, dont la largeur à mi-hauteur est inférieure à la moitié de la taille du pas de temps temporel où cet effet peut être négligé.

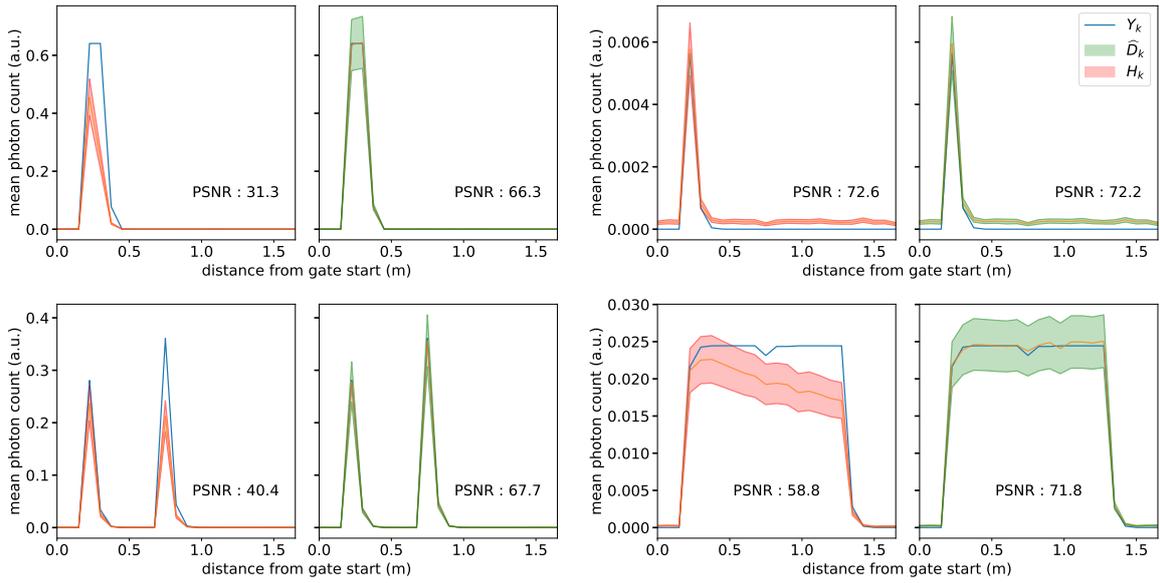


FIGURE 2.5. Comparaison de l'histogramme moyen et de la forme d'onde moyenne reconstruite (Section 2.2.1) à partir de 100 signaux simulés. Chaque sous-image correspond à un pixel différent de la caméra GmAPD. Le bleu représente la vérité terrain des signaux, le rouge la moyenne des histogrammes et son intervalle de confiance à 95% et le vert la moyenne des signaux estimés et son intervalle de confiance à 95%. La porte distance commence à 13 km de la caméra.

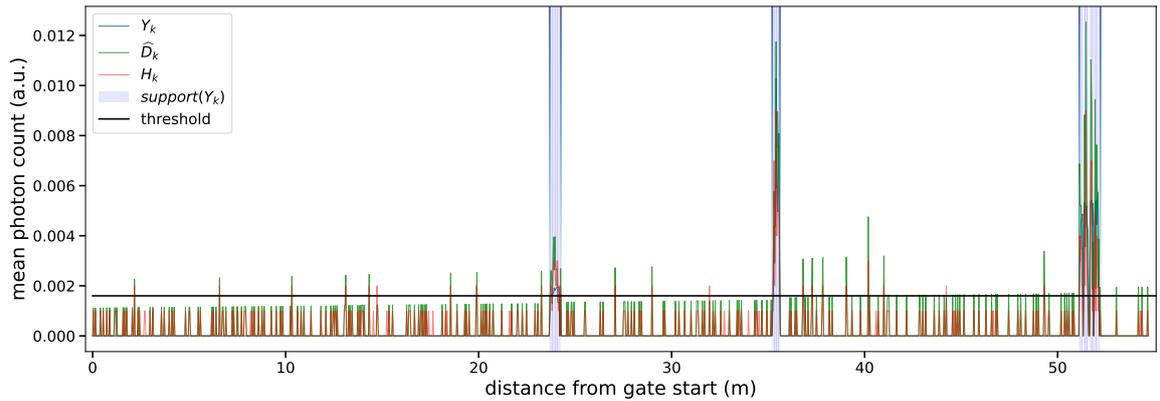


FIGURE 2.6. Illustration de l'effet du bruit sur un signal à longue portée. Les courbes bleue, rouge et verte représentent respectivement la forme d'onde réelle Y_k , le signal estimé \hat{D}_k tel que défini dans la Section 2.2.1 et l'histogramme H_k . La porte temporelle commence à 10 km de la caméra.

Un nouveau cas avec des cibles multiples le long de la ligne de visée et une longueur de porte accrue est considéré pour illustrer l'effet du bruit sur l'estimation du support. La Figure 2.6 met en évidence un inconvénient de notre méthode : le bruit augmente le long de la porte. Dans ce cas, il est particulièrement difficile de distinguer le signal du bruit. Par exemple, le signal à 23 m est au même niveau que le bruit à 40 m. L'utilisation d'un simple seuil est insuffisante pour assurer à la fois un faible taux de fausses alarmes et un bon taux de détection.

2.4.2 Estimation de la fonction du support

Le support de fonction estimé est comparé à la vérité terrain et à la méthode basée sur l'histogramme dans la Figure 2.7 pour les cas (B), (C) et (D).

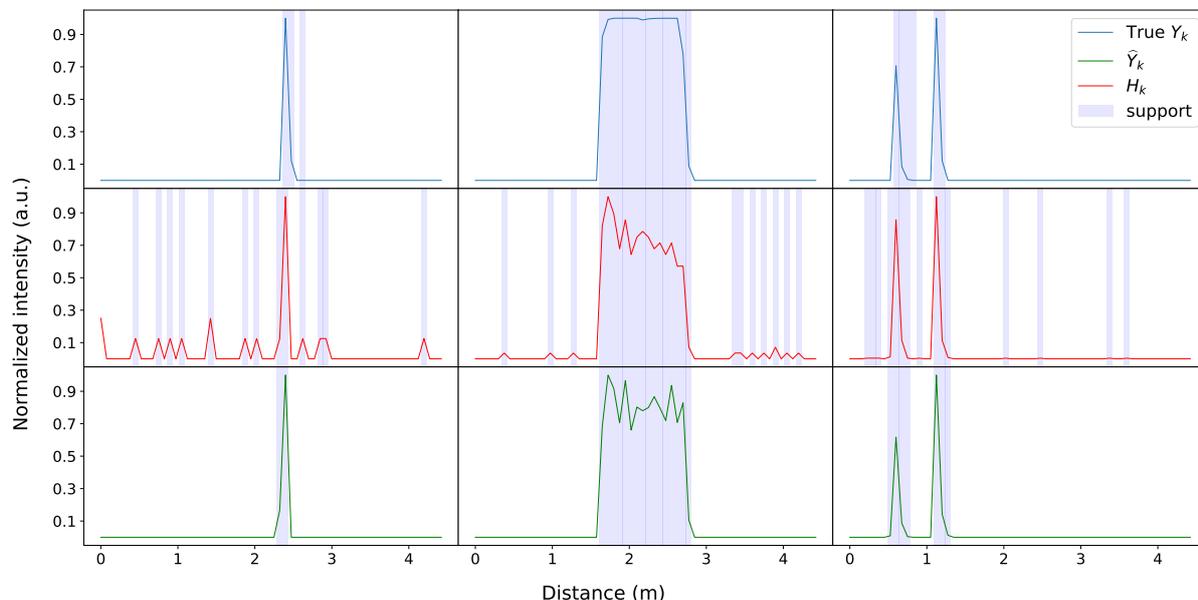


FIGURE 2.7. De haut en bas : la « vérité terrain » Y_k , l'histogramme brut H_k et \hat{Y}_k le support estimé à l'aide du test défini dans la Section 2.2.2 appliqué à la forme d'onde estimée \hat{D}_k . Les pas de temps remplis se trouvent dans le support de la fonction. Chaque colonne représente le signal traité pour différents pixels GmAPD. De gauche à droite : cas (B), (C) et (D). La normalisation *Min-Max*, par colonne, est appliquée aux trois signaux. La porte temporelle commence à 13 km de la caméra.

Chaque point de la forme d'onde estimée est considéré comme une observation statistique, étiquetée « in » ou « out » par rapport à la vérité terrain. Ainsi, nous construisons la matrice de confusion (voir Tableau 2.1) pour évaluer la qualité de l'estimation. Dans la plupart des cas, le support est correctement reconstruit avec une légère tendance à l'élargissement autour des pics.

L'estimation du support par cette méthode augmente le nombre de vrais positifs/négatifs et diminue le nombre de faux positifs/négatifs. Aucun seuil absolu n'est nécessaire. Seul un niveau de signification pour le test doit être fixé. Arbitrairement, nous l'avons fixé à 0,1%. Notre méthode fournit une meilleure estimation du support que la méthode classique de l'histogramme, avec un nombre plus élevé de vrais positifs et plus faible de faux négatifs. Les statistiques dans la Tableau 2.1 sont faites pour tous les pixels, tous les pas de temps et tous les motifs. Notre méthode offre une plus grande confiance statistique que l'évaluation du support pour chaque motif considéré séparément les uns des autres.

Predicted \ True	out	in
out	216464	53181
in	393	1322

(a)

Predicted \ True	out	in
out	263652	5992
in	565	1151

(b)

Predicted \ True	out	in
out	269274	371
in	164	1551

(c)

TABLE 2.1. Comparaison des matrices de confusion, (a) est la matrice de confusion de $support(H_k)$ par rapport à $support(Y_k)$; (b) est la matrice de confusion de $support(H_k)$ utilisant un seuil par rapport à $support(Y_k)$ et (c) est la matrice de confusion de $support(\hat{Y}_k)$ par rapport à $support(Y_k)$. La teinte verte correspond aux vrais positifs/négatifs et la rouge aux faux positifs/négatifs. La saturation diminue avec la qualité de l'estimation.

Nous appliquons cette étape au cas présenté dans la Figure 2.6. La Figure 2.8 illustre que notre approche fonctionne bien même avec un rapport signal/bruit faible. Ici, les supports à 23 m et 40 m sont correctement évalués.

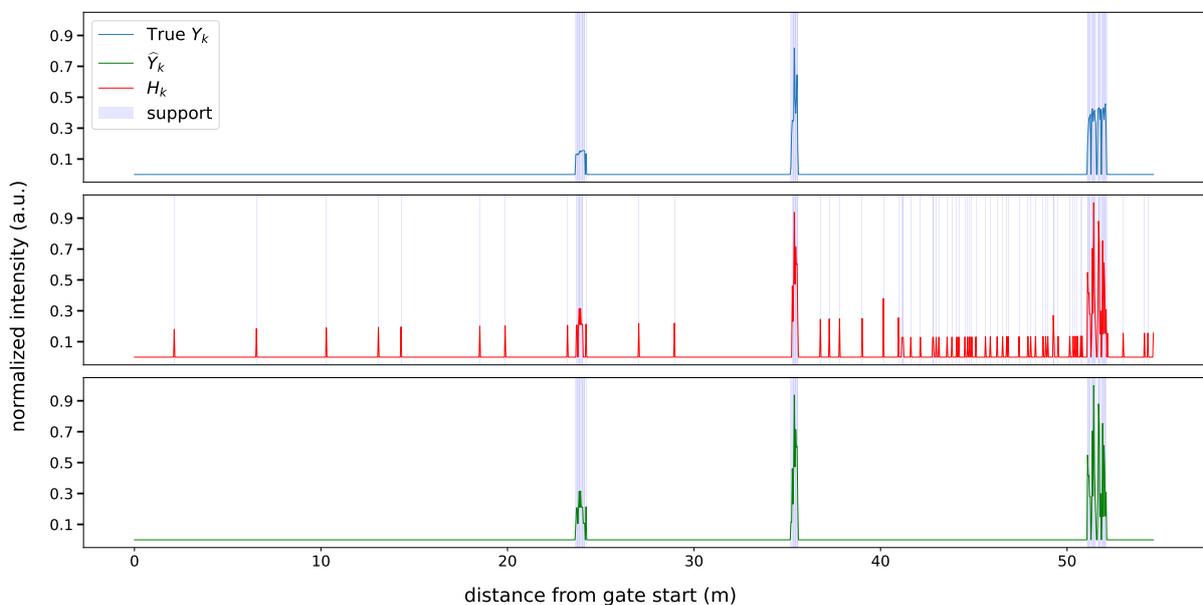


FIGURE 2.8. De haut en bas : « vérité terrain » Y_k , forme d'onde estimée \hat{D}_k avec seuillage, puis support estimé \hat{Y}_k appliqué à la forme d'onde estimée \hat{D}_k . Les pas de temps remplis sont dans le support de la fonction. Le signal est normalisé avec la normalisation *Min-Max*. La porte temporelle commence à 10 km de la caméra.

2.4.3 Chaîne de traitement complète

Les résultats de la chaîne de traitement complète ont été présentés dans plusieurs conférence [153, 155]. Comme le montre la Figure 2.9, CRISPS-3D permet de reconstruire de nombreux détails de la scène par rapport aux données brutes en augmentant le nombre de points. Il est néanmoins possible de remarquer des artefacts de reconstruction (trous, points fantômes) provenant de la résolution du problème d'optimisation du Compressive

Sensing. Les fréquences spatiales les plus élevées (mire éventail) et les surfaces planes dont la section transversale est inférieure à la taille d'un pixel projeté du DMD ne sont pas récupérées (scène synthétique). Notre méthode présente de nets avantages pour la récupération des détails en reconstruisant des bords mieux définis comme dans les cas (A), (C) et (D) (voir Figures 2.9a et 2.9b) ou les lettres « ONERA » (voir Figures 2.9c et 2.9d). Un autre avantage concerne les objets à haute fréquence spatiale (voir devise française) et avec des variations de profondeur (voir buste de Marianne). Ces objets sont identifiables dans les Figures 2.9b et 2.9d alors qu'ils ne l'étaient pas dans les Figures 2.9a et 2.9c. La résolution latérale est augmentée d'un facteur 8 pour chaque direction latérale (Figure 2.9), la résolution en profondeur reste inchangée.

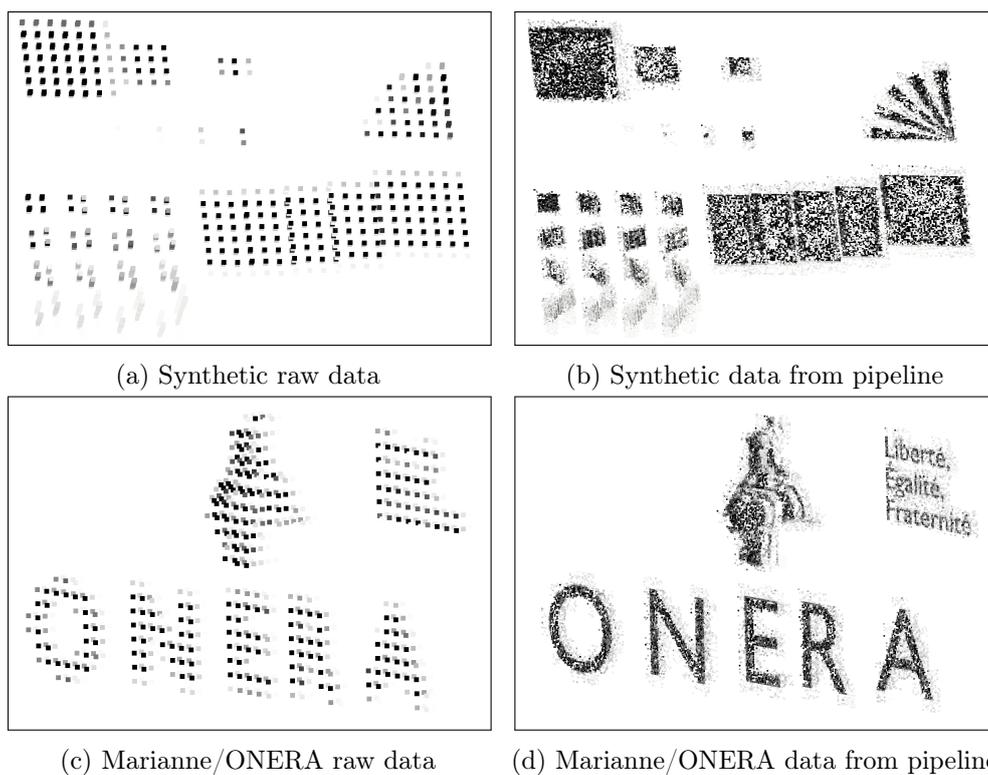


FIGURE 2.9. Résultats obtenus après : (a) et (c) une acquisition LiDAR avec seulement les points labélisés comme provenant d'une impulsion laser ; (b) et (d) une acquisition *Compressive Sensing* avec tous les motifs, traitée avec CRISPS-3D, sans aucun post-traitement supplémentaire. L'échelle de gris représente l'intensité normalisée de la chaîne de traitement complète.

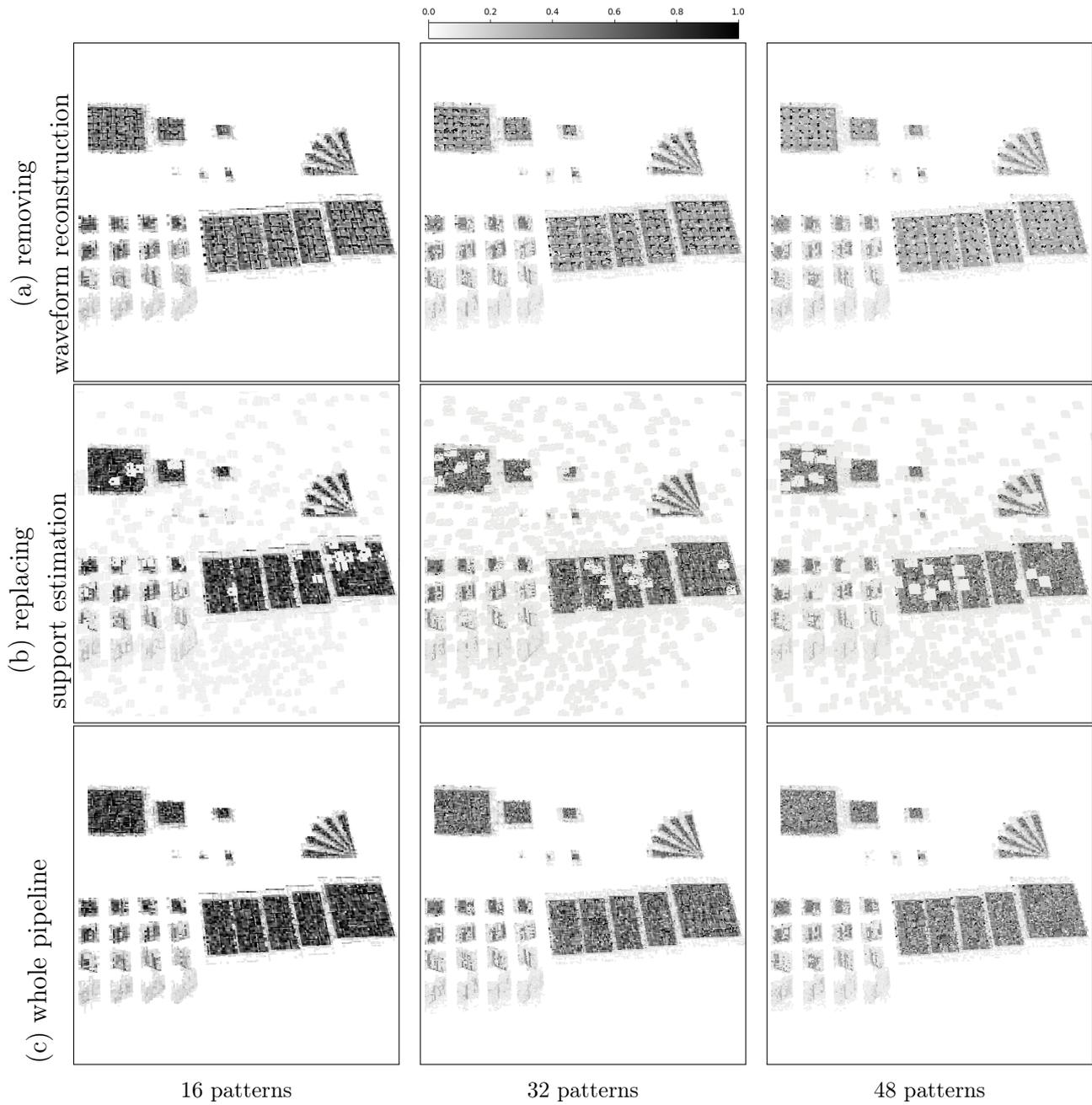


FIGURE 2.10. De haut en bas, résultats obtenus après : suppression de l'étape de reconstruction de la forme d'onde, remplacement de l'étape d'estimation du support par un seuil choisi pour éviter plus de pertes pour la reconstruction du signal et, utilisation de CRISPS-3D. De gauche à droite, résultats en utilisant : la scène synthétique avec 16, 32, 48 mesures de motifs. L'échelle de gris représente l'intensité normalisée par la chaîne de traitement complète pour une scène donnée.



FIGURE 2.11. De haut en bas, résultat obtenu après : (a) suppression de l'étape de reconstruction de la forme d'onde, (b) remplacement de l'étape d'estimation du support par un seuil choisi pour éviter plus de pertes pour la récupération du signal et, (c) utilisation de l'ensemble de la chaîne de traitement. De gauche à droite, résultat utilisant la scène ONERA avec 16, 32, 48 mesures de motifs. L'échelle de gris représente l'intensité normalisée par le pipeline pour une scène donnée.

Dans les Figures 2.10 et 2.11, nous étudions la variation de la qualité du processus de reconstruction en supprimant différentes étapes de CRISPS-3D sur les différentes scènes et différents nombres de mesures de motifs. Sans la reconstruction de la forme d'onde, l'intensité globale diminue et des artefacts supplémentaires apparaissent. Ces artefacts dépendent de la parcimonie et de la base de mesure. Dans la deuxième ligne, un seuil

optimal a été défini pour limiter la perte de signal avant la reconstruction par *Compressive Sensing* et ne permet pas de supprimer toutes les acquisitions dues aux bruits. Comme l'algorithme *Compressive Sensing* classique, notre algorithme permet une acquisition d'images 3D avec une résolution latérale accrue en utilisant seulement quelques mesures compressées. On peut remarquer que la qualité diminue plus rapidement par rapport au taux de compression que les autres algorithmes classiques de CS [55]. La diminution de la qualité avec le taux de compression ne diminue pas linéairement avec la taille de l'image [129]. Cependant, pour souligner l'intérêt de notre méthode, il suffit de 16 mesures de formes d'ondes complètes avec un réseau GmAPD de $32 \text{ pixels} \times 32 \text{ pixels}$ pour reconstruire une image 3D ayant une résolution latérale de $256 \text{ pixels} \times 256 \text{ pixels}$. De plus, l'utilisation de CRISPS-3D permet de limiter les artefacts indésirables issus de la récupération de *Compressive Sensing* tout en effectuant un débruitage efficace.

2.5 Synthèse

Ce chapitre décrit la chaîne de traitement complète CRISPS-3D qui permet d'augmenter la résolution latérale des systèmes LiDAR 3D à base de capteurs GmAPD en utilisant une approche de *Compressive Sensing*. CRISPS-3D tient compte de la nature du système en utilisant une approche statistique pour débruiter et estimer la forme d'onde corrigée avant de résoudre le problème d'optimisation *Compressive Sensing*. La faisabilité de notre méthode est démontrée numériquement sur des données simulées. CRISPS-3D a été validé uniquement sur des cibles statiques. Dans notre cas, la résolution latérale est augmentée par un facteur 64 tout en préservant la résolution en profondeur et en minimisant les artefacts. Cette chaîne de traitement complète est adaptée pour des acquisitions de jour (également de nuit) à très longue portée et potentiellement embarquée. Cependant, CRISPS-3D ne permet pas la reconstruction du signal obtenu avec une impulsion dont la largeur temporelle est supérieure à la taille d'un pas de temps. On verra dans le chapitre suivant comment étendre cette limite.

Chapitre 3

Prise en compte dans CRISPS-3D d'impulsions dont la longueur temporelle est supérieure à la taille d'un pas de temps

Points clefs

- Tel que formulé dans le chapitre 2, notre algorithme CRISPS-3D est défini pour des impulsions laser courtes ($l_{fwhm} \ll \delta_t$) et des niveaux de signal allant jusqu'à 0,8 photon par image et par pixel en moyenne.
- Sans modification de la distance d'acquisition (10 km à 20 km) et de l'énergie par impulsion laser, CRISPS-3D permet en déconvoluant le signal de prendre en compte des impulsions plus longues (jusqu'à 2 ns).
- La limite sur la longueur de l'impulsion n'est pas induite par la déconvolution en elle-même mais par la méthode de correction du DTE.

Contributions principales

- Identification du domaine de validité de CRISPS-3D, explication des limitations actuelles et potentiel de la chaîne de traitement après déconvolution.
- Définition d'un algorithme d'estimation de profondeur exploitant la forte parcimonie des acquisitions LiDAR par déconvolution. L'algorithme est spécifiquement adapté à notre chaîne de traitement CRISPS-3D.
- Proposition d'amélioration de l'algorithme de déconvolution pour l'estimation de profondeur par régularisation temporelle.

L'objectif de ce chapitre est de développer une méthode pour prendre en compte des impulsions dont la largeur temporelle est supérieure à la durée d'un pas de temps. En effet, cette limitation combinée avec les autres contraintes (c.f. Section 1.1.1) peut fortement réduire le nombre de sources laser admissibles pour l'acquisition, en particulier vis-à-vis de contraintes d'embarquabilités. Le plus souvent, il est difficile de maintenir un faible SWaP-C, une énergie d'impulsion et une fréquence de répétition constantes tout en diminuant la largeur temporelle d'une impulsion à des durées sub-nanosecondes.

3.1 Utilisation d'impulsions larges avec déconvolution

Dans cette section on cherche à lever la limitation sur la largeur des impulsions utilisées.

3.1.1 Motivation et contexte

Dans un système LiDAR 3D, en mesurant le temps de retour d'une impulsion, la distance de l'objet au capteur peut être déduite (voir Section 1.1). Par nature les LiDAR 3D permettent d'acquérir des surfaces représentées par des nuages de points (discrets) en 3D, parfois assimilées à des *2D-manifolds* (variétés 2D) dans un espace 3D. De ce fait, il est souhaitable que le nuage de points reconstruit représente un ensemble de surfaces et ne soit pas dense. Cependant, du fait de la convolution entre l'impulsion et les surfaces dans la scène, plus une impulsion est longue plus les surfaces sont « épaisses » (voir Figure 3.1). Malgré une bonne résolution temporelle des capteurs GmAPD, une impulsion trop longue ne permet pas toujours de distinguer deux objets pour des pas de temps voisins. Ainsi, l'intérêt principal de l'estimation de la profondeur est de maintenir la résolution temporelle du système en gardant une bonne fidélité à la scène de par les mesures.

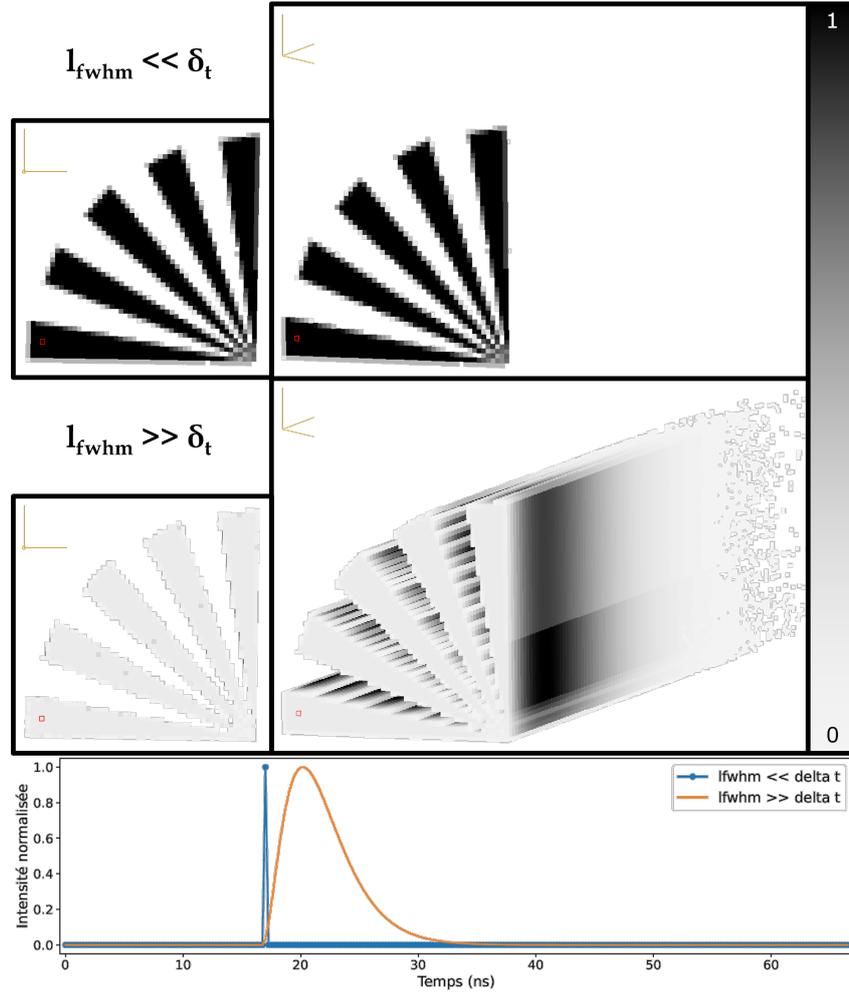


FIGURE 3.1. Exemples de reconstruction d'un quart de mire de Siemens avec deux largeurs d'impulsions différentes (de haut en bas $l_{fwhm} \ll \delta_t$ puis $l_{fwhm} \gg \delta_t$). Les intensités sont normalisées sur chaque scène, avec à gauche une vue de dessus et une vue oblique à droite, l'orientation des axes est donnée en haut à gauche de chaque point de vue. Les courbes représentent la distribution des intensités au cours du temps pour le pixel coloré en rouge.

3.1.2 Définition du problème de déconvolution avec *Compressive Sensing*

Dans cette section, il est supposé que le DTE peut-être efficacement compensé. Avant de s'intéresser à la façon de résoudre ce problème d'estimation de profondeur, il faut le définir rigoureusement. Pour cela, le problème à résoudre est décliné de l'équation lidar abordée dans la Section 1.1.2. La puissance collectée sur le capteur (i, j) pour le motif m au pas de temps n , est défini par :

$$P_{i,j,n}^m = \sum_{k=0}^{N_{DMD}} (\Phi_{i,j})_{m,k} \frac{2A_0 C_a \xi_{\lambda_l}}{c} \int_0^\infty \int_0^{\delta_t} P_E(n\delta_t + t - T) dt \frac{O_{i,j,k}(\frac{cT}{2})}{A_L(\frac{cT}{2})T^2} dT$$

L'information $O_{i,j,k}$ souhaitée sur la scène est convoluée temporellement à l'impulsion laser. Afin de résoudre le problème de déconvolution, certaines hypothèses sont faites :

- La forme de l'impulsion laser P_E est connue par étalonnage.
- Par nature de l'acquisition LiDAR 3D, $O_{i,j,k}$ est fortement parcimonieuse en profondeur.

Cela nous permet d'inverser la convolution entre l'impulsion et les objets pour récupérer des informations en profondeur plus précises. Nous posons le problème d'estimation de profondeur de la façon suivante :

$$X_{i,j,\cdot,k}(n) = \sum_{l=0}^{N_{bin}} \frac{2A_0 C_a \xi_{\lambda_l}}{c} \int_{n\delta_t}^{(n+1)\delta_t} P_E(t - l\delta_t) dt O_{i,j,k}^A(l\delta_t)$$

On approxime $\int_{(n-l)\delta_t}^{(n-l+1)\delta_t} P_E(t) dt$ par $P_{tot} h_{n,l} = P_E((n-l)\delta_t) \approx \int_{(n-l)\delta_t}^{(n-l+1)\delta_t} P_E(t) dt$ où $\sum_n h_{n,l} = 1$, $P_{tot} = \int P_E(t) dt$ et $\delta_t \ll l_{fhw}$, ainsi :

$$X_{i,j,\cdot,k}(n) = \frac{2A_0 C_a \xi_{\lambda_l} P_{tot}}{c} \sum_{l=0}^{N_{bin}} h_{n,l} \times O_{i,j,k}^A(l\delta_t) \quad (3.1)$$

$$X_{i,j,\cdot,k}(n) \propto \sum_{l=0}^{N_{bin}} h_{n,l} \times O_{i,j,k}^A(l\delta_t) \quad (3.2)$$

$$X_{i,j,\cdot,k} \propto h \times O_{i,j,k}^A \quad (3.3)$$

Où $X_{i,j,\cdot,k} \in \mathbb{R}^{N_{bin}}$ représente la reconstruction par *Compressive Sensing* à chaque pas de temps sur le pixel du DMD numéro k et du pixel (i, j) du plan focal, $h \in \mathbb{R}^{N_{bin} \times N_{bin}}$ représente la forme de l'impulsion laser émise, normalisée et décalée d'un pas de temps entre chaque colonne et $O_{i,j,k}^A \in \mathbb{R}^{N_{bin}}$ la variable d'intérêt.

3.1.3 Solutions existantes au problème de déconvolution parcimonieuse dans le cadre d'acquisitions LiDAR 3D

Il n'existe pas de solutions dans la littérature explicitant comment déconvoluer un signal provenant d'une acquisition LiDAR 3D avec *Compressive Sensing* et prenant en compte le DTE. Cependant, certaines méthodes peuvent-être extrapolées en modifiant leur cadre d'application.

Dans le cas où le *Compressive Sensing* n'est pas considéré et que le DTE peut être ignoré, il existe un certain nombre de méthodes traitant d'ores et déjà du sujet pour les LiDAR à retour d'onde complète.

Dans le cadre de la déconvolution en traitement d’image, deux algorithmes sont utilisés en majorité, le *Filtre de Wiener* [162] et l’algorithme de *Richardson-Lucy* [123]. Le filtre de Wiener déconvolue un signal en estimant un filtre prenant en compte l’impact du bruit sur le signal, l’idée étant de déconvoluer le signal en divisant la transformée de Fourier du signal par celle du filtre sur les fréquences où le SNR est bon. Cette méthode est efficace si le bruit et le signal sont indépendants (*e.g.* bruit gaussien additif). L’algorithme de Richardson-Lucy construit un algorithme itératif basé sur une interprétation probabiliste de la convolution, où le filtre donne la distribution que suit le signal d’origine pour obtenir celui d’arrivé.

Matched filtering est un des algorithmes les plus utilisés pour estimer la distance tout en débruitant le signal. Le principe étant d’évaluer la corrélation croisée entre le signal reçu et la forme de l’impulsion ou son logarithme pour garder soit uniquement le maximum soit l’ensemble des pas de temps au-dessus d’un certain seuil. Cette méthode est assez populaire [6, 98] du fait de sa faible complexité algorithmique $O(N_{bin} \log(N_{bin}))$.

En supposant que la forme de l’impulsion soit une loi normale dont les paramètres sont connus, Mallet *et al.* [94] ont identifié trois méthodes dans leur état-de-l’art en 2009 : (i) Moindres carrées non-linéaires avec l’algorithme de Levenberg–Marquardt [75], (ii) Estimateur du maximum de vraisemblance avec l’algorithme d’expectation-maximisation [116] et (iii) la méthode de Reversible Jump Monte Carlo Markov Chain (RJCMCMC) [39, 72] (voir [94] pour plus de détails).

La méthode de RJCMCMC n’est pas limitée aux formes d’impulsions gaussiennes. Cette approche a notamment été utilisée pour estimer l’intensité et la profondeur d’acquisitions de systèmes LiDAR dotés de caméras à comptage de photons [6, 70, 145]. Cette extension des méthodes dites de *Monte Carlo Markov Chain* (MCMC), basée sur le formalisme bayésien, permet d’inférer les paramètres (positions et intensités des points du nuages) dans un espace de dimension variable (nombres de points reconstruits variables).

Le problème de déconvolution est mal-conditionné. De faibles modifications de la mesure entraînent de fortes modifications sur la reconstruction [23]. Une idée pour pallier le mauvais conditionnement du problème de déconvolution (Équation (3.1)) est de le régulariser pour accroître la stabilité de la solution [56]. Les acquisitions LiDAR étant assimilables à des surfaces (*2D-manifolds*) en 3D, O^A est particulièrement parcimonieux. Prendre en compte l’hypothèse de parcimonie sur $O_{i,j,k}$ dans l’Équation (3.1) permet de retrouver un problème de *Matching Pursuit* (MP) :

$$\begin{cases} \min_{O \in \mathbb{R}^{N_{bin}}} \|O\|_{l_0} \\ s.c. \|X_{i,j,k} - hO\|_2^2 \leq \varepsilon, \varepsilon \geq 0 \end{cases} \quad (3.4)$$

Cette idée de prendre en compte la parcimonie a déjà été exploitée dans plusieurs travaux, notamment dans Manipop [145], RT3D [144] et par Shin et al. [137].

Dans Manipop, un modèle Bayésien est défini en prenant en compte les corrélations spatiales avec des points et en favorisant la parcimonie. Ensuite le *Maximum A Posteriori* (MAP) est estimé en utilisant une méthode de RJMCMC modifiée spécifiquement pour le modèle.

Suite à Manipop, une version optimisée nommée RT3D a été développée [144]. Elle est basée sur un algorithme de *plug-and-play* [152] utilisant des algorithmes de débruitage pour régulariser l’optimisation. RT3D obtient des résultats similaires à Manipop mais en gagnant 2 à 3 ordres de grandeur sur le temps de calcul. La solution du problème de MP sert d’initialisation à l’algorithme de débruitage en profondeur. La distance est ensuite débruitée en utilisant un *a priori* local de régularité spatiale sur la profondeur. Pour cela, des sphères algébriques sont ajustées localement et les points du nuage sont recalés sur ces sphères (APSS [103, 144]).

De manière similaire, Shin et al. [137] utilisent cette méthode mais avec du *Non-Negative Orthogonal Matching Pursuit* (NNOMP) afin de détecter des regroupements d’objets en profondeur dans des nuages de points 3D avant de minimiser la TV localement pour débruiter la reconstruction. Ces méthodes exploitant la parcimonie utilisent aussi des APD mais n’augmentent pas la résolution en utilisant du *Compressive Sensing* et ne prennent pas en compte le DTE.

Bien que les méthodes présentées dans cette section fonctionnent, il nous faut maintenant prendre en compte l’application du *Compressive Sensing* ainsi que la correction du DTE. Certains algorithmes [71, 121] présentés dans la Section 2.2.1 permettent de prendre en compte le DTE et d’estimer la position en profondeur mais n’appliquent pas de *Compressive Sensing* et n’augmentent pas la résolution latérale du capteur.

3.1.4 Limitation des durées d’impulsions possibles due au DTE

Lorsque les impulsions deviennent trop longues, la correction du DTE présentée dans le chapitre 2 n’est plus suffisante et entraîne trop d’erreurs pour pouvoir reconstruire correctement le nuage de points. En pratique, nous avons pu vérifier que si la somme cumulée de l’histogramme sur une colonne est trop proche de 1, cela empêche de reconstruire le signal complet (voir Figure 3.2).

Lorsque la somme cumulée approche 1, la réduction du signal due à l'effet du DTE est telle que plus aucun autre signal ou presque n'est détecté. En utilisant des impulsions plus longues, le capteur peut ne plus détecter la fin de la queue d'une impulsion réfléchiée. Cette impulsion tronquée induit de fortes erreurs dans les algorithmes de déconvolution qui utilisent une correspondance avec la forme complète de l'impulsion pour estimer la profondeur.

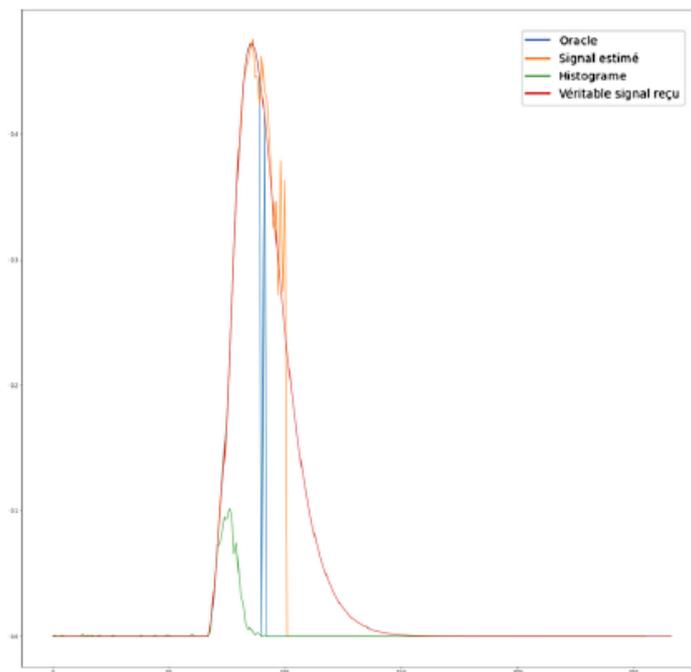


FIGURE 3.2. Exemple de cas dans lequel la correction du DTE n'est plus suffisante, l'Oracle correspond au signal détectable par la caméra pour le motif avec tous les miroirs tournés vers la caméra.

En reprenant l'équation utilisée pour corriger le DTE (voir ANNEXE A), l'instabilité de la reconstruction est prévisible lorsque la somme cumulée de l'histogramme approche 1. En effet, lorsque cette somme tend vers 1, l'estimateur du signal tend vers 0 et il n'est plus défini lorsque la somme cumulée atteint 1 (division par 0).

De plus, du fait de la différence entre le nombre de pixels du DMD en position « on » sur la première modulation et les autres (2 fois plus), le signal est tronqué plus tôt (différence entre l'oracle et le signal estimé sur la Figure 3.2). Ce décalage induit des erreurs de reconstruction dès l'étape du *Compressive Sensing* pour tous les pas de temps après que le signal estimé soit tronqué.

Le domaine d'application de cette méthode est limité par l'énergie totale par impulsion. Dans notre cas, en conservant le même niveau d'énergie et la même distance, le nuage de points peut être reconstruit sans perte avec des durées d'impulsions allant jusqu'à 2 ns

(empiriquement) contre 0,25 ns dans le chapitre 2. Ainsi, la limitation sur la longueur de l'impulsion induite par le DTE dépend de l'énergie totale reçue par impulsion. Cette énergie varie en fonction des paramètres du système d'acquisitions considéré, et des propriétés de la scène comme défini dans l'équation LiDAR, chapitre 1.

De ce fait, l'estimation de profondeur par déconvolution parcimonieuse permet d'étendre le domaine d'application de notre algorithme CRISPS-3D à des impulsions plus longues mais reste limité par le DTE. Pour aller plus loin, et étendre le domaine d'application de CRISPS-3D à des formes d'impulsions plus longues que 2 ns, il faut considérer des algorithmes avec plus d'*a priori* comme ceux mis en place par Rapp et al. ou Heide et al. [71, 121] et corriger le DTE tout en utilisant la forme de l'impulsion (potentiellement tronquée de par la nature du capteur).

Maintenant que le domaine d'application de CRISPS-3D pour la longueur des impulsions est identifié, nous définissons un algorithme pour déconvoluer le signal.

3.2 Définition d'un algorithme de déconvolution pour CRISPS-3D

3.2.1 Déconvolution *via* une approximation parcimonieuse

Application de l'estimation de profondeur dans CRISPS-3D

L'objectif du chapitre est d'utiliser la connaissance sur la forme de l'impulsion pour estimer la profondeur avec du *Compressive Sensing* dans CRISPS-3D. Pour cela plusieurs stratégies ont été envisagées. Elles dépendent notamment de la position de l'étape d'estimation de la profondeur dans CRISPS-3D par rapport à celle du *Compressive Sensing*.

Avant Compressive Sensing

L'avantage majeur de déconvoluer le signal avant le *Compressive Sensing* est que le support temporel du signal peut être réduit fortement pour chaque motif. Une réduction du support implique une diminution du nombre d'images 2D à reconstruire avec le *Compressive Sensing* pour former l'image 3D finale. En revanche, en raison de la parcimonie de O^A en profondeur et du fait que le *Compressive Sensing* soit appliqué à chaque pas de temps, une faible erreur de position pour un motif donné peut engendrer de fortes instabilités lors de la reconstruction.

Les méthodes existantes qui permettent d'estimer la position sont soit trop imprécises (corrélation croisée [98]) soit trop lentes (en même temps que le DTE [121]) pour notre application. En revanche, le nombre d'itérations de l'algorithme est réduit au nombre de pixels de la caméra multiplié par le nombre de motifs.

Pendant *Compressive Sensing*

L'idée ici est inspirée des travaux de Chen et *al.* [41] et Bogdan et *al.* [129] qui ont intégré avec succès une fonction d'étalement du point ou *Point Spread Function* (PSF) dans l'algorithme de CS. Dans notre cas, l'idée est que l'impulsion laser se comporte comme un filtre convolué au signal et donc par analogie comme une PSF. Pour reconstruire le signal, le problème de déconvolution peut être formulé comme un problème de reconstruction directement en 3D avec une PSF spécifique s'appliquant uniquement dans la profondeur. Actuellement, c'est la solution qui semble est la plus stable numériquement. Résoudre le problème de CS directement en 3D n'a pas été effectué jusque-là. Cependant les temps d'exécution de cette classe d'algorithmes peuvent-être élevés. En effet, les temps d'exécutions augmentent avec le nombre de voxels à évaluer et diminuent avec la parcimonie. Cependant la parcimonie étant plus importante en profondeur, la vitesse d'exécution est incertaine. De plus, l'utilisation de la PSF ne garantit pas nécessairement le maintien de l'incohérence entre la base de parcimonie et la base de mesure. Comme expliqué dans la Section 1.3, avoir une forte incohérence est nécessaire pour assurer la reconstruction exacte d'un signal s'il est non bruité et à minima borné par la norme du bruit sinon.

Après *Compressive Sensing*

L'estimation de la profondeur des objets de la scène *a posteriori* est intéressante notamment car elle permet l'application de méthodes déjà existantes (Section 3.1.3). En revanche, elle a deux limitations principales : (i) elle nécessite d'être appliquée une fois pour chaque pixel de la caméra et (ii) la qualité de l'estimation varie en fonction de la qualité de la reconstruction du *Compressive Sensing*. Cependant, le nombre d'itérations nécessaires pour cette méthode est notamment compensé par la plus forte parcimonie qui peut être utilisée pour accélérer certains algorithmes (*c.f.* Section 1.3.1). De plus, dans ce cas, nous nous ramenons au cadre classique de l'estimation de la profondeur de données LiDAR 3D (ici les intensités étant localement obtenues avec du *Compressive Sensing*). Bien que dépendante de la qualité de la reconstruction par *Compressive Sensing*, la méthode est plus stable que l'utilisation « avant *Compressive Sensing* ». L'erreur sur la résolution latérale n'est plus dépendante de l'estimation de la distance contrairement à l'application « avant *Compressive Sensing* ».

Dans les cas de l'estimation de la position « avant » ou « pendant le *Compressive Sensing* », plusieurs solutions sont possibles. Une première est de ne plus raisonner sur le *Compressive Sensing* en 2D uniquement pour augmenter la résolution latérale, mais plutôt de résoudre le problème en 3D directement. Cependant pour cela, il faut être capable de moduler le signal temporellement afin de lui appliquer la base de mesure 3D. Une idée pour moduler le signal temporellement est de recomposer une base d'Hadamard en 3D à partir des acquisitions de base d'Hadamard en 2D sur l'ensemble des pas de temps [165]. Le problème est ensuite résolu directement dans un espace 3D.

Une seconde solution consiste à s’inspirer des travaux de Tachella et *al.* sur Manipop [145] pour mettre en place une méthode d’estimation bayésienne qui englobe les différents problèmes. Il faut alors ajouter la modélisation du DTE, par exemple comme défini par Rapp [121], ainsi que le *Compressive Sensing* par exemple avec le BCS [10] qui suivent un paradigme similaire. Ces méthodes ont des temps d’exécution variant de quelques secondes à plusieurs heures pour une seule image, ce qui est limitant pour des applications nécessitant un temps de reconstruction rapide. Dans le cadre d’acquisitions LiDAR 3D avec des contraintes différentes, Tachella et *al.* [144] ont proposé un algorithme rapide basé sur l’architecture *plug-and-play* [152] pour optimiser Manipop, en remplaçant les distributions *a priori* par des outils de débruitage. Une approche similaire pourrait permettre de développer cette méthode avec des temps de traitement raisonnables et compatibles des applications temps réel. C’est une piste intéressante nécessitant beaucoup de modifications de redéfinir une nouvelle chaîne de traitement ainsi que des outils de débruitages en position et en intensité pour le *Compressive Sensing* et le DTE qui n’existe pas actuellement. Cette approche ne sera pas plus développée dans cette thèse mais reste une perspective prometteuse.

Du fait de garanties théoriques existantes et d’une bonne stabilité avec les méthodes de déconvolution [123], nous avons choisi de nous concentrer sur l’utilisation de la déconvolution « après » le *Compressive Sensing* dans CRISPS-3D. Les avantages et inconvénients des différentes approches sont résumés dans la Tableau 3.1.

Avant <i>Compressive Sensing</i>	Pendant <i>Compressive Sensing</i>	Après <i>Compressive Sensing</i>
<ul style="list-style-type: none"> ✗ Modification de l’algorithme de <i>Compressive Sensing</i> dans CRISPS-3D ou une précision trop importante ✗ Les méthodes existantes ne sont pas assez précises ou très lentes ✓ Réduction du temps de calcul à l’étape du <i>Compressive Sensing</i> ✓ Nécessite peu d’itérations de l’algorithme 	<ul style="list-style-type: none"> ✗ Modification de l’algorithme de <i>Compressive Sensing</i> dans CRISPS-3D ✗ Perte d’assurance de reconstruction (potentiel diminution de l’incohérence entre Φ et Ψ) ✗ Temps de calcul incertain ✓ Algorithme potentiellement plus stable numériquement 	<ul style="list-style-type: none"> ✗ Dépend fortement de la qualité de reconstruction du <i>Compressive Sensing</i> ✗ Nécessite beaucoup d’itérations de l’algorithme ✓ Compatible à l’application de méthodes de déconvolution de l’état de l’art ✓ Meilleure parcimonie en profondeur
Solution[Perspective]		Solution
Développement d’algorithmes de <i>Compressive Sensing</i> appliqués directement sur des images 3D / Formulation bayésienne global permettant de prendre en compte la déconvolution d’ <i>a priori</i> .		Définition / sélection d’une méthode rapide et efficace tirant parti de la parcimonie

TABLE 3.1. Choix du positionnement de l’estimation de la profondeur dans CRISPS-3D par rapport au *Compressive Sensing*

3.2.2 Amélioration de la méthode de déconvolution

Une fois ramené au problème de MP (Équation (3.1)), nous ajoutons une contrainte supplémentaire pour améliorer la qualité de la reconstruction [137]. Par définition $O_{i,j,k}^A$ ne peut pas être négatif. Nous souhaitons résoudre le problème suivant :

$$\begin{cases} \min_{O \in \mathbb{R}^{N_{bin}}} \|O\|_{l_0} \\ s.c. \|X_{i,j,\cdot,k} - hO\|_2^2 \leq \varepsilon, \varepsilon \geq 0 \\ O \geq 0 \end{cases} \quad (3.5)$$

Pour cela, le *Fast Non-Negative Orthogonal Matching Pursuit* (FNNOMP) [164] est sélectionné pour être intégré à notre chaîne de traitement CRISPS-3D. Le FNNOMP est une version accélérée du NNOMP utilisée par Shin [137] et qui assure aussi la non-négativité de l'estimateur. En pratique X est entaché d'une erreur de reconstruction, d'une erreur de mesure et de bruits (bruit solaire et DCR principalement, c.f. Section 1.1.1). Cela tend à rendre le problème de déconvolution plus instable comme illustré dans la littérature [162].

Une façon de limiter l'impact des erreurs et du bruit est de s'intéresser à la forme temporelle de l'impulsion. Elle est souvent régulière pour $\delta_t \ll l_{f_{hmv}}$ en raison du processus physique qui la génère. Elle est le plus souvent approximée par une gaussienne ou une gaussienne déformée (c.f. Section 1.1.1). Comme P_E est régulière et que $X_{i,j,\cdot,k}$ est une convolution de P_E avec une autre fonction, $X_{i,j,\cdot,k}$ doit être aussi régulier [22].

De plus, comme la fonction attendue hO est particulièrement régulière, $X_{i,j,\cdot,k}$ peut être régularisé afin de diminuer l'impact des erreurs de reconstruction sur la déconvolution. Pour gagner en régularité sur $X_{i,j,\cdot,k}$, le signal est filtré en profondeur par un noyau gaussien [160]. Ainsi, nous nous ramenons au problème suivant :

$$\begin{cases} \min_{O \in \mathbb{R}^{N_{bin}}} \|O\|_{l_0} \\ s.c. \|K_\sigma * X_{i,j,\cdot,k} - hO\|_2^2 \leq \varepsilon, \varepsilon \geq 0 \\ O \geq 0 \end{cases} \quad (3.6)$$

avec K_σ un noyau gaussien de variance σ^2 à définir. Le paramètre σ est défini de manière inversement proportionnelle à la confiance pour la reconstruction $X_{i,j,\cdot,k}$, un σ plus faible implique une plus forte confiance en $X_{i,j,\cdot,k}$.

Régulariser la fonction par un noyau gaussien peut cependant introduire un biais, en particulier dans les zones de plus fort gradient de $X_{i,j,\cdot,k}$. Ce biais sera plutôt négatif

autour des maxima et positif proche du front montant de la fonction. L'introduction d'un biais sur la reconstruction dépend aussi fortement de σ par nature du noyau gaussien. Ainsi, dans le cas où les impulsions sont courtes ($\delta_t < l_{f_{hmv}}$), il faut choisir un σ faible (e.g. $\approx 0,5$ pour $l_{f_{hmv}}/\delta_t=3$) et lorsque que $\delta_t \ll l_{f_{hmv}}$ un σ plus élevé (e.g. ≈ 1 pour $l_{f_{hmv}}/\delta_t=10$).

3.2.3 Validation de la méthode sur des signaux simulés

Dans cette section, nous montrons quelques exemples de déconvolution sur des données simulées après reconstructions avec le *Compressive Sensing*. Sur la Figure 3.3, nous utilisons l'algorithme de Richardson-Lucy qui estime les paramètres de la transformée de Fourier du signal divisée par celle du filtre en considérant un bruit de Poisson. Pour Richardson-Lucy, comme le signal est assez mal conditionné, beaucoup de points se retrouvent avec des valeurs très faibles mais le SNR est similaire à celui obtenu avec FNNOMP. Sur la Figure 3.4, nous nous concentrons sur l'utilisation de la méthode décrite dans la Section 3.2.2 avec FNNOMP et un filtre gaussien, en considérant les cas où 1, 2 ou 3 surfaces par colonne sont reconstruites. Dans la Figure 3.4 pour (e) et (f), le « 3+ » indique que l'algorithme ne donne pas de résultat différent avec $K_{max} > 3$ et K_{max} le nombre de surfaces maximum reconstruites avec FNNOMP. En effet, dans ce cas l'algorithme s'arrête soit parce que le résidu est trop faible soit car aucun point ne peut être ajouté dans le support sans donner un résultat négatif.

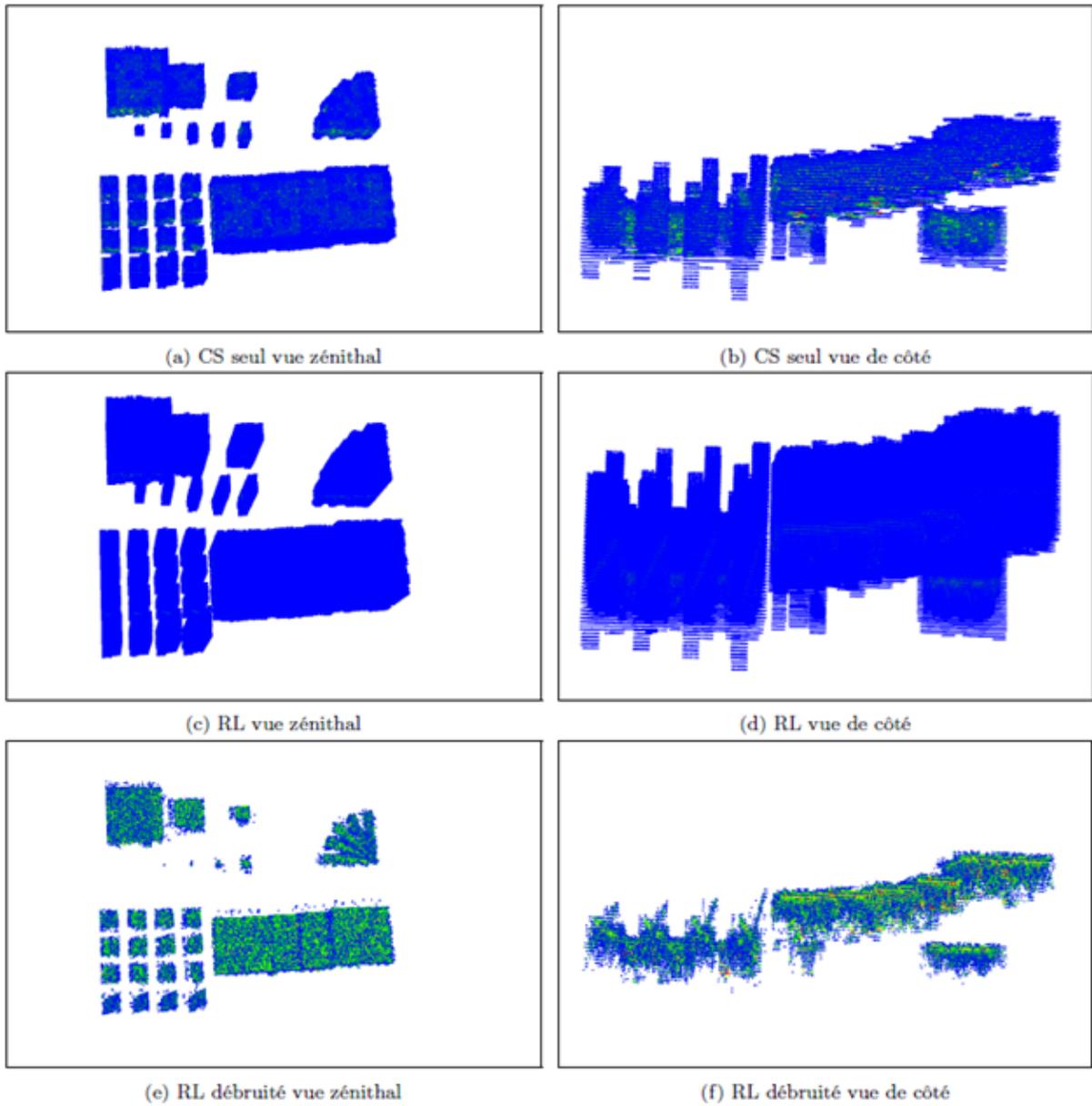


FIGURE 3.3. Comparaison des résultats de la déconvolution *a posteriori* en fonction des différents algorithmes utilisés. Chaque ligne représente le même nuage de points avec des points de vue différents par colonne (vue de dessus à gauche, de côté à droite). Les valeurs sont normalisées entre les nuages de points.

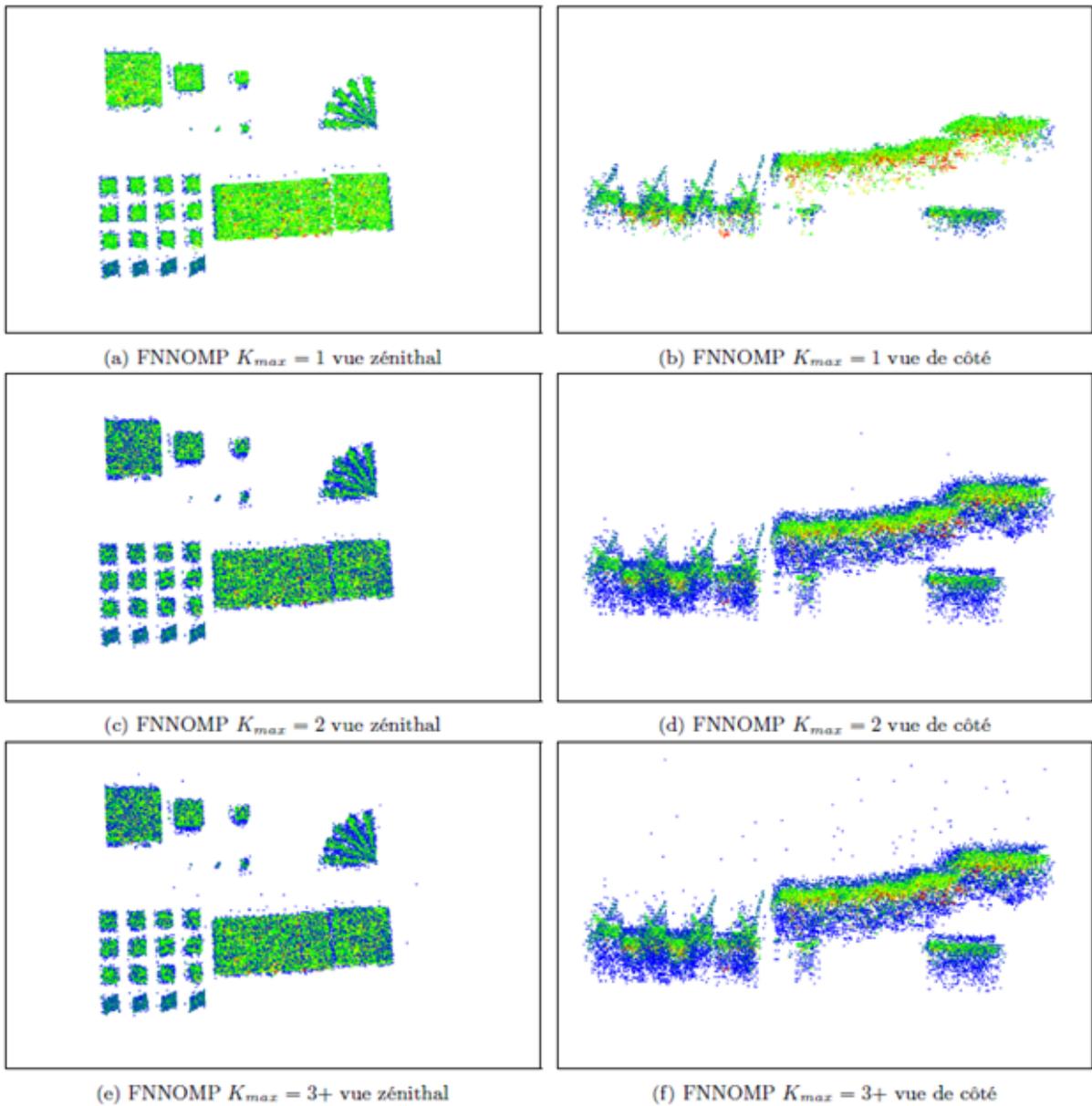


FIGURE 3.4. Comparaison des résultats de la déconvolution *a posteriori* en fonction des différents algorithmes utilisés. Chaque ligne représente le même nuage de points avec des points de vue différents par colonne (vue de dessus à gauche, de côté à droite). Les valeurs sont normalisées entre les nuages de points.

Afin de comparer uniquement les résultats en profondeur, les valeurs du nuage de points intégrés sur l'axe y sont représentées en suivant le plan de coupe en rouge illustré sur la Figure 3.5.

L'algorithme de Richardson-Lucy tend à produire beaucoup de points en profondeur par rapport au MP (reconstruction peu parcimonieuse) comme illustré sur la Figure 3.6. Cependant, il est plus précis en intensité (une ligne ressort principalement). L'algorithme FNNOMP quant à lui produit des nuages de points parcimonieux en profondeur et donc plus fidèles à des acquisitions LiDAR. Cependant, prendre en compte plusieurs surfaces sur une même colonne (en profondeur, $K_{max} > 1$), comme c'est le cas au bord des marches, introduit de fortes erreurs de reconstruction (ainsi que beaucoup de points non-désirables). Dans la Figure 3.6, nous illustrons empiriquement que l'utilisation de noyaux gaussiens pour améliorer la régularité affecte peu les résultats, même si elle permet de mieux reconstruire le signal dans certains cas (*c.f.* Figure 3.7).

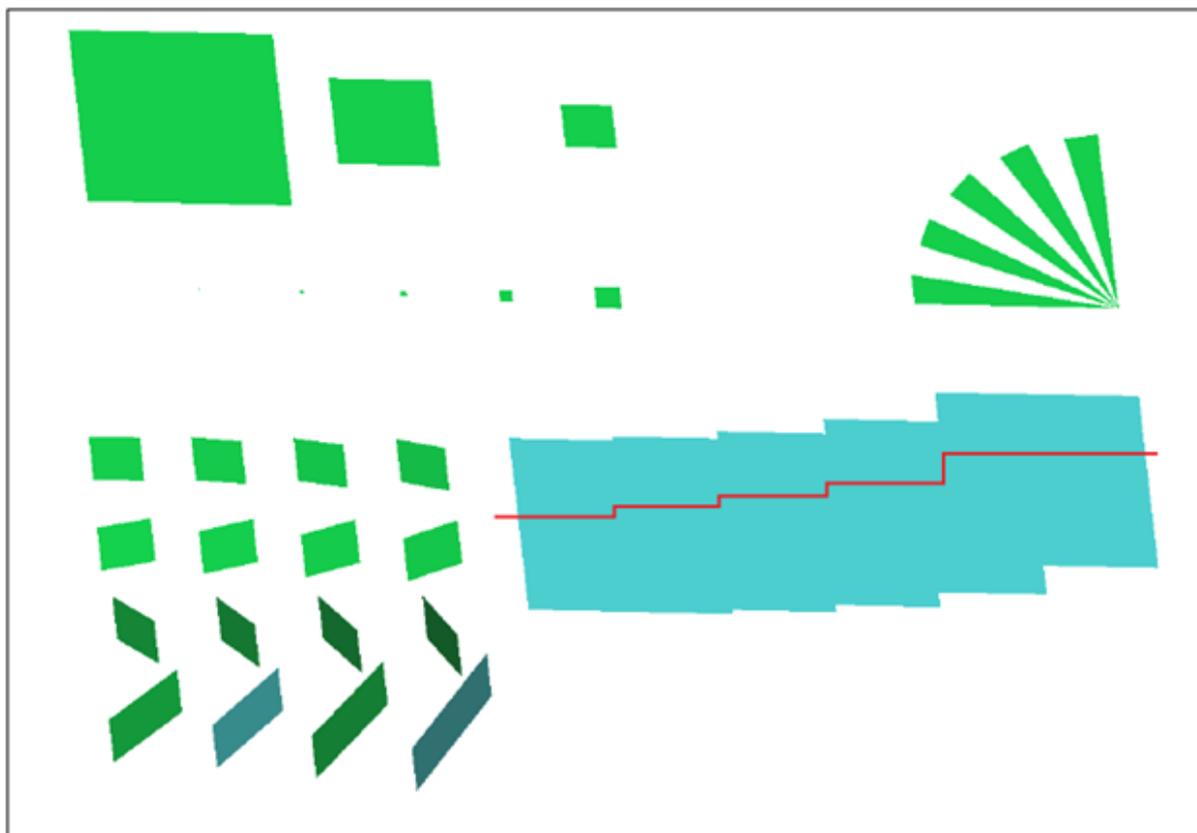


FIGURE 3.5. Scène avec le plan de coupe en rouge, les résultats seront ensuite intégrés selon l'axe y et uniquement sur les « marches »

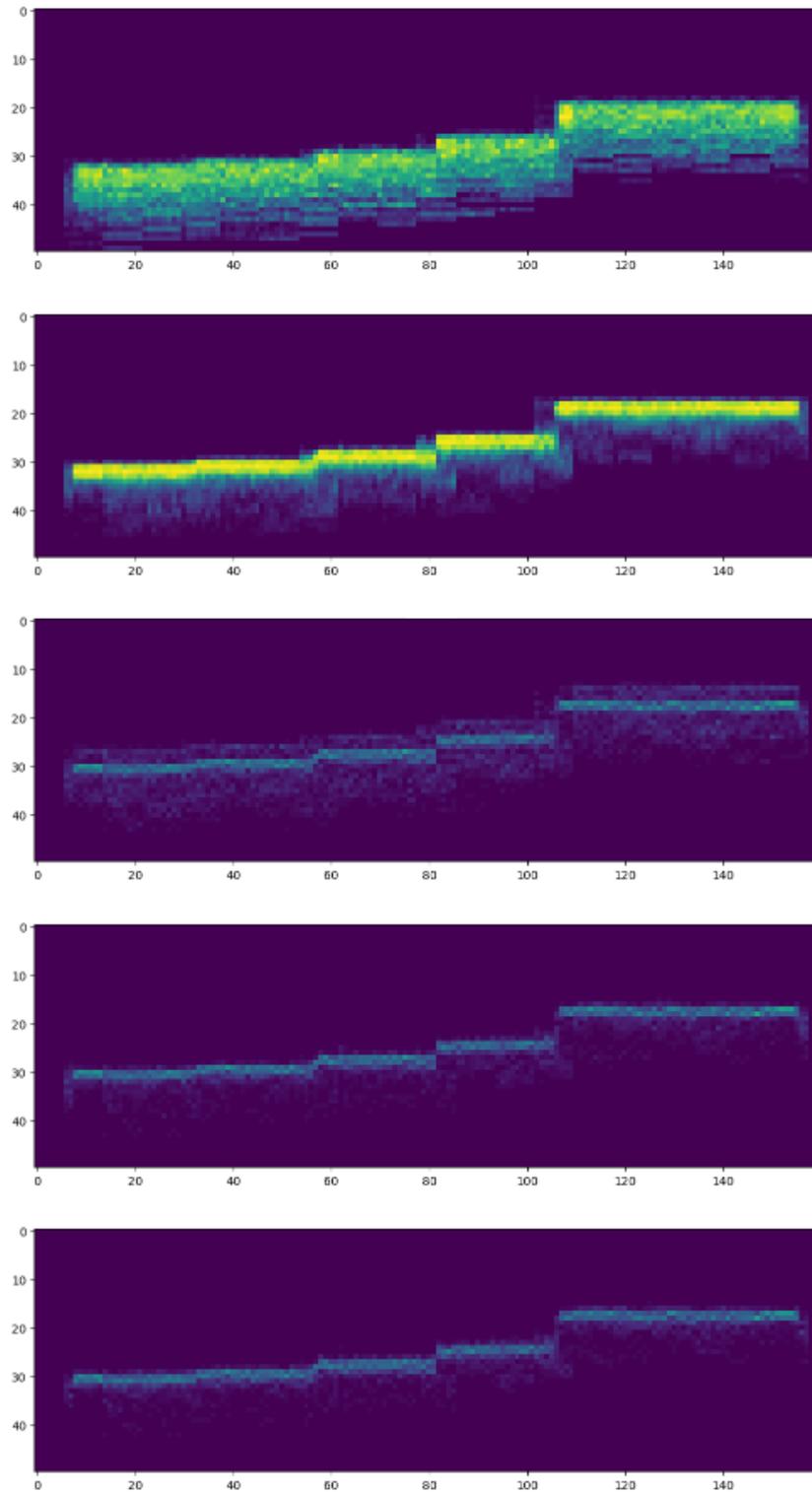


FIGURE 3.6. Résultats intégrés selon l'axe y en suivant le plan de coupe Figure 3.5 puis normalisé de haut en bas : CRISPS-3D seul ; Algorithme de Richardson-Lucy ; FNNOMP sans noyau gaussien et $K_{max} = 2$; FNNOMP sans noyau gaussien et $K_{max} = 1$; FNNOMP avec noyau gaussien et $K_{max} = 1$

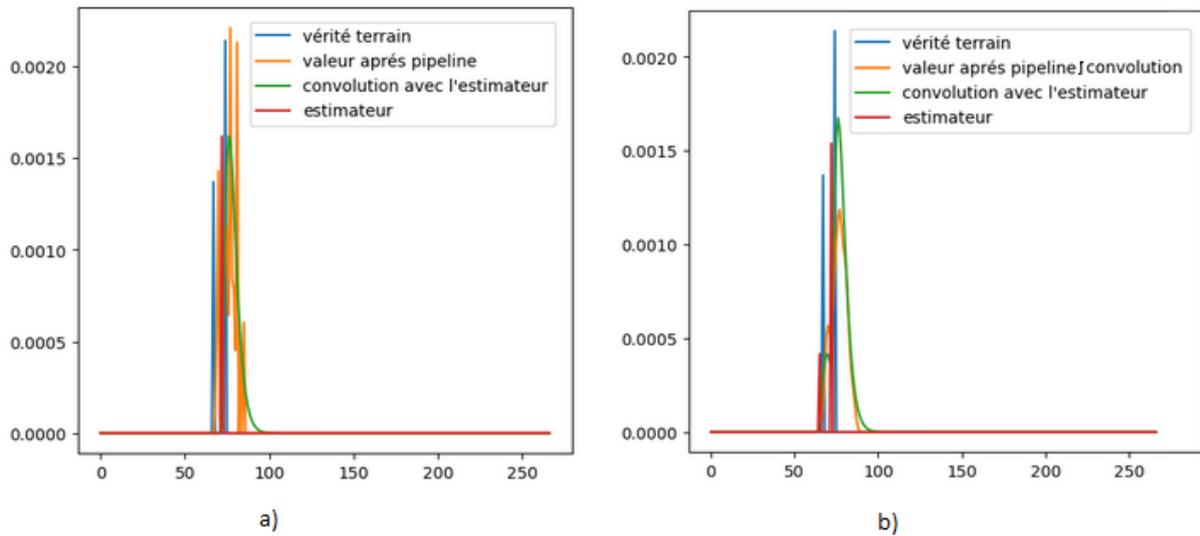


FIGURE 3.7. Comparaison d’une onde complète reconstruite sur une colonne sans (a) et avec (b) ajout de la convolution pour lisser les mesures dans un cas où une amélioration est notable.

Utiliser cette méthode permet de prendre en compte l’utilisation d’impulsions plus longues que la durée d’un pas de temps. Cette approche permet de reconstruire correctement la scène dans le cas où une seule surface par pixel détecteur est suffisante (après *Compressive Sensing*), mais nécessite un débruitage en profondeur pour les scènes comprenant un plus grand nombre de surfaces.

3.3 Ajout d’hypothèses sur la régularité spatiale pour améliorer les résultats

Une fois le signal déconvolué, la distance estimée manque toujours de précision en profondeur. En effet, un même plan est défini dans plusieurs voxels (voir Figure 3.6).

Les méthodes de l’état-de-l’art qui estiment la profondeur réelle à partir d’acquisitions en 3D, considèrent souvent [71, 137, 144] une méthode de débruitage sur la position en profondeur pour obtenir de meilleurs résultats. En suivant ce principe, l’ajout d’une méthode de débruitage sur la profondeur pourrait permettre de meilleurs résultats.

En pratique, une méthode de débruitage est obtenue en formulant une hypothèse sur la régularité du nuage de points. Le problème de débruitage peut alors être résolu en utilisant différents algorithmes de régularisation spatiale : APSS [103], TV [35] ...

Dans la littérature et pour une application à des LiDAR 3D, nous identifions en particulier les travaux de :

- [144] Tachella et *al.* avec l'APSS
- [137] Shin et *al.* avec la TV en profondeur
- [71] Heide et *al.* avec la TV en profondeur et en réflectance

Bien que l'intérêt d'introduire une étape de débruitage exploitant la régularité spatiale de la scène a été identifiée, pour l'instant cette étape n'est pas implémentée dans notre chaîne de traitement CRISPS-3D. Il sera intéressant dans de futurs travaux de recherche d'évaluer la pertinence de cette stratégie et de l'intégrer à CRISPS-3D pour en améliorer la précision en profondeur.

3.4 Synthèse

L'utilisation d'impulsions plus longues dans CRISPS-3D par rapport à sa première version tout en limitant l'erreur sur la reconstruction finale est possible en ajoutant un algorithme d'estimation de la profondeur. Pour cela, notre chaîne de traitement CRISPS-3D met en place un algorithme de déconvolution inspiré par le *Matching Pursuit*. Cet algorithme est défini en exploitant la connaissance de la forme de l'impulsion. D'autre part, les acquisitions LiDAR 3D sont par nature positives et particulièrement parcimonieuses en profondeur. Cette idée bien que déjà présente dans la littérature a été complétée par une version optimisée des algorithmes proposés. En suivant l'idée que la forme d'impulsion initiale est très régulière, une méthode a été proposée pour améliorer la déconvolution mais sans amélioration globales des résultats. La limite du domaine d'application, notamment induite par le DTE, a été définie. Cette limite sur la longueur d'impulsion est définie par rapport à l'énergie reçue qui dépend de nombreux paramètres (énergie d'impulsions, distance de la scène, taille de pupille, transmission atmosphérique, etc. . .) de sorte que le capteur ne soit pas saturé au cours du temps durant l'acquisition.

Chapitre 4

Compressive Sensing et mouvement

Points clefs

- Mouvement des cibles ou du porteur non pris en compte. Nécessité de le considérer pour se rapprocher de cas opérationnels.
- Le mouvement peut être estimé *a posteriori* avec des algorithmes disponibles dans la littérature avec pour hypothèse le déplacement lent des objets dans la scène.
- Lorsque la vitesse est trop importante par rapport à la fréquence d'acquisition, les objets sont déformés au cours de l'acquisition des motifs entraînant des erreurs importantes à la reconstruction.

Contributions principales

- Modélisation du problème de *Compressive Sensing* avec du mouvement potentiellement entre chaque acquisition d'un motif.
- Définition d'une méthode permettant de reconstruire une image, tout en estimant la déformation induite par le mouvement subit durant l'acquisition avec modulation pour le *Compressive Sensing*. Cette méthode de reconstruction s'inspire de différentes propriétés de la reconstruction par *Compressive Sensing* et suppose une déformation régulière.

4.1 Estimation du mouvement *a posteriori*

Cette section présente les travaux sur l'estimation de mouvement entre des nuages de points réalisés conjointement avec A. Goldwurm lors de son stage de Master 2 à l'ONERA dont j'étais l'encadrant.

Une première façon d'aborder la dynamique sur des nuages de points 3D avec du *Compressive Sensing*, est de considérer un mouvement suffisamment lent par rapport à l'acquisition pour pouvoir considérer les acquisitions des motifs pour une reconstruction donnée par *Compressive Sensing* comme statiques.

L'intérêt de considérer les acquisitions comme statiques au cours du *Compressive Sensing* est de pouvoir obtenir une estimation du mouvement entre 2 images sans artefact dû à la reconstruction (voir Section 4.2). De plus, des bibliothèques de calcul existent dans l'état-de-l'art pour estimer un déplacement.

Il s'agit d'un problème de *scene flow*, c'est-à-dire d'estimation de mouvement en trois dimensions entre des nuages de points 3D ou des structures 3D de façon plus générale.

Il existe un certain nombre de méthodes dans la littérature [76, 93] traitant du sujet. Récemment, des approches de *scene flow* fondées sur le transport optimal dans l'espace des caractéristiques [61, 118, 136] ont été définies.

Cependant, ces méthodes sont sensibles aux occlusions et aux acquisitions partielles [136]. D'autres méthodes existent notamment dans la communauté *computer graphics* [102], permettant de gérer les occlusions et acquisitions partielles mais ne permettent pas de prendre en compte de fortes déformations des objets comme le permet le transport optimal.

4.1.1 Méthodes de recalage utilisant le transport optimal pour l'estimation de mouvement 3D

Feydy et *al.* [61] proposent une approche du problème de recalage entre deux nuages de points fondée sur le transport optimal en utilisant comme termes d'attache aux données la distance de Wasserstein. Cette distance consiste à déterminer un plan de transport entre deux mesures. Un plan de transport entre deux mesures p et q est une bijection mesurable d'un espace de mesure vers un autre qui lorsqu'elle est appliqué à p est égale à q et inversement. L'algorithme a été testé sur des nuages de points 3D acquis avec un LiDAR, la méthode utilise des plans 2D sur plusieurs profondeurs plutôt que de considérer le nuage de points dans son ensemble [68].

La méthode de FLOT proposée par Puy et *al.* [118] est fondée sur le transport optimal ainsi que sur des réseaux de neurones utilisés pour extraire des caractéristiques de la scène

(*features*). Cette méthode correspond à une implémentation dont l'exécution est rapide et parallélisable pour accélérer la vitesse de calcul. De plus, Puy et *al.* [118] fournissent les poids du réseau de neurones pré-entraînés ainsi qu'une base de données de nuage de points 3D permettant d'entraîner le modèle de réseau de neurones.

Shen et *al.* [136] ont proposé une méthode qui semblait également adaptée pour ce type de problème de *scene flow*. Elle repose sur une combinaison entre réseaux de neurones et modèles de déformation pour estimer le mouvement. Elle introduit en plus une étape de pré-alignement *via* une transformation affine pour faciliter la prédiction et une étape postérieure de recalage pour affiner l'estimation.

Finalement, nous avons choisi FLOT pour le recalage de points *a posteriori* d'acquisitions LiDAR 3D, notamment de par sa rapidité d'exécution et la disponibilité de poids du réseau de neurones pré-entraînés. Dans la suite de la section, cette méthode est détaillée puis quelques résultats obtenus sur des acquisitions LiDAR 3D simulées sont donnés .

FLOT est défini pour des nuages de points sans intensité, en utilisant uniquement des coordonnées spatiales en entrée. Dans notre implémentation, nous changeons cependant légèrement la formulation en considérant que chaque point est défini par ses coordonnées spatiales en 3D ainsi qu'une intensité (ou une masse au sens du transport optimal). Cette méthode a pour bloc principal un module de transport optimal permettant de trouver la correspondance entre des points dans l'espace. Ce bloc peut être utilisé seul, en prenant en entrée directement les nuages de points p et q , donnant une correspondance directe entre les points permettant d'obtenir une estimation du mouvement entre ces nuages de points. Cependant, de meilleurs résultats sont obtenus lorsqu'elle est décomposée en trois étapes successives, comme précisée sur la Figure 4.1.

Deux nuages de points $p, q \in \mathbb{R}^{n \times 3}$, sont considérés comme représentant tous deux la même scène à deux instants t et $t + 1$, et ayant les intensités $a, b \in \mathbb{R}^n$. Un premier bloc de réseaux de neurones convolutif prenant en entrée les deux nuages de points p et q permet l'extraction de caractéristiques pour chacun de ces nuages de points. En traitement d'images et en apprentissage machine, les caractéristiques sont des informations sur le contenu d'une image, comme par exemple la structure d'une certaine région. Ici, les caractéristiques peuvent être des structures spécifiques dans l'image telles que des points, des bords ou des objets dépendant notamment de la profondeur du réseau [1]. Ce module extrait donc des caractéristiques et les utilise pour le calcul du coût pour le transport optimal entre les deux nuages de points. Le deuxième bloc est celui du transport optimal prenant en entrée les deux nuages de points ainsi que les caractéristiques extraites. Il en ressort la correspondance entre les nuages de points avec la matrice de coût calculée grâce aux caractéristiques. C'est pourquoi ce module peut également être utilisé seul, en prenant une matrice de coût plus commune comme par exemple celle des distances

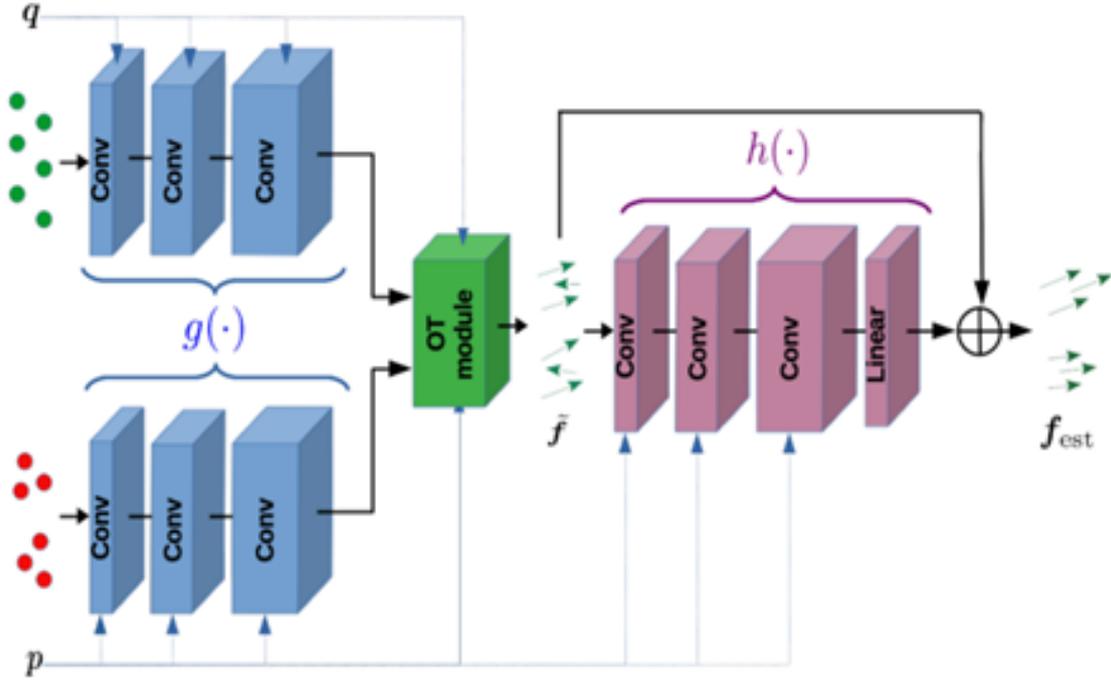


FIGURE 4.1. Illustration du fonctionnement de FLOT pour résoudre le problème de *scene flow* [118]

euclidiennes. Finalement, la dernière étape correspond à l'estimation du mouvement à partir de la correspondance, affinée avec un réseau de neurones résiduel. Dans le cas de FLOT, le réseau de neurones est un réseau ajoutant des « raccourcis ». Le nuage de points est réutilisé à différentes couches du réseau pour calculer non plus des caractéristiques mais des résidus afin d'affiner l'estimation du mouvement, de sorte que $f_{est} = \tilde{f} + h(\tilde{f})$ (voir Figure 4.1).

Étape du calcul du plan de transport avec transport optimal

Le but de FLOT est donc de déterminer, à partir de deux nuages de points $p, q \in \mathbb{R}^{n \times 3}$, le champ de vecteurs $f \in \mathbb{R}^{n \times 3}$ qui représente les déplacements de chaque point entre les deux images. Ainsi, appliquer f au premier nuage de points doit permettre de retrouver le second. Le champ de vecteurs est défini tel que $p + f = Pq$, avec $P \in \{0, 1\}^{n \times n}$ une matrice de permutation. Donc pour chaque point p_i ayant une masse a_i , chaque q_j ayant une masse b_j , reçoit la masse a_i de p_i si $p_i + f_i = q_j$, soit si $P_{i,j} = 1$. Pour trouver ce champ de vecteurs le plan de transport entre p et q est calculé, soit la matrice $T \in \mathbb{R}_+^{n \times n}$ qui doit satisfaire :

$$T = \operatorname{argmin}_{U \in \mathbb{R}_+^{n \times n}} \sum_{i,j} U_{i,j} C_{i,j} \quad (4.1)$$

où U est admissible, c'est-à-dire que la conservation des masses est respectée donc que $UI_m = a$ et $U^T I_n = b$.

Cependant, tel que défini dans l'Équation (4.1) le problème de transport optimal est un problème de programmation linéaire dans un espace potentiellement grand (nuage de points 3D). Afin d'obtenir de meilleures propriétés de convergence M. Cuturi [49] introduit une régularisation entropique dans le problème de transport qui est repris dans la formulation de Puy et *al.* [118]. Cela permet en particulier d'obtenir un problème strictement convexe et permet l'utilisation de l'algorithme de Sinkhorn qui est particulièrement efficace [49]. De plus, dû à la nature des acquisitions et du mouvement, il peut arriver que des objets disparaissent ou apparaissent dans la scène entre acquisitions. Ces variations entraînent une violation de la contrainte de conservation de la masse. C'est pourquoi Puy et *al.* [118] proposent, afin de contourner ce problème, l'ajout de termes de divergence permettant de relâcher la contrainte de conservation de masse. En reprenant la formulation de Puy et *al.* [118] en introduisant les intensités des nuages de points, le problème de transport optimal régularisé et relaxé est donc :

$$T = \operatorname{argmin}_{U \in \mathbb{R}_+^{n \times n}} \sum_{i,j}^n [U_{i,j} C_{i,j} + \varepsilon U_{i,j} (\log(U_{i,j}) - 1)] + \lambda KL(U \mathbf{1}, a) + \lambda KL(U^T \mathbf{1}, b) \quad (4.2)$$

Le terme désigné par KL est la divergence de Kullback-Leibler qui est, en théorie des probabilités et de l'information, une fonction de mesure de dissimilarités entre deux distributions de probabilités. Elle est définie, pour deux probabilités discrètes P et Q , par $KL(P, Q) = \sum_i P_i \log(\frac{P_i}{Q_i})$. Il s'agit d'une formulation connue dans le domaine du transport optimal entre deux mesures de masses éventuellement différentes avec au plus n composants, où les contraintes marginales du transport optimal standard sont assouplies *via* la divergence de Kullback-Leibler [109], le *Unbalanced Optimal Transport* (UOT). Un algorithme de Sinkhorn modifié permet de résoudre ce type de problème de l'UOT régularisé par entropie [42, 117]. La matrice de coût C dans la formulation du transport peut être définie de plusieurs manières. Pour cela, la distance euclidienne entre chaque point $C_{i,j} = \|p_i - q_j\|^2$ est considérée [158]. Cependant, de meilleurs résultats sont observés en prenant une matrice de coût définie à partir des caractéristiques extraites dans FLOT [118]. On reprend donc le réseau $g : \mathbb{R}^{n \times 3} \rightarrow \mathbb{R}^{n \times c}$ prenant en entrée les nuages de points, les caractéristiques se notent donc $g(p)_i$ et $g(q)_j$. La matrice de coût est définie, d'après [118] par :

$$C_{i,j} = \left(1 - \frac{g(p)_i^T g(q)_j}{\|g(p)_i\|_2 \|g(q)_j\|_2} \right) \times i_{\|\cdot\|_2 \leq d_{max}}(p_i - q_j) \quad (4.3)$$

$$\text{avec } i_{\|\cdot\|_2 \leq d_{max}}(p_i - q_j) = \begin{cases} 1 & \text{si } \|(p_i - q_j)\|_2 \leq d_{max} \\ +\infty & \text{else} \end{cases}$$

Autrement dit, $i_{\|\cdot\|_2 \leq d_{max}}(p_i - q_j)$ est une indicatrice permettant de retenir un coût entre p_i et q_j seulement si la distance euclidienne entre ces deux points est inférieure à d_{max} qui est fixée par rapport à l'échelle des distances des nuages de points de nos données. L'indicatrice empêche donc la correspondance entre des points trop éloignés. L'autre partie du terme du coût détermine une similarité entre les caractéristiques.

Explication plus détaillée des paramètres de la méthode

Dans cette méthode, deux paramètres importants sont présents. Tout d'abord, un paramètre de régularisation du transport $\varepsilon > 0$ qui permet la mise en place d'un schéma numérique. Plus ce terme ε est faible, plus précis sera le transport et plus instable sera le schéma numérique. Ensuite, le paramètre λ contrôle la régularisation de la masse. En effet, il permet de contrôler à quel point la masse transportée d'une distribution à l'autre varie. Lorsque ce paramètre est très élevé, on peut considérer qu'il impose une conservation de masse stricte. De plus, dans l'algorithme de Sinkhorn, la régularisation est contrôlée par le facteur $\lambda/(\lambda + \varepsilon)$. Et lorsque λ est très grand par rapport à ε , ce facteur tend vers 1. C'est pourquoi il est difficile de trouver un équilibre entre ces deux paramètres. Puy et al. [118], ont entraînés ces paramètres en même temps que les poids du réseau de neurones. Comme les modèles de réseaux de neurones avec des poids pré-entraînés sont utilisés, les paramètres pré-entraînés seront également réutilisés. L'implémentation de l'algorithme de FLOT avec des intensités ayant été réalisée par A. Goldwurm [68], elle ne sera pas plus détaillée dans cette thèse.

4.1.2 Application de méthode de FLOT avec intensités sur des acquisitions LiDAR 3D simulées

Un jeu de données est d'abord défini à partir de la scène présentée sur la Figure 4.2. Cette scène représente un cube disposé à côté du buste de la Marianne, le cube fait 1 m de côté et est espacé de 1 m par rapport au buste de la Marianne. Dans la scène, le cube est composé de 12 facettes de l'ordre du mètre-carré et Marianne d'environ 10^5 facettes de l'ordre du centimètre-carré voir du millimètre-carré.

Un ensemble de 90 nuages de points est défini à partir de cette scène. Entre chaque nuage de points, une rotation de la scène centrée sur le centre du cube est appliquée. Cette rotation correspond à une rotation du cube sur lui-même et une rotation du buste autour du cube. La rotation entre 2 nuages de points est de 1° , les déplacements entre les images sont donc faibles. Ceci, permet de tester graduellement l'algorithme de FLOT sur des déplacements afin d'identifier les limites de la méthode.

Un premier test de recalage (voir Figure 4.3) est réalisé entre le premier et le dixième nuage de points, ce qui correspond à une rotation totale de 9° .

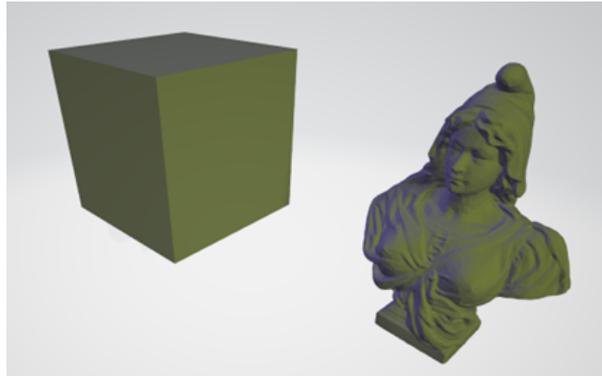


FIGURE 4.2. Scène simplifiée représentant deux objets en trois dimensions : un buste de la Marianne et un cube

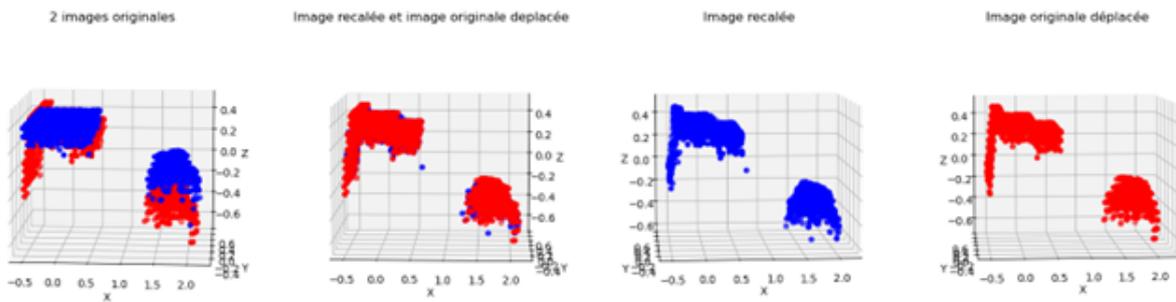


FIGURE 4.3. Recalage des nuages de points 1 et 10, le nuage de points en bleu correspondant à l'image recalée et le nuage de points en rouge l'image d'origine

Le premier nuage de points en bleu est correctement recalé par rapport au second nuage de points qui est initialement décalé de 9° . La Figure 4.4 illustre les erreurs commises avec l'algorithme de Sinkhorn. La convergence est atteinte en quelques itérations (< 10).

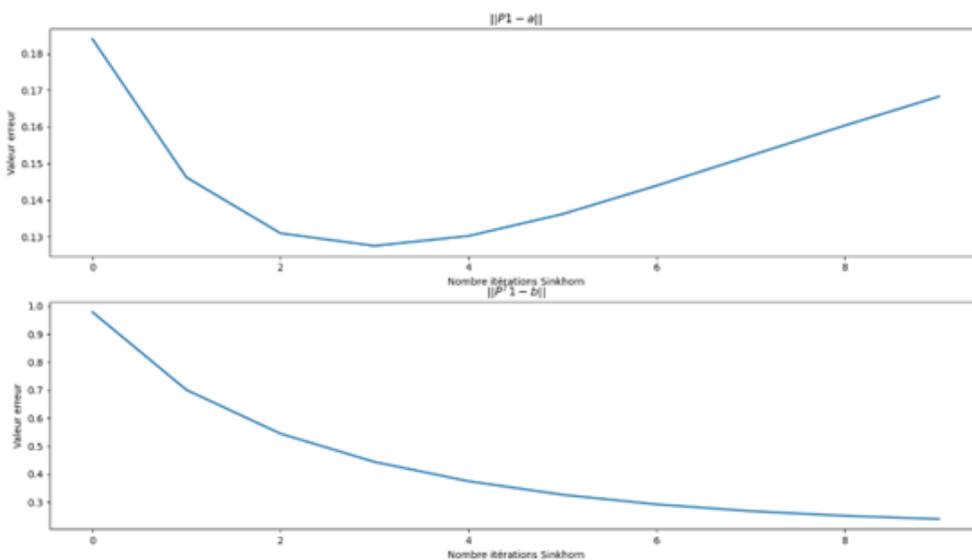


FIGURE 4.4. Erreur dans l'algorithme Sinkhorn pour les nuages de points 1 et 10

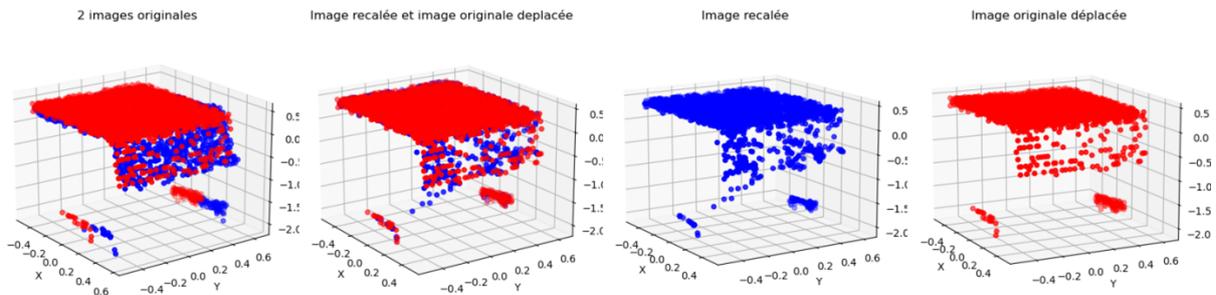


FIGURE 4.5. Recalage des nuages de points 80 et 90

Pour une rotation de 10° entre 80° et 90° , le nuage de points bleu se déplace correctement par rapport au nuage de points rouge. Le nuage de points initial (en bleu) recouvre correctement le nuage de points déplacé (en rouge) après application de l'estimation du champ de vecteurs de déplacement. Les erreurs correspondantes et commises avec l'algorithme de Sinkhorn (Figure 4.6) ont une convergence légèrement différente et notamment sur $\|P1 \check{a}\|$.

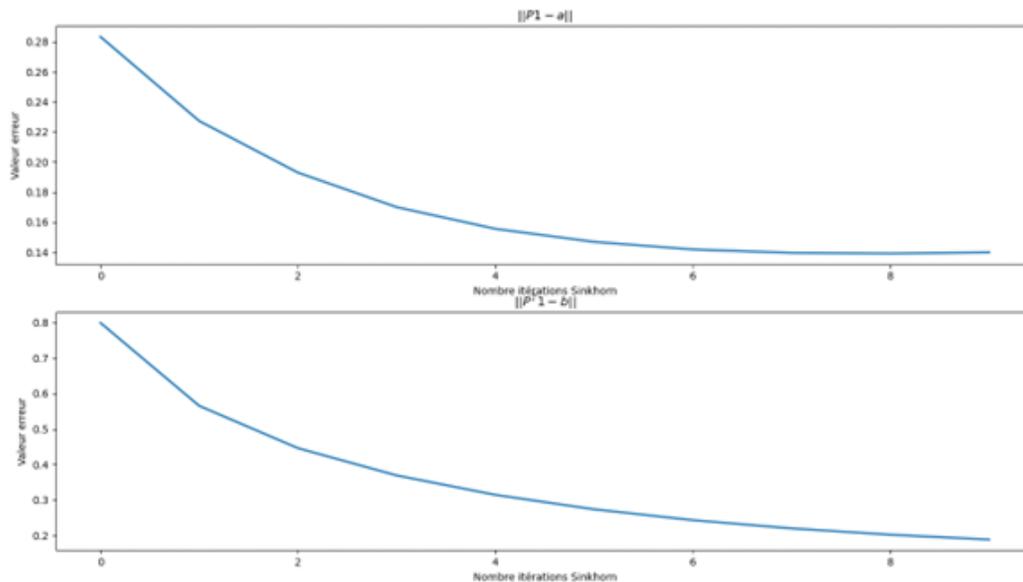


FIGURE 4.6. Erreur dans l'algorithme Sinkhorn pour les nuages de points 80° et 90°

Pour aller plus loin, le jeu de données avec des déplacements plus importants peut être testé comme par exemple entre la première et la 15^{ème} image. La rotation globale est de 14° . La Figure 4.7 illustre les résultats obtenus avant et après recalage *via* notre algorithme. Des points « fantômes » de correspondance erronée entre les deux formes sont visibles.

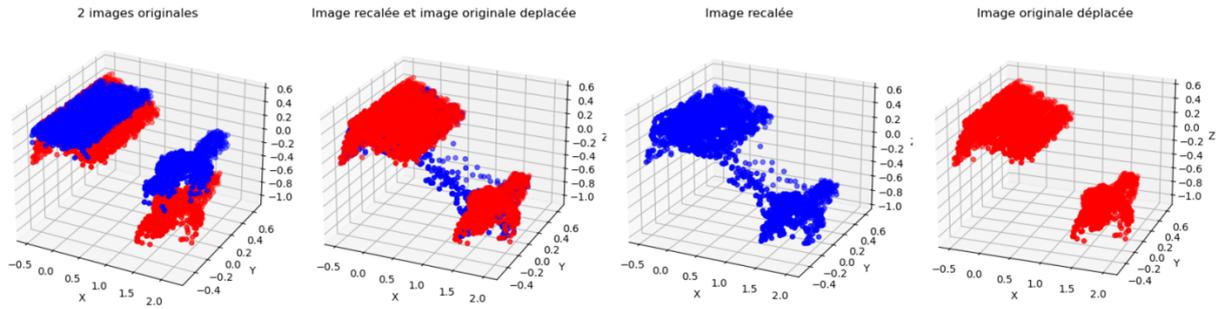


FIGURE 4.7. Recalage des nuages de points correspondant à 1° et 15°

Dans FLOT [118], le terme d_{max} est identifié comme le paramètre qui permet à l'indicatrice d'empêcher de trouver la correspondance entre des points trop éloignés. En effet, ce terme d_{max} doit être initialement fixé à une échelle qui correspond à l'espace maximal entre les points sur un même objet.

Dans le cas de ce jeu de données (cube + buste), la distance entre les deux objets est d'environ 1 m. Le paramètre d_{max} doit donc être fixé comme tel. En reprenant les simulations avec le paramètre d_{max} modifié et en représentant sur la Figure 4.8 le résultat du recalage pour les nuages de points issus de la première et de la 30^{ème} image. La rotation globale équivalente est fixée à 29° .

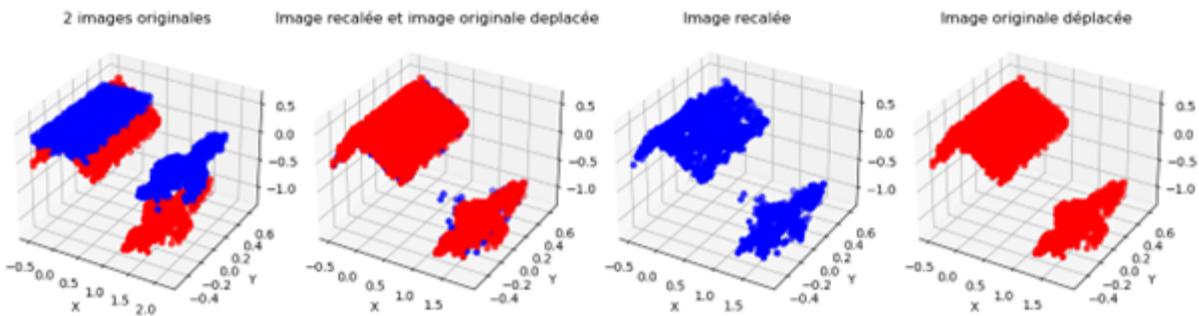


FIGURE 4.8. Recalage des nuages de points correspondant à 1° et 30°

La précision est meilleure avec une convergence rapide en quelques itérations *via* notre algorithme de Sinkhorn. Ainsi les erreurs de correspondances sont diminuées même si des déplacements plus importants sont considérés (voir Figure 4.9).

Après quelques itérations l'erreur reste relativement constante, et l'algorithme de Sinkhorn a convergé. La solution trouvée est donc proche de la solution optimale.

L'estimation du mouvement et de la scène sur des données modulées pour le *Compressive Sensing* est un sujet novateur et très peu traité dans la littérature. La suite du chapitre s'attèle à définir et résoudre cette problématique.

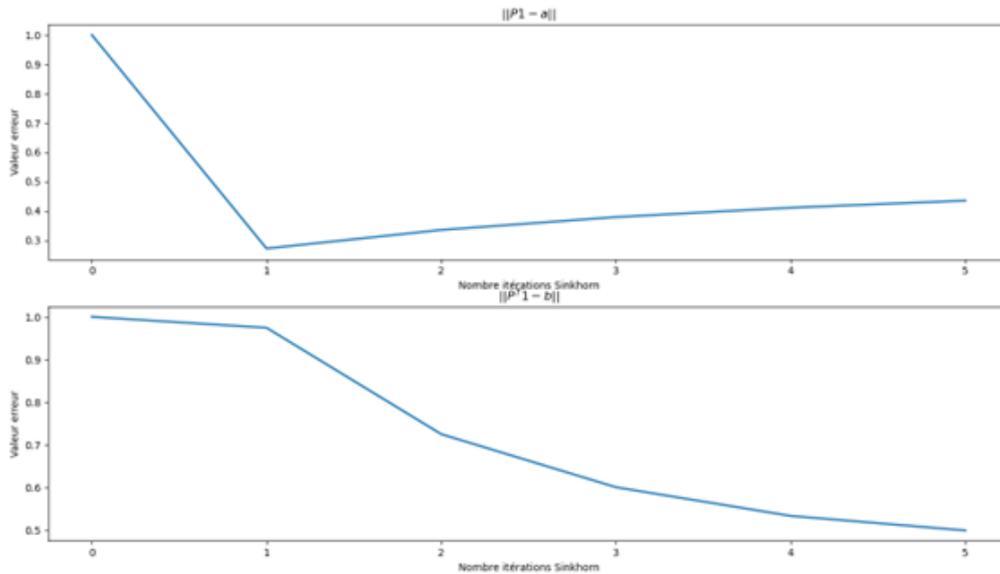


FIGURE 4.9. Erreur dans l'algorithme Sinkhorn pour les nuages de points à 1° et 30°

4.2 *Compressive sensing* et mouvement entre l'acquisition des motifs

Cette section est consacrée à l'étude du mouvement relatif des cibles ou du porteur dans le cadre du *Compressive Sensing*. Afin de simplifier le problème de reconstruction et de se concentrer sur la prise en compte du mouvement durant l'acquisition des motifs, le problème est ramené à un problème à deux dimensions. Cette simplification est équivalente à une acquisition en trois dimensions avec une déformation latérale, uniquement dans le plan.

4.2.1 Modélisation du problème de *Compressive Sensing* simplifié prenant en compte le mouvement

Tout d'abord, afin de modéliser le problème mathématique et d'observer les effets liés aux mouvements, un certain nombre d'hypothèses sont faites :

- i Les déplacements sont inférieurs à la taille d'un pixel du DMD projeté à la distance d'observation pour la durée d'acquisition et pour un motif.
- ii Il n'y a pas de bruit et les turbulences sont toujours ignorées afin que les erreurs et les artefacts liés à la reconstruction ne soient causés que par le *Compressive Sensing* et/ou le mouvement.
- iii Le mouvement (des cibles et/ou du porteur) n'est pas trop erratique de sorte qu'une approximation affine locale n'introduise pas trop d'erreurs.
- iv Un seul pixel est considéré et un objet ne sort pas ou n'entre pas dans le champ de vue.

En outre, les contraintes en termes de temps d'intégration, de temps d'exécution ou encore d'embarquabilité ne seront plus considérées.

Le problème de *Compressive Sensing* « classique » est repris comme point de départ :

$$Y = \Phi X, \text{ et } X = \Psi s \text{ avec } s \text{ } K\text{-sparse} \quad (4.4)$$

Le mouvement est introduit comme :

$$X_t = X_0 \circ \phi_t^{-1} = \phi_t \cdot X_0 \quad (4.5)$$

Avec ϕ_t la transformation représentant le mouvement de X_0 vers X_t .

Ainsi en ajoutant la variable de temps à la mesure, nous obtenons :

$$Y_t = \Phi_t X_t, \text{ et } X_t = \Psi s_t \text{ avec } s_t \text{ } K\text{-sparse} \quad (4.6)$$

Dans cette section, t est utilisé pour représenter le temps discrétisé *via* l'indice des motifs acquis. En remplaçant X_t dans l'Équation (4.6) :

$$Y_t = \Phi_t D_t X_0, \text{ et } X_0 = \Psi s_0 \text{ avec } s_0 \text{ } K\text{-sparse} \quad (4.7)$$

où D_t est une matrice de taille $N \times N$, appliquant le mouvement ϕ_t à X_0 ($D_t \approx X_0 \circ \phi_t^{-1}$), en approchant la valeur grâce à une interpolation si ϕ_t ne déplace pas X_0 exactement sur la grille.

Ainsi, si D_t est connue pour tout t et que $\phi_t D_t$ suit la RIP, le problème redevient un problème de *Compressive Sensing* « classique ». Actuellement, nous n'avons pas pu déterminer de conditions sur ϕ_t pour assurer que $\phi_t D_t$ suit cette propriété. L'étude de ces conditions est laissée en perspectives et sera nécessaire au déploiement d'algorithmes basés sur cette modélisation. Cependant, un exemple de reconstruction par *Compressive Sensing* avec et sans correction grâce à D_t est donnée sur la Figure 4.10 pour illustrer le potentiel de cette approche.

L'estimation obtenue en résolvant le problème de *Compressive Sensing* avec $\phi_t D_t$ tend à être bruitée par des artefacts à haute fréquence spatiale. L'estimation peut alors être débruitée en la régularisant avec la TV [35]. Ceci est illustré Figure 4.10.

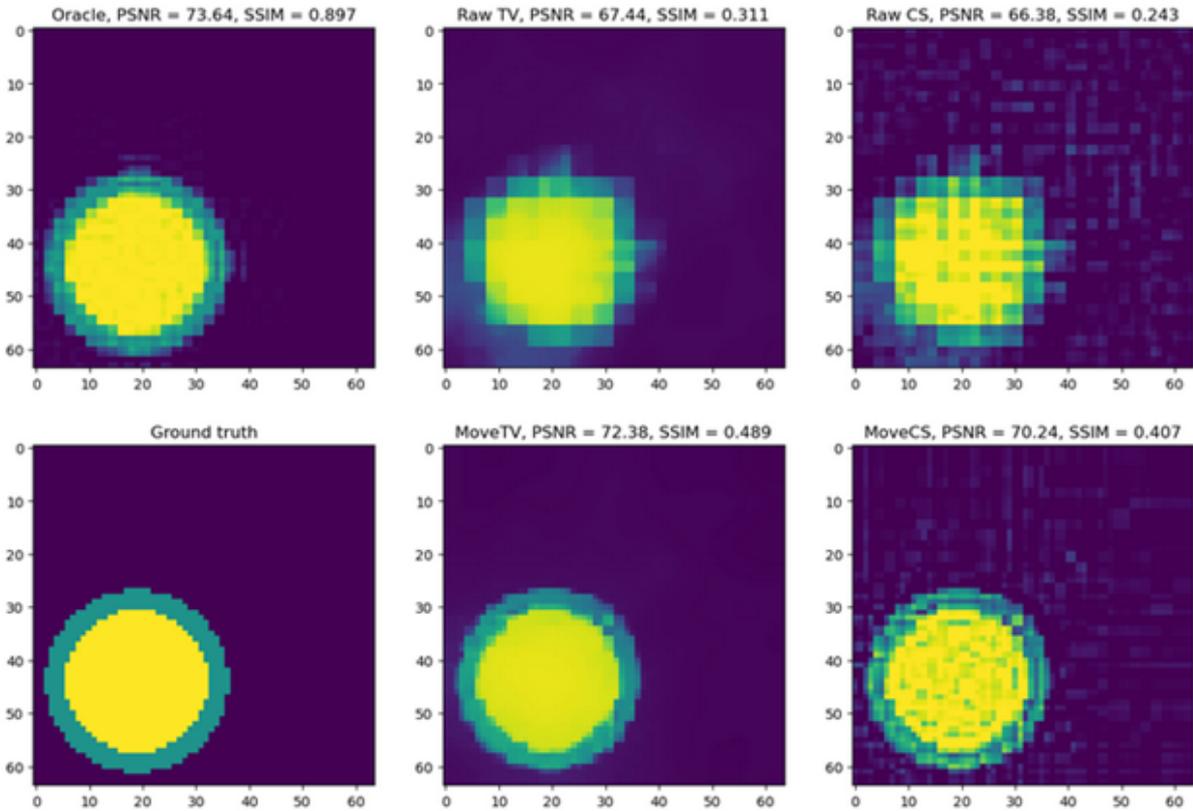


FIGURE 4.10. Comparaison entre les résultats de différentes reconstructions par *Compressive Sensing* (taux de compression à 25%) avec et sans information sur le mouvement. De gauche à droite et de haut en bas : la reconstruction avec *Compressive Sensing* s'il n'y avait pas de mouvement ; la reconstruction sans D_t et après régularisation avec la TV pour retirer les artefacts ; le *Compressive Sensing* seul sans D_t ; la vérité terrain ; la reconstruction avec D_t et après régularisation avec la TV ; le *Compressive Sensing* seul avec D_t .

4.2.2 Définition d'une stratégie de reconstruction d'image en utilisant l'approche *Large Deformation Diffeomorphic Metric Mapping*

Le problème de base étant posé, une méthode spécifique doit être définie afin de le résoudre sans information sur D_t . Pour cela, certaines spécificités du *Compressive Sensing* avec des bases Hadamard peuvent être utilisées à notre avantage. En utilisant uniquement les premiers éléments les moins résolus de la base, cela permet de reconstruire une image avec moins de motifs [3, 142]. En exploitant cette idée avec (iii) Section 4.2.1, un maillage est défini sur le mouvement afin de l'approcher sur l'ensemble des acquisitions localement, le pas du maillage correspond à un nombre m de motifs.

Les motifs acquis sont définis comme suit :

Soit $t \in [0; M]$ avec M le nombre de motifs acquis et $t_0 = 0, \dots, t_n = M - m$, n mailles et $m \ll \frac{M}{n+1}$, tel que :

$$\tilde{\Phi}_t = \begin{cases} \Phi_{t_i-t}, & \text{si } t \in [t_i, t_i + m] \forall i \\ \Phi_{t-i \times m}, & \text{si } t \in [t_i + m, t_{i+1}] \forall i \end{cases} \quad (4.8)$$

L'idée étant de reconstruire une image moins résolue à chaque maille en utilisant m motifs (voir Figure 4.11), puis d'estimer $\phi_{t_i}^{-1}$ de proche en proche. Une fois $\phi_{t_i}^{-1}$ estimé, le D_t est calculé pour tous les $t \notin [t_i, t_i + m]$ avec $i > 0$. Enfin la reconstruction est obtenue en utilisant le *Compressive Sensing* avec la base déformée.

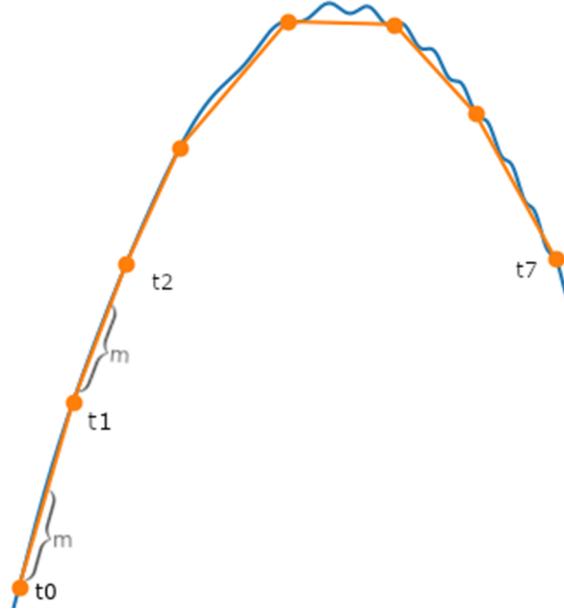


FIGURE 4.11. Illustration de la stratégie de maillage du mouvement. Le mouvement d'un objet est représenté en bleu et les points à partir desquels le mouvement est estimé, en orange.

Localement, l'estimation du mouvement entre deux points du maillage est réalisée dans notre cas à partir de l'algorithme LDDMM (*Large Deformation Diffeomorphic Metric Mapping*) [13].

Définition de LDDMM

Pour deux images X_{t_1} et X_{t_2} , le principe de LDDMM est de chercher une transformation (*mapping*) ϕ_t dans l'espace des difféomorphismes, telle que $X_{t_2} = X_{t_1} \circ \phi_1^{-1}$ et $\phi_0 = Id$, sans hypothèse de faible déformation.

Un difféomorphisme est une application bijective, différentiable sur son ensemble de départ et dont la réciproque est différentiable d'arrivé [104]. En reconstruisant un ϕ_t

difféomorphique, une transformation régulière et inversible est obtenue et son inverse est régulière produisant des transformations préservant les courbes et les surfaces [13]. Une façon de conceptualiser ces transformations est d’imaginer la grille représentant l’image comme un tissu élastique (voir Figure 4.12).

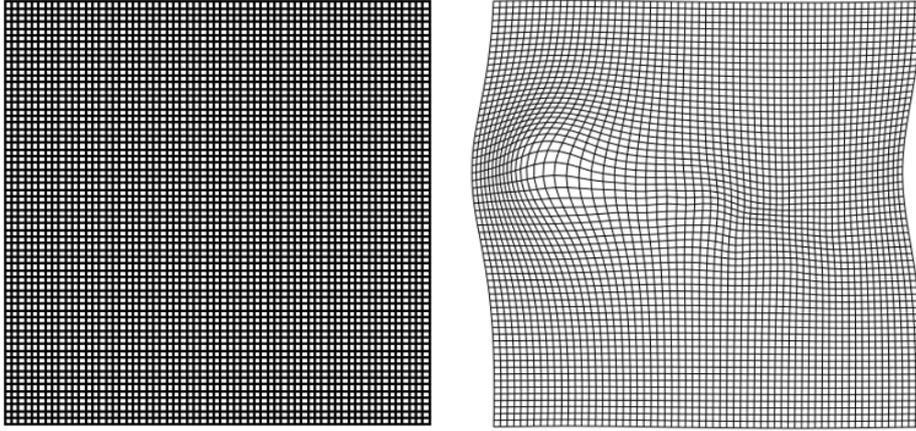


FIGURE 4.12. Exemple de transformation obtenue avec LDDMM [13]

Dans LDDMM, la transformation ϕ_1^{-1} est obtenue en résolvant un problème d’optimisation (voir l’Équation (4.9)). Pour cela, ϕ_1 est défini comme le flux d’un champ de vitesses v_t (à $t = 1$), dépendant du temps [13], le temps étant défini sur l’intervalle $[0; 1]$ de l’image initiale à l’image déformée. Le lien entre v_t et ϕ est défini par l’équation différentielle $\phi'_t = v_t(\phi_t)$ dans l’espace des transformations. Ainsi, ϕ_1 peut être obtenu par la relation $\phi_1 = \phi_0 + \int v_t(\phi_t) dt$.

Dans les travaux de Beg et *al.* [13], le problème d’optimisation est formulé d’un point de vue variationnel par rapport à l’espace Riemannien V des champs de vitesses réguliers afin de retrouver une transformation elle aussi régulière. En ajoutant un terme d’attache aux données dans L_2 et la norme de v dans V intégrée sur l’ensemble des temps, le problème d’optimisation devient alors :

$$\tilde{v}_t = \underset{v: \phi'_t = v_t(\phi_t)}{\operatorname{argmin}} \left(\int_0^1 \|v_t\|_V^2 dt + \frac{1}{\sigma^2} \|X_{t_i} \circ \phi_1^{-1} - X_{t_{i+1}}\|_{L^2}^2 \right) \quad (4.9)$$

La définition de V peut varier et changer les propriétés de l’algorithme, ainsi que de la transformation estimée. Comme démontré par Risser et *al.* [124], le niveau de régularité contraint l’espace de Hilbert V , il peut s’exprimer à travers un noyau de convolution des déformations K . En un temps t , la norme $\|v_t\|_V^2$ peut alors être calculée comme égale à $\langle F(v_t)F(K)^{-1}, F(v_t) \rangle_{L^2}$ où $F(\cdot)$ est une transformée de Fourier.

Une implémentation numérique est proposée par Beg et *al.* [13] pour résoudre le problème (Équation (4.9)) en utilisant une descente de gradients avec un pas de plus profond de descente. Dans ce cadre, le problème est discrétisé sur le temps et l’espace. En particulier,

le champ de vecteurs de vitesses v_t est calculé sur une grille spatialement homogène des points, qui correspond généralement aux pixels de l'image. Le temps t est aussi discrétisé sur une grille de temps homogène $t_i = i\Delta_t$, $i \in [0, T]$ avec $\Delta_t = \frac{1}{T}$. Pour chaque temps t_i , le calcul du gradient de l'Équation (4.9) par rapport aux champs de vitesses discrétisés v_i nécessite de calculer une transformation *forward* ($X_0 \circ \phi_{t_i}^{-1} = X_{t_i}$) et une transformation *backward* ($X_1 \circ \phi_{t_i} = X_{t_i}$). Les détails du calcul sont définis dans les travaux de Beg et al. [13]. Ces transformations sont utilisées pour estimer la déformation à des temps intermédiaires. Elles sont respectivement calculées en intégrant v du temps 0 au temps t_i , et en intégrant $-v$ du temps 1 au temps t_i .

Stratégie expérimentale de reconstruction de mouvement avec *Compressive Sensing*

En reprenant LDDMM tel que défini par Beg et al., les transformations *forward* (ϕ_t^{-1}) et *backward* (ϕ_t) sont estimées pour tout t entre deux images. Un algorithme est défini pour estimer le mouvement et l'image de la façon suivante :

- (1) Calcul d'une approximation grossière de X à chaque temps t_i avec *Compressive Sensing* (sans D_t). Les m premiers motifs après t_i (voir Figure 4.11) sont utilisés afin d'estimer X_{t_i} avec du *Compressive Sensing* pour chaque t_i . A cette étape, la présence de mouvement est ignorée. Comme les premiers motifs favorisent les reconstructions moins détaillées [3], une version peu résolue des objets peut être estimée.
- (2) X_{t_i} est régularisé en utilisant le débruitage par TV [35]. Bien que X_{t_i} soit plus précis en estimant le mouvement, cette méthode tend à créer de nombreux artefacts. En suivant cette observation, X_{t_i} est régularisé par TV.
- (3) Estimation de la déformation entre X_{t_i} et $X_{t_{i+1}}$ calculée grâce à LDDMM. Ce calcul nous renseigne sur une transformation *forward* et une transformation *backward* qui représentent ϕ_t . LDDMM est calculé avec une discrétisation temporelle de $t_{i+1} - t_i$ pour chaque pas de temps t_i , permettant ensuite de définir un D_t par rapport à l'ensemble des patterns (pour chaque t_i).
- (4) Définition de la déformation D_t pour chaque t_i en composant (appliquant de manière) les opérateurs *backward* pour les temps inférieurs à i et *forward* pour les temps supérieurs à i . (voir Figure 4.13)

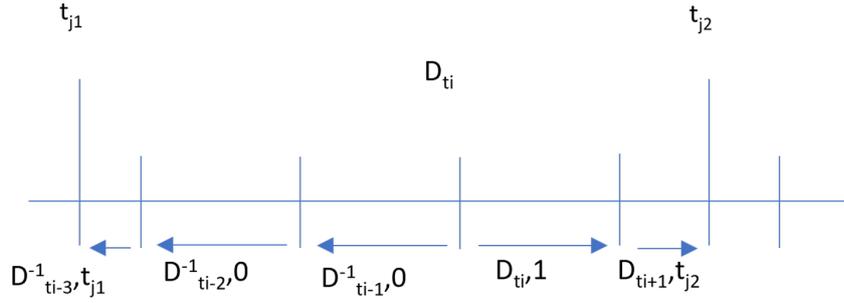


FIGURE 4.13. Exemple de composition des déformations pour une maille t_i et deux temps t_{j_1} , t_{j_2} . Pour le t_{j_1} la déformation en t_i est donc $D_{t_{i-3}, t_{j_1}}^{-1} \circ D_{t_{i-2}, 0}^{-1} \circ D_{t_{i-1}, 0}^{-1}$ et pour t_{j_2} , $D_{t_i, 1} \circ D_{t_{i+1}, t_{j_2}}$.

- (5) Une estimation de X_{t_i} est calculée pour chaque temps t_i avec *Compressive Sensing* en utilisant le D_t correspondant, en reprenant l'idée qu'une déformation peut être appliquée à la base de mesure pour reconstruire X_{t_i} plus précisément.
- (6) X_{t_i} est régularisé en utilisant le débruitage par TV [35]. De même que pour (2), les X_{t_i} sont régularisés par TV.
- (7) Tant que le critère d'arrêt n'est pas valide, il faut revenir à (2) sinon aller à (8). Pour l'instant, tant que la notion de convergence n'est pas bien établie pour cet algorithme, le critère d'arrêt correspond à un nombre d'itérations.
- (8) Renvoyer X_{t_i} et D_t . Pour finir, l'algorithme renvoie X_{t_i} pour chaque i ainsi que D_t .

Comme cet algorithme ne corrige pas juste le mouvement au cours de l'acquisition mais donne une estimation de X_{t_i} et D_t , en prenant le meilleur des X_{t_i} *a posteriori* et en lui appliquant la déformation D_t correspondante, une approximation de X pour tous les t est donnée.

Ainsi, dans l'état actuel des travaux, cet algorithme est à considérer comme étant une ouverture nouvelle vers un projet de recherche et non comme étant une solution en soit. Pour l'instant, son efficacité n'est démontrée que sur un cas particulier simplifié. Pour généraliser son utilisation, il faudrait justifier théoriquement et à minima : une garantie de convergence globale de l'algorithme ; une forme de garantie sur le *Compressive Sensing* avec une base modifiée pour prendre en compte les déformations ; borner l'erreur de reconstruction éventuellement avec et sans bruit. D'un point de vue pratique, il serait aussi intéressant d'évaluer le comportement de l'algorithme en fonction des déformations appliquées, des différents niveaux de bruits ainsi que la sensibilité des différents des paramètres.

4.2.3 Résultats dans le cas 2D

En restant dans le cas d'images 2D, quelques résultats sont présentés sur les différentes étapes de l'algorithme. L'algorithme est appliqué sur l'image définie dans la Figure 4.10 et déplacée en diagonale jusqu'à l'angle supérieur droit après quoi, le disque longe le bord droit de l'image. Nous avons choisi d'utiliser $m = 64$ motifs avec 5 mailles, ce qui donne expérimentalement de bons résultats en appliquant (1) et (2) (voir Figure 4.14) ainsi que (3) (voir Figure 4.15).

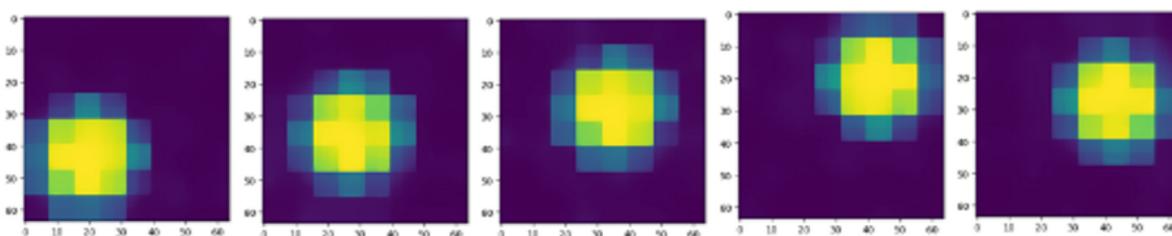


FIGURE 4.14. Estimation de la scène en ignorant le mouvement (1) et (2) pour 5 temps qui augmentent de gauche à droite.

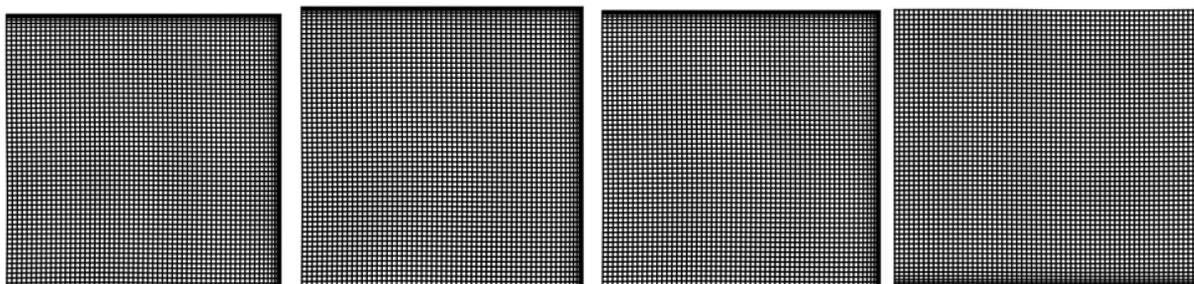


FIGURE 4.15. Estimation du mouvement entre t_i et t_{i+1} à la sortie de LDDMM

Dans la Figure 4.15, la déformation est bien retrouvée à chaque pas de temps avec une translation vers l'angle supérieur droit sur les trois premières déformations évaluées et vers le bas uniquement ensuite.

En appliquant la déformation à la base de mesure puis en estimant X_{t_i} , on obtient des résultats entachés d'une forte erreur avec de nombreux artefacts (voir Figure 4.16).

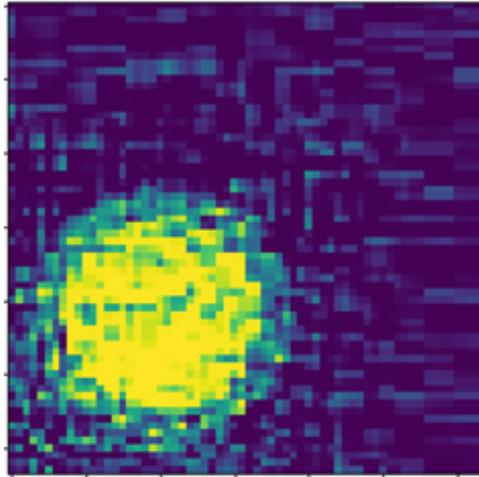


FIGURE 4.16. Estimation de X_{t_0} après estimation du mouvement avec D_t

Après 15 itérations de la boucle (2)-(7), l'algorithme est stoppé. Certaines propriétés des reconstructions sont notables. En particulier, de meilleurs résultats sont obtenus sur les premiers maillages (voir Figure 4.17) et les temps comportant des objets proches des bords sont moins bien résolus que les autres temps. Cette observation est également identifiée sur la carte d'erreur (voir Figure 4.17). L'algorithme tend à reconstruire des intensités non présentes proches des bords de l'image. En effet aux temps 0, 4 et 5 lorsque les objets sont proches des bords, les intensités entre le bord et l'image augmentent. Cet algorithme semble globalement sous-estimer la valeur des intensités. Cela est dû au décalage régularisé et parfois imprécis. Les frontières des objets sont plus susceptibles d'introduire des erreurs.

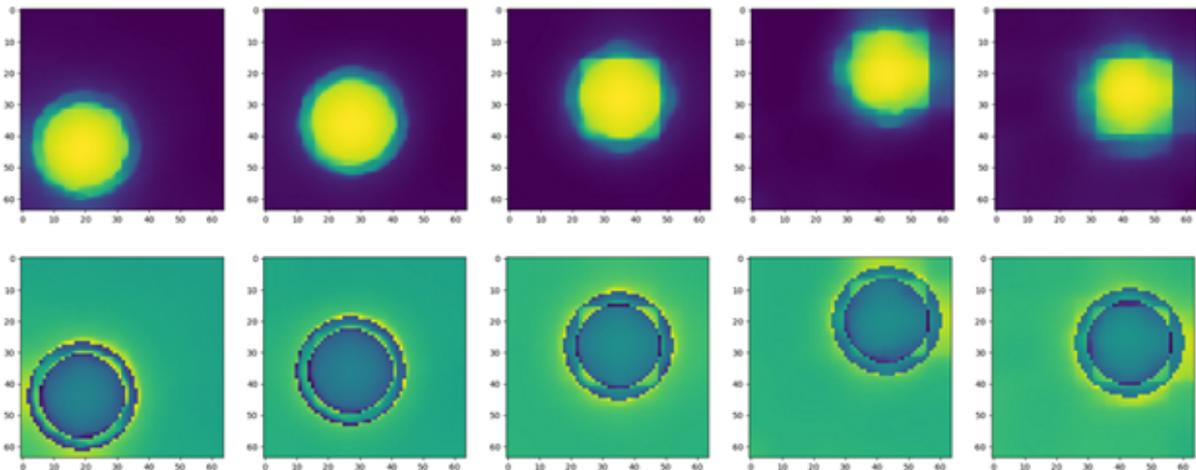


FIGURE 4.17. Résultat de l'algorithme de *Compressive Sensing* en mouvement dans un cas simple. Première ligne : Image reconstruite après 15 itérations de l'algorithme. Seconde ligne : Erreur de reconstruction (image estimée – vérité terrain), les valeurs négatives resp. positives correspondent aux zones où l'intensité de l'image estimée est inférieure resp. supérieure à l'image reconstruite.

4.3 Synthèse

Lorsqu'un mouvement est suffisamment lent par rapport à la vitesse d'acquisition, il existe déjà un certain nombre de méthodes à l'état-de-l'art permettant de l'estimer correctement. Dans le cas contraire, le *Compressive Sensing* ne peut plus être appliqué sans modification supplémentaire. Afin d'y remédier, une méthode d'estimation du mouvement a été développée. Elle est fondée sur l'algorithme LDDMM, l'ajout de régularisations ainsi que l'incohérence imparfaite entre la base de mesure et la base parcimonieuse conduisant à une répartition inégale de l'information. La méthode, nécessitant tout de même des garanties théoriques, donne de nombreuses perspectives intéressantes pour le développement de méthodes de *Compressive Sensing* en mouvement. Elle sera mise en œuvre et testée dans la perspective de prochains travaux de recherche appliquée en interne ONERA.

Conclusion

Tout le long de ce manuscrit, nous avons démontré qu’il est possible d’augmenter la résolution latérale de plans focaux 3D pour l’acquisition LiDAR 3D en utilisant le *Compressive Sensing* avec une latence aussi faible que possible sous un ensemble de contraintes. Les acquisitions sont à très longue portée (supérieure à 10 km). Le système met en œuvre une matrice de capteurs GmAPD en temps limité avec un faible SWaP-C et un niveau de signal très faible (inférieur à un photon en moyenne). Les travaux réalisés sont déclinés selon 2 axes de recherche principaux :

1. **Développement de la chaîne de traitement CRISPS-3D (*Compressive Reconstruction of Images for Single Photon System in 3D*) pour augmenter la résolution de plans focaux 3D dans le cas statique.**

(Chapitre 2 et 3)

2. **Généralisation à la chaîne complète de traitement pour des applications en mouvement (porteur et cibles).**

(Chapitre 4)

Afin de réaliser notre objectif, un certain nombre d’outils et de modules sont développés pour la chaîne CRISPS-3D. Tout d’abord une extension des travaux de Liu et *al.* [91] permet de compenser l’effet du temps mort lorsque le capteur n’est pas saturé dans le temps. La compensation de ces effets permet de considérer une gamme dynamique plus élevée et de minimiser l’effet des artefacts qui sont liés à la reconstruction du *Compressive Sensing* à chaque pas de temps.

Ensuite, une nouvelle stratégie de débruitage est définie en exploitant les spécificités d’un système LiDAR 3D : en particulier, la différence entre la fréquence d’impulsion de la source laser et la fréquence d’acquisition de la caméra. En comparant statistiquement avec un test non-paramétrique ces acquisitions, le support de l’onde complète est estimé. En ne gardant que les pas de temps dans le support, le bruit s’en retrouve fortement diminué. Dans CRISPS-3D, une implémentation relativement standard du *Compressive Sensing* a été implémentée dans une version optimisée.

Enfin, un algorithme de déconvolution est introduit dans la chaîne CRISPS-3D afin d’inverser la convolution de la réponse impulsionnelle avec la scène. Pour tirer parti de la forte parcimonie en profondeur, le problème de déconvolution est formulé comme un problème de *matching pursuit*. La nouveauté dans ce cadre n’est pas l’utilisation du *matching pursuit* pour la déconvolution mais plutôt le questionnement sur son intégration dans la chaîne CRISPS-3D avec du *Compressive Sensing*. Fondée sur l’idée que le retour obtenu devrait être régulier, une amélioration de cet algorithme est proposée en régularisant le signal. Cependant, bien que cela améliore la reconstruction de certains points, l’effet sur l’ensemble du nuage de points est faible.

Dans une deuxième partie plus expérimentale, l’estimation du mouvement 3D entre deux

nuages de points est défini *via* le prisme d’algorithmes de type *scene flow* basés sur le transport optimal. Afin d’intégrer cette idée dans notre chaîne de traitement CRISPS-3D, il faut considérer des mouvements de cibles et/ou du porteur faibles par rapport à la fréquence d’acquisition.

L’estimation et la correction de la dynamique dans CRISPS-3 avec des mouvements suffisamment importants pour que la scène soit modifiée entre l’acquisition de deux éléments de la base de mesure est un challenge. En ajoutant des hypothèses pour simplifier le problème, un algorithme d’estimation de la dynamique est défini. Cette méthode nécessite toutefois le développement d’un cadre théorique pour être justifiée mais elle ouvre la voie à l’application du *Compressive Sensing* sur des scènes se déformant rapidement au cours du temps, plus rapidement que la fréquence d’acquisition. On se rapproche ainsi d’un cadre opérationnel qu’il conviendra de valider dans de prochains travaux à court terme.

Perspectives

Les travaux conduits dans cette thèse nous ont permis d’identifier plusieurs pistes d’amélioration des outils développés. Il est évident que l’introduction du *Compressive Sensing* pour l’étude de signaux LiDAR 3D est un bénéfice pour améliorer la résolution latérale de ces imageurs tout en maintenant une bonne résolution en profondeur. L’état-de-l’art sur la technologie des capteurs impose d’introduire des traitements spécifiques qui devront être implémentés dans les futures générations de systèmes LiDAR 3D. Ils devront notamment permettre un traitement proche du temps réel. Des expériences à plus ou moins longs termes permettront de valider les outils développés dans cette thèse. Nos travaux ouvrent les portes vers de nombreuses perspectives qui peuvent être résumées comme suit :

- Au cours de cette thèse, un ensemble de méthodes et d’algorithmes ont été définis et développés. L’ensemble des tests ont été réalisés sur des données simulées, une première perspective à court terme est de valider l’ensemble des travaux sur des données dites « terrains » en développant un imageur 3D intégrant une acquisition par *Compressive Sensing*. Cet instrument est en cours de développement à l’ONERA et fournira bientôt les premières images expérimentales.
- Plusieurs éléments ont été négligés pour définir notre chaîne de traitement CRISPS-3D. En particulier, les effets liés aux turbulences atmosphériques ou aux aérosols présents sur la ligne de visée dont la prise en compte numérique est un véritable challenge en soit. De même, une étude précise de l’effet des bruits tels que l’*afterpulse* et le *cross-talk* serait pertinente.
- Dans le chapitre 3, le choix a été fait de déconvoluer le signal *a posteriori* de la résolution du *Compressive Sensing*. Cependant, déconvoluer le signal en même temps que le DTE en amont du *Compressive Sensing* ou après le DTE et en même temps

que le *Compressive Sensing* sont deux méthodes prometteuses. Elles nécessitent de développer une approche très différente basée par exemple sur une modélisation Bayésienne impliquant ses propres défis en termes de gestion de la mémoire et des temps d'exécution ou des apports théoriques supplémentaires pour appliquer le *Compressive Sensing* directement en 3D avec des bases de mesures modifiées.

- Dans le cadre Bayésien, il est intéressant d'unifier les différentes modélisations développées sur le DTE sur le *Compressive Sensing* avec BCS et sur la reconstruction de nuages de points 3D d'acquisition par des capteurs GmAPD afin d'obtenir un modèle qui permettrait de résoudre l'ensemble de la reconstruction obtenue avec CRISPS-3D. Une implémentation spécifique est probablement nécessaire à la façon de RT3D, un algorithme *plug-and-play* développé pour Manipop.
- Pour la mise en œuvre du *Compressive Sensing* en mouvement, notre approche donne des résultats prometteurs mais nécessite d'être validée théoriquement sur plusieurs points dont notamment sur la garantie de reconstruction avec des bases déformées ainsi que sur la convergence de l'algorithme global avec éventuellement une erreur bornée.
- Du fait d'un manque de métriques adaptées à l'identification (voir Section 1.1.1), nous avons évalué quantitativement les étapes de la chaîne de traitement CRISPS-3D et qualitativement les nuages de points. La définition d'une nouvelle métrique équivalente au critère de Johnson est nécessaire. Il faudra l'adapter aux nuages de points provenant d'acquisitions LiDAR 3D (*2D-manifolds* dans un espace 3D, voir Section 3.1.1). Une fois définie, la métrique permettra une évaluation quantitative des résultats obtenus pour mieux se comparer aux méthodes de l'état de l'art. En suivant, cette logique, nous pourrons aussi réaliser une analyse de sensibilité, notamment vis-à-vis des paramètres.
- À plus longs termes, nous identifions des travaux complémentaires. Nous avons fait d'autres hypothèses dans nos travaux, notamment sur l'absence d'occlusions ou de rotations ainsi que sur l'utilisation de deux dimensions plutôt que trois. Étendre la méthode à ces cas est nécessaire pour un usage proche de conditions opérationnelles.
- Un dernier point qui n'a été que très peu abordé dans ce manuscrit concerne l'usage des algorithmes d'apprentissage de type *machine learning*. En particulier, l'apprentissage profond est de plus en plus fréquent associé au *Compressive Sensing*. Il est important d'évaluer son apport sur le sujet de la reconstruction 3D avec et sans mouvement.

Annexe A

Équivalence des corrections du DTE

Dans cette annexe, l'objectif est de démontrer que la correction du DTE défini dans Section 2.2.1 est équivalente au minimum de vraisemblance ainsi qu'à la formulation de Coates [44, 45]. Pour commencer, la vraisemblance est calculée :

$$\begin{aligned}
 \ln L(Y_0, \dots, Y_{N_{bins}} \mid z_0, \dots, z_{N_{acq}}) &= \sum_{l=0}^{N_{acq}} \ln(\mathbb{P}(z = z_l)) \\
 &= \sum_{l=0}^{N_{acq}} \ln \left((1 - e^{-Y_{z_l}}) \prod_{j=0}^{z_l-1} e^{-Y_j} \right) \\
 &= \sum_{l=0}^{N_{bins}} H_l \left[\ln \left((1 - e^{-Y_l}) \prod_{j=0}^{l-1} e^{-Y_j} \right) \right] \\
 &= \sum_{l=0}^{N_{bins}} H_l \left[\ln(1 - e^{-Y_l}) + \sum_{j=0}^{l-1} \ln(e^{-Y_j}) \right] \\
 &= \sum_{l=0}^{N_{bins}} H_l \left[\ln(1 - e^{-Y_l}) - \sum_{j=0}^{l-1} Y_j \right]
 \end{aligned}$$

Ensuite, la vraisemblance est maximisée par rapport au paramètre Y , donc son gradient est calculé :

$$\begin{aligned}
 \frac{\partial \ln L(Y_0, \dots, Y_{N_{bins}} \mid z_0, \dots, z_{N_{acq}})}{\partial Y_k} &= \frac{\partial}{\partial Y_k} \left(\sum_{l=0}^{N_{bins}} H_l \left[\ln(1 - e^{-Y_l}) - \sum_{j=0}^{l-1} Y_j \right] \right) \\
 &= \sum_{l=0}^{N_{bins}} H_l \frac{\partial}{\partial Y_k} \left[\ln(1 - e^{-Y_l}) - \sum_{j=0}^{l-1} Y_j \right] \\
 &= \sum_{l=0}^{N_{bins}} H_l \left[\frac{\partial}{\partial Y_k} (\ln(1 - e^{-Y_l})) - \sum_{j=0}^{l-1} \frac{\partial}{\partial Y_k} (Y_j) \right] \\
 &= \sum_{l=0}^{N_{bins}} H_l \left[\frac{e^{-Y_l}}{1 - e^{-Y_l}} \mathbb{1}_{k=l} - \sum_{j=0}^{l-1} \mathbb{1}_{j=k} \right] \\
 &= \frac{H_k e^{-Y_k}}{1 - e^{-Y_k}} - \sum_{l=0}^{N_{bins}} H_l \sum_{j=0}^{l-1} \mathbb{1}_{j=k}
 \end{aligned}$$

$$\begin{aligned}
&= \frac{H_k}{e^{Y_k} - 1} - \sum_{l=0}^{N_{bins}} H_l \mathbb{1}_{l>k} \\
&= \frac{H_k}{e^{Y_k} - 1} - \sum_{l=k+1}^{N_{bins}} H_l \\
&= \frac{H_k}{e^{Y_k} - 1} - \sum_{l=k+1}^{N_{bins}} H_l
\end{aligned}$$

Nous cherchons Y tel que le gradient soit nul :

$$\begin{aligned}
&\frac{\partial \ln L(Y_0, \dots, Y_{N_{bins}} \mid z_0, \dots, z_{N_{acq}})}{\partial Y_k} = 0 \\
&\Leftrightarrow \frac{H_k}{e^{Y_k} - 1} - \sum_{l=k+1}^{N_{bins}} H_l = 0 \\
&\Leftrightarrow \frac{H_k}{e^{Y_k} - 1} = \sum_{l=k+1}^{N_{bins}} H_l \\
&\Leftrightarrow e^{Y_k} - 1 = H_k \left[\sum_{l=k+1}^{N_{bins}} H_l \right]^{-1} \\
&\Leftrightarrow e^{Y_k} = H_k \left[\sum_{l=k+1}^{N_{bins}} H_l \right]^{-1} + 1 \\
&\Leftrightarrow Y_k = \ln \left(1 + H_k \left[\sum_{l=k+1}^{N_{bins}} H_l \right]^{-1} \right) \\
&\Leftrightarrow Y_k = \ln \left(1 + H_k \left[1 - \sum_{l=0}^k H_l \right]^{-1} \right)
\end{aligned}$$

Le résultat est bien un maximum :

$$\begin{aligned}
&\frac{\partial \ln L(Y_0, \dots, Y_{N_{bins}} \mid z_0, \dots, z_{N_{acq}})}{\partial^2 Y_k} = \frac{\partial}{\partial Y_k} \left(\frac{H_k}{e^{Y_k} - 1} - \sum_{l=k+1}^{N_{bins}} H_l \right) \\
&= \frac{\partial}{\partial Y_k} \left(\frac{H_k}{e^{Y_k} - 1} \right) = -\frac{H_k e^{Y_k}}{(e^{Y_k} - 1)^2} \leq 0
\end{aligned}$$

Ainsi, en reprenant le maximum de vraisemblance, l'estimateur de Coates [44, 45] et l'estimateur définit dans la Section 2.2.1 :

$$\begin{aligned}
Y_k^{MLE} &= \ln \left(1 + H_k \left[1 - \sum_{l=0}^k H_l \right]^{-1} \right) \\
Y_k^{Coates} &= -\ln \left(1 - H_k \left[1 - \sum_{l=0}^{k-1} H_l \right]^{-1} \right) \\
\widehat{Y}_k &= -\ln \left(1 - H_k e^{\sum_{j=0}^{k-1} \widehat{Y}_j} \right)
\end{aligned}$$

Tous ces estimateurs sont strictement équivalents, on montre d'abord que $\widehat{Y}_k = Y_k^{Coates}$:

$$\begin{aligned}
\widehat{Y}_k &= -\ln \left(1 - H_k \left[e^{-\sum_{j=0}^{k-1} \widehat{Y}_j} \right]^{-1} \right) \\
&= -\ln \left(1 - H_k \left[e^{-\sum_{j=0}^{k-1} \widehat{Y}_j} + \sum_{l=0}^{k-2} e^{-\sum_{j=0}^l \widehat{Y}_j} - \sum_{l=0}^{k-2} e^{-\sum_{j=0}^l \widehat{Y}_j} \right]^{-1} \right) \\
&= -\ln \left(1 - H_k \left[\sum_{l=0}^{k-1} e^{-\sum_{j=0}^l \widehat{Y}_j} - \sum_{l=1}^{k-1} e^{-\sum_{j=0}^{l-1} \widehat{Y}_j} \right]^{-1} \right) \\
&= -\ln \left(1 - H_k \left[e^{-\widehat{Y}_0} + \sum_{l=1}^{k-1} \left(e^{-\sum_{j=0}^l \widehat{Y}_j} - e^{-\sum_{j=0}^{l-1} \widehat{Y}_j} \right) \right]^{-1} \right) \\
&= -\ln \left(1 - H_k \left[e^{-\widehat{Y}_0} + \sum_{l=1}^{k-1} \left(e^{-\sum_{j=0}^l \widehat{Y}_j} - e^{-\sum_{j=0}^{l-1} \widehat{Y}_j} \right) \right]^{-1} \right) \\
&= -\ln \left(1 - H_k \left[e^{-\widehat{Y}_0} + \sum_{l=1}^{k-1} e^{-\sum_{j=0}^{l-1} \widehat{Y}_j} \left(e^{-\widehat{Y}_l} - 1 \right) \right]^{-1} \right) \\
&= -\ln \left(1 - H_k \left[1 - H_0 + \sum_{l=1}^{k-1} e^{-\sum_{j=0}^{l-1} \widehat{Y}_j} \left(1 - H_l \times e^{\sum_{j=0}^{l-1} \widehat{Y}_j} - 1 \right) \right]^{-1} \right) \\
&= -\ln \left(1 - H_k \left[1 - H_0 - \sum_{l=1}^{k-1} H_l \times e^{\sum_{j=0}^{l-1} \widehat{Y}_j} \times e^{-\sum_{j=0}^{l-1} \widehat{Y}_j} \right]^{-1} \right) \\
&= -\ln \left(1 - H_k \left[1 - \sum_{l=0}^{k-1} H_l \right]^{-1} \right) \\
&= Y_k^{Coates}
\end{aligned}$$

Enfin on montre que $Y_k^{MLE} = Y_k^{Coates}$:

$$\begin{aligned}
Y_k^{MLE} &= \ln \left(1 + H_k \left[1 - \sum_{l=0}^k H_l \right]^{-1} \right) \\
&= \ln \left(\left[1 - \sum_{l=0}^k H_l + H_k \right] \left[1 - \sum_{l=0}^k H_l \right]^{-1} \right) \\
&= -\ln \left(\left[1 - \sum_{l=0}^{k-1} H_l \right]^{-1} \left[1 - \sum_{l=0}^k H_l \right] \right) \\
&= -\ln \left(\left[1 - \sum_{l=0}^{k-1} H_l \right]^{-1} \left[1 - \sum_{l=0}^{k-1} H_l - H_k \right] \right) \\
&= -\ln \left(1 - H_k \left[1 - \sum_{l=0}^{k-1} H_l \right]^{-1} \right) \\
&= Y_k^{Coates}
\end{aligned}$$

D'où $Y_k^{MLE} = Y_k^{Coates} = \widehat{Y}_k$

Bibliographie

- [1] feature visualization. <https://distill.pub/2017/feature-visualization/>. Accessed: 2023-08-01.
- [2] Is lidar dangerous for our eyes? <https://photonicsreport.com/blog/is-lidar-dangerous-for-our-eyes/>. Accessed: 2023-08-01.
- [3] B. Adcock, A. C. Hansen, C. Poon, and B. Roman. Breaking the coherence barrier: A new theory for compressed sensing. In Forum of mathematics, sigma, volume 5, page e4. Cambridge University Press, 2017.
- [4] M. A. Ainslie. Principles of sonar performance modelling, volume 707. Springer, 2010.
- [5] D. L. Aleva, G. G. Kuperman, and A. F. R. L. W.-P. A. O. H. E. DIRECTORATE. Effects of scene modulation image blur and noise upon human target acquisition performance. Technical report, Technical report, DTIC Document, 1997.
- [6] Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin. Robust bayesian target detection algorithm for depth imaging from sparse single-photon data. IEEE Transactions on Computational Imaging, 2(4):456–467, Dec 2016.
- [7] G. Anna, L. Hespel, N. Riviere, D. Hamoir, and B. Tanguy. Physical modelling of point-cloud (3d) and full-wave-form (4d) laser imaging. In Electro-Optical Remote Sensing, Photonic Technologies, and Applications IV, volume 7835, pages 95 – 109. SPIE, Oct 2010.
- [8] Y. Arjoun, N. Kaabouch, H. El Ghazi, and A. Tamtaoui. Compressive sensing: Performance comparison of sparse recovery algorithms. In 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), volume 1, pages 1–7. IEEE, Jan 2017.
- [9] B. Aull. Geiger-mode avalanche photodiode arrays integrated to all-digital cmos circuits. Sensors, 16(4):495, Apr 2016.
- [10] S. D. Babacan, R. Molina, and A. K. Katsaggelos. Bayesian compressive sensing using laplace priors. IEEE Transactions on Image Processing, 19(1):53–63, 2010.
- [11] J. Barzilai and J. M. Borwein. Two-point step size gradient methods. IMA journal of numerical analysis, 8(1):141–148, 1988.
- [12] W. Becker, A. Bergmann, G. L. Biscotti, and A. Rueck. Advanced time-correlated single photon counting techniques for spectroscopy and imaging in biomedical systems. In Commercial and Biomedical Applications of Ultrafast Lasers IV, volume 5340, pages 104–112. SPIE, 2004.
- [13] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International journal of computer vision, 61:139–157, 2005.

- [14] E. Bernard. Comparaison théorique et expérimentale des performances apres traitement de l'imagerie active et de l'IR2 en conditions dégradées. PhD thesis, UNIVERSITE DE TOULOUSE III-PAUL SABATIER, 2015.
- [15] E. Bernard, N. Riviere, M. Renaudat, P. Guiset, M. Pealat, and E. Zenou. Experiments and models of active and thermal imaging under bad weather conditions. In Electro-Optical Remote Sensing, Photonic Technologies, and Applications VII; and Military Applications in Hyperspectral Imaging and High Spatial Resolution Sensing, volume 8897, pages 14 – 25. SPIE, Oct 2013.
- [16] M. Bland. An introduction to medical statistics. Oxford university press, 2015.
- [17] T. Blu, P.-L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot. Sparse sampling of signal innovations. IEEE Signal Processing Magazine, 25(2):31–40, March 2008.
- [18] T. Blumensath and M. E. Davies. Iterative thresholding for sparse approximations. Journal of Fourier analysis and Applications, 14(5-6):629–654, 2008.
- [19] T. Blumensath and M. E. Davies. Iterative hard thresholding for compressed sensing. Applied and Computational Harmonic Analysis, 27(3):265–274, 2009.
- [20] T. Blumensath, M. E. Davies, G. Rilling, Y. Eldar, and G. Kutyniok. Greedy algorithms for compressed sensing., 2012.
- [21] C. P. Bradley, S. S. Mukherjee, A. D. Reinhardt, P. F. McManamon, A. O. Lee, and V. Dhulla. 3d imaging with 128×128 eye safe ingaas p-i-n lidar camera. In Laser Radar Technology and Applications XXIV, volume 11005, pages 273 – 279. SPIE, May 2019.
- [22] K. Bredies, D. A. Lorenz, and P. Maass. Mathematical concepts of multiscale smoothing. Applied and Computational Harmonic Analysis, 19(2):141–161, 2005.
- [23] K. Bredies and H. K. Pikkarainen. Inverse problems in spaces of measures. ESAIM: Control, Optimisation and Calculus of Variations, 19(1):190–218, 2013.
- [24] X. Briottet, L. Hespel, and N. Riviere. Imagerie laser 3d à plan focal. Techniques de l'ingénieur, Oct 2020.
- [25] A. Bucchianico. Combinatorics, computer algebra and the wilcoxon-mann-whitney test. Journal of Statistical Planning and Inference, 79(2):349–364, Jul 1999.
- [26] K. by Lumibird. Pefl-kult pulsed erbium fiber laser brochure v1.0, Feb 2019.
- [27] B. R. Call, D. Kelley, D. G. Fried, C. Reichert, K. Reichel-Vischi, and A. Eldredge. Low SWaP, commercially-available Geiger-mode lidar system. In G. W. Kameron, L. A. Magruder, and M. D. Turner, editors, Laser Radar Technology and Applications XXVII, volume 12110, page 1211008. International Society for Optics and Photonics, SPIE, 2022.
- [28] E. Candes, J. Romberg, et al. l1-magic: Recovery of sparse signals via convex programming. URL: www. acm. caltech. edu/l1magic/downloads/l1magic. pdf, 4(14):16, 2005.

- [29] E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory, 52(2):489–509, Jan 2006.
- [30] E. J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? IEEE Transactions on Information Theory, 52(12):5406–5425, Dec 2006.
- [31] E. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. Inverse Problems, 23(3):969–985, Apr 2007.
- [32] E. J. Candès, J. K. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. Communications on Pure and Applied Mathematics, 59(8):1207–1223, Mar 2006.
- [33] T. Carlsson, O. Steinvall, and D. Letalick. Signature simulation and signal analysis for 3-d laser radar. Sweden: Swedish Defence Research Agency, 2001.
- [34] M. Carvalho, B. Le Saux, P. Trouvé-Peloux, A. Almansa, and F. Champagnat. Deep depth from defocus: how can defocus blur improve 3d estimation using dense neural networks? In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, pages 0–0, 2018.
- [35] A. Chambolle. An algorithm for total variation minimization and applications. Journal of Mathematical imaging and vision, 20:89–97, 2004.
- [36] R. Chartrand and V. Staneva. Restricted isometry properties and nonconvex compressive sensing. Inverse Problems, 24(3):035020, may 2008.
- [37] R. Chartrand and W. Yin. Iteratively reweighted algorithms for compressive sensing. In 2008 IEEE international conference on acoustics, speech and signal processing, pages 3869–3872. IEEE, 2008.
- [38] A. F. Chase, D. Z. Chase, J. F. Weishampel, J. B. Drake, R. L. Shrestha, K. C. Slatton, J. J. Awe, and W. E. Carter. Airborne lidar, archaeology, and the ancient maya landscape at caracol, belize. Journal of Archaeological Science, 38(2):387–398, 2011.
- [39] A. Chauve, C. Mallet, F. Bretar, S. Durrieu, M. Pierrot-Deseilligny, and W. Puech. Processing Full-Waveform Lidar Data: Modelling Raw Signals. In ISPRS Workshop Laser Scanning and SilviLaser (LS SL), volume 36 (Part 3/W52), pages 102–107, Espoo, Finland, Sept. 2007. [Departement_IRSTEA]DS [TR1_IRSTEA]METHODO / SYNERGIE.
- [40] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. SIAM review, 43(1):129–159, 2001.
- [41] Z. Chen, A. Basarab, and D. Kouamé. Compressive deconvolution in medical ultrasound imaging. IEEE Transactions on Medical Imaging, 35(3):728–737, March 2016.

- [42] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. Scaling algorithms for unbalanced optimal transport problems. Mathematics of Computation, 87(314):2563–2609, 2018.
- [43] K. Chrzanowski. Review of night vision technology. Opto-Electronics Review, 21(2):153–181, 2013.
- [44] P. B. Coates. The correction for photon ‘pile-up’ in the measurement of radiative lifetimes. Journal of Physics E: Scientific Instruments, 1(8):878, Aug 1968.
- [45] P. B. Coates. Pile-up corrections in the measurement of lifetimes. Journal of Physics E: Scientific Instruments, 5(2):148, Feb 1972.
- [46] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa. Avalanche photodiodes and quenching circuits for single-photon detection. Appl. Opt., 35(12):1956–1976, Apr 1996.
- [47] A. Coyac, L. Hespel, N. Riviere, and X. Briottet. Performance assessment of simulated 3d laser images using geiger-mode avalanche photo-diode: tests on simple synthetic scenarios. In Electro-Optical and Infrared Systems: Technology and Applications XII; and Quantum Information Science and Technology, volume 9648, pages 22 – 31. SPIE, Oct 2015.
- [48] A. Coyac, N. Riviere, L. Hespel, and X. Briottet. Comparison of simulated and experimental 3d laser images using a gmapd array: application to long range detection. In Laser Radar Technology and Applications XXI, volume 9832, pages 46 – 53. SPIE, May 2016.
- [49] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. Advances in neural information processing systems, 26, 2013.
- [50] G. Davis and M. Mallat, S. and Avellaneda. Adaptive greedy approximations. Constructive Approximation, 13(1):57–98, Mar 1997.
- [51] I. De Gélis, S. Lefèvre, T. Corpetti, T. Ristorcelli, C. Thénoz, and P. Lassalle. Benchmarking change detection in urban 3d point clouds. In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pages 3352–3355, July 2021.
- [52] D. Donoho. Compressed sensing. IEEE Transactions on Information Theory, 52(4):1289–1306, 2006.
- [53] D. L. Donoho and M. Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization. Proceedings of the National Academy of Sciences, 100(5):2197–2202, 2003.
- [54] M. Douglass. Dmd reliability: a mems success story. In Reliability, Testing, and Characterization of Memes/Moems II, volume 4980, pages 1–11. SPIE, 2003.
- [55] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single-pixel imaging via compressive sampling. IEEE Signal Processing Magazine, 25(2):83–91, Mar 2008.

- [56] V. Duval and G. Peyré. Sparse regularization on thin grids i: the lasso. Inverse Problems, 33(5):055008, 2017.
- [57] G. d’Ockham. Quaestiones et decisiones in quatuor libros sententiarum cum centilogo theologico. Livre, 2, 1319.
- [58] M. P. Edgar, S. Johnson, D. B. Phillips, and M. J. Padgett. Real-time computational photon-counting lidar. Optical Engineering, 57(3):1 – 7, Dec 2017.
- [59] M. P. Edgar, M.-J. Sun, G. M. Gibson, G. C. Spalding, D. B. Phillips, and M. J. Padgett. Real-time 3d video utilizing a compressed sensing time-of-flight single-pixel camera. In Optical Trapping and Optical Micromanipulation XIII, volume 9922, pages 171 – 178. SPIE, Sep 2016.
- [60] T. Felix, Martinez. Maritime protection of critical infrastructure assets in the Campeche sound. PhD thesis, Naval Postgraduate School, Dec 2005.
- [61] J. Feydy, B. Charlier, F.-X. Vialard, and G. Peyré. Optimal transport for diffeomorphic registration. In Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20, pages 291–299. Springer, 2017.
- [62] A. S. for Testing and Materials. Standard tables for reference solar spectral irradiances: Direct normal and hemispherical on 37° tilted surface (standard g173-03). 2020.
- [63] H. Gao, B. Cheng, J. Wang, K. Li, J. Zhao, and D. Li. Object classification using cnn-based fusion of vision and lidar in autonomous vehicle environment. IEEE Transactions on Industrial Informatics, 14(9):4224–4231, Sep. 2018.
- [64] Y.-x. Gao, H. Sun, T. Zhang, and L. Du. Research on compressive sensing reconstruction algorithm based on total variation model. In LIDAR Imaging Detection and Target Recognition 2017, volume 10605, page 1060515. International Society for Optics and Photonics, 2017.
- [65] R. Garg and R. Khandekar. Gradient descent with sparsification: an iterative algorithm for sparse recovery with restricted isometry property. In Proceedings of the 26th annual international conference on machine learning, pages 337–344, 2009.
- [66] P. Gatt, S. Johnson, and T. Nichols. Geiger-mode avalanche photodiode lidar receiver performance characteristics and detection statistics. Appl. Opt., 48(17):3261–3276, Jun 2009.
- [67] G. M. Gibson, S. D. Johnson, and M. J. Padgett. Single-pixel imaging 12 years on: a review. Opt. Express, 28(19):28190–28208, Sep 2020.
- [68] A. Goldwurm. Estimation de dynamique de nuages de points 3d. Rapport de Master 2, disponible à l’ONERA, 2022.

- [69] K. J. Gordon, P. A. Hiskett, and R. A. Lamb. Advanced 3d imaging lidar concepts for long range sensing. In Advanced Photon Counting Techniques VIII, volume 9114, pages 49 – 55. SPIE, Jun 2014.
- [70] A. Halimi, R. Tobin, A. McCarthy, S. McLaughlin, and G. S. Buller. Restoration of multilayered single-photon 3d lidar images. In 2017 25th European Signal Processing Conference (EUSIPCO), pages 708–712. IEEE, 2017.
- [71] F. Heide, S. Diamond, D. B. Lindell, and G. Wetzstein. Sub-picosecond photon-efficient 3d imaging using single-photon sensors. Scientific Reports, 8(1):17726, Dec 2018.
- [72] S. Hernandez-Marin, A. M. Wallace, and G. J. Gibson. Bayesian analysis of lidar signals with multiple returns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(12):2170–2180, 2007.
- [73] L. Hespel, N. Riviere, M. Fraces, P. E. Dupouy, A. Coyac, P. Barillot, S. Fauquex, A. Plyer, M. Tauvy, M. Jacquart, I. Vin, E. Nascimben, C. Perez, J. P. Velayguet, and D. Gorce. 2d and 3d flash laser imaging for long-range surveillance in maritime border security: detection and identification for counter uas applications. In Laser Radar Technology and Applications XXII, volume 10191, pages 52 – 62. SPIE, May 2017.
- [74] C. Higham, R. Murray-Smith, M. Padgett, and M. Edgar. Deep learning for real-time single-pixel video. Scientific Reports, 8(1):2369, Feb 2018.
- [75] M. A. Hofton, J.-B. Minster, and J. B. Blair. Decomposition of laser altimeter waveforms. IEEE Transactions on geoscience and remote sensing, 38(4):1989–1996, 2000.
- [76] B. K. Horn and B. G. Schunck. Determining optical flow. Artificial intelligence, 17(1-3):185–203, 1981.
- [77] G. A. Howland, P. B. Dixon, and J. C. Howell. Photon-counting compressive sensing laser radar for 3d imaging. Appl. Opt., 50(31):5917–5920, Nov 2011.
- [78] G. A. Howland, D. J. Lum, M. R. Ware, and J. C. Howell. Photon counting compressive depth mapping. Opt. Express, 21(20):23822–23837, Oct 2013.
- [79] Y. Hu, Q. Chen, S. Feng, and C. Zuo. Microscopic fringe projection profilometry: A review. Optics and Lasers in Engineering, 135(1):106192, Dec 2020.
- [80] J. Iglhaut, C. Cabo, S. Puliti, L. Piermattei, J. O’Connor, and J. Rosette. Structure from motion photogrammetry in forestry: A review. Current Forestry Reports, 5:155–168, 2019.
- [81] M. A. Itzler, M. Entwistle, M. Owens, K. Patel, X. Jiang, K. Slomkowski, S. Rangwala, P. F. Zalud, T. Senko, J. Tower, and J. Ferraro. Geiger-mode avalanche photodiode focal plane arrays for three-dimensional imaging lidar. In Infrared Remote Sensing and Instrumentation XVIII, volume 7808, pages 75 – 88. SPIE, Aug 2010.

- [82] M. A. Itzler, U. Krishnamachari, Q. Chau, X. Jiang, M. Entwistle, M. Owens, and K. Slomkowski. Statistical analysis of dark count rate in geiger-mode apd fpas. In Electro-Optical Remote Sensing, Photonic Technologies, and Applications VIII; and Military Applications in Hyperspectral Imaging and High Spatial Resolution Sensing II, volume 9250, pages 1 – 12. SPIE, Oct 2014.
- [83] L. J. Hornbeck. From cathode rays to digital micromirrors : A history of electronic projection display technology. TI Technical Journal, 1998.
- [84] S. Ji, Y. Xue, and L. Carin. Bayesian compressive sensing. IEEE Transactions on signal processing, 56(6):2346–2356, 2008.
- [85] J. Johnson. Analysis of Image Forming Systems. In R. B. Johnson and W. L. Wolfe, editors, Selected papers on infrared design. Part I and II, volume 513 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, page 761, Jan. 1985.
- [86] K. H. Jones and J. N. Gross. Reducing size, weight, and power (swap) of perception systems in small autonomous aerial systems. In 14th AIAA Aviation Technology, Integration, and Operations Conference, page 2705, 2014.
- [87] R. Leroy, P. Trouvé-Peloux, F. Champagnat, B. L. Saux, and M. Carvalho. Pix2point: Learning outdoor 3d using sparse point clouds and optimal transport, 2021.
- [88] C. Li. An efficient algorithm for total variation regularization with applications to the single pixel camera and compressive sensing. PhD thesis, 2010.
- [89] C. Li. Compressive Sensing for 3D Data Processing Tasks Applications, Models and Algorithms. PhD thesis, Rice University, Apr 2011.
- [90] Z.-P. Li, X. Huang, Y. Cao, B. Wang, Y.-H. Li, W. Jin, C. Yu, J. Zhang, Q. Zhang, C.-Z. Peng, F. Xu, and J.-W. Pan. Single-photon computational 3d imaging at 45km. Photon. Res., 8(9):1532–1540, Sep 2020.
- [91] X. Liu, Y. Ma, S. Li, J. Yang, Z. Zhang, and X. Tian. Photon counting correction method to improve the quality of reconstructed images in single photon compressive imaging systems. Opt. Express, 29(23):37945–37961, Nov 2021.
- [92] X. Liu, J. Shi, L. Sun, Y. Li, J. Fan, and G. Zeng. Photon-limited single-pixel imaging. Opt. Express, 28(6):8132–8144, Mar 2020.
- [93] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In IJCAI’81: 7th international joint conference on Artificial intelligence, volume 2, pages 674–679, 1981.
- [94] C. Mallet and F. Bretar. Full-waveform topographic lidar: State-of-the-art. ISPRS Journal of Photogrammetry and Remote Sensing, 64(1):1–16, 2009.
- [95] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. The Annals of Mathematical Statistics, 18(1):50 – 60, Mar 1947.

- [96] Y. C. E. Mark A. Davenport, Marco F. Duarte and G. Kutyniok. Introduction to Compressed sensing. Cambridge University Press, 2011.
- [97] E. Marques, Crespo, N. Maciel, L. Naviner, H. Cai, and J. Yang. A review of sparse recovery algorithms. IEEE Access, 7:1300–1322, 2019.
- [98] A. McCarthy, N. J. Krichel, N. R. Gemmell, X. Ren, M. G. Tanner, S. N. Dorenbos, V. Zwiller, R. H. Hadfield, and G. S. Buller. Kilometer-range, high resolution depth imaging via 1560 nm wavelength single-photon detection. Opt. Express, 21(7):8904–8915, Apr 2013.
- [99] P. F. McManamon. Lidar technologies and systems. (No Title), 2019.
- [100] R. M. Measures. Laser remote sensing: fundamentals and applications. J. Wiley, 1984.
- [101] N. Megiddo. On finding primal-and dual-optimal bases. ORSA Journal on Computing, 3(1):63–65, 1991.
- [102] N. Mellado, D. Aiger, and N. J. Mitra. Super 4pcs fast global pointcloud registration via smart indexing. Computer Graphics Forum, 33(5):205–215, 2014.
- [103] N. Mellado, G. Guennebaud, P. Barla, P. Reuter, and C. Schlick. Growing least squares for the analysis of manifolds in scale-space. In Computer Graphics Forum, volume 31, pages 1691–1701. Wiley Online Library, 2012.
- [104] M. I. Miller and L. Younes. Group actions, homeomorphisms, and matching: A general framework. International Journal of Computer Vision, 41:61–84, 2001.
- [105] D. Monnin, A. L. Schneider, F. Christnacher, and Y. Lutz. A 3d outdoor scene scanner based on a night-vision range-gated active imaging system. In Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT’06), volume 1, pages 938–945. IEEE, Jun 2006.
- [106] M. Nalamati, A. Kapoor, M. Saqib, N. Sharma, and M. Blumenstein. Drone detection in long-range surveillance videos. In 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), volume 1, pages 1–6. IEEE, Sep 2019.
- [107] B. K. Natarajan. Sparse approximate solutions to linear systems. SIAM Journal on Computing, 24(2):227–234, Jul 1995.
- [108] D. Needell and J. Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. Applied and Computational Harmonic Analysis, 26(3):301–321, 2009.
- [109] Q. M. Nguyen, H. H. Nguyen, Y. Zhou, and L. M. Nguyen. On unbalanced optimal transport: Gradient methods, sparsity and approximation error. arXiv preprint arXiv:2202.03618, 2022.
- [110] H. Nyquist. Certain topics in telegraph transmission theory. Transactions of the American Institute of Electrical Engineers, 47(2):617–644, 1928.

- [111] M. E. O'Brien and D. G. Fouche. Simulation of 3d laser radar systems. Lincoln Laboratory Journal, 15(1):37–60, Sep 2005.
- [112] G. Pallier and J.-F. Poisson. Beam shaping to scale up microprocessing. PhotonicsViews, 20(1):32–35, 2023.
- [113] A. M. Pawlikowska, A. Halimi, R. A. Lamb, and G. S. Buller. Single-photon three-dimensional imaging at up to 10 kilometers range. Opt. Express, 25(10):11919–11931, May 2017.
- [114] D. W. Peaceman and H. H. Rachford, Jr. The numerical solution of parabolic and elliptic differential equations. Journal of the Society for industrial and Applied Mathematics, 3(1):28–41, 1955.
- [115] W. B. Pennebaker and J. L. Mitchell. JPEG: Still image data compression standard. Springer Science & Business Media, 1992.
- [116] Å. Persson, U. Söderman, J. Töpel, and S. Ahlberg. Visualization and analysis of full-waveform airborne laser scanner data. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, 36(3/W19):103–108, 2005.
- [117] K. Pham, K. Le, N. Ho, T. Pham, and H. Bui. On unbalanced optimal transport: An analysis of sinkhorn algorithm. In International Conference on Machine Learning, pages 7673–7682. PMLR, 2020.
- [118] G. Puy, A. Boulch, and R. Marlet. Flot: Scene flow on point clouds guided by optimal transport. In European conference on computer vision, pages 527–544. Springer, 2020.
- [119] T. Rajeev. Scanning lidar in advanced driver assistance systems and beyond: Building a road map for next-generation lidar technology. IEEE Consumer Electronics Magazine, 5(3):48–54, July 2016.
- [120] J. Rapp. Probabilistic modeling for single-photon lidar. PhD thesis, 2020.
- [121] J. Rapp, Y. Ma, R. M. A. Dawson, and V. K. Goyal. High-flux single-photon lidar. Optica, 8(1):30–39, Jan 2021.
- [122] M. A. Richards, J. Scheer, W. A. Holm, and W. L. Melvin. Principles of modern radar. 2010.
- [123] W. H. Richardson. Bayesian-based iterative method of image restoration*. J. Opt. Soc. Am., 62(1):55–59, Jan 1972.
- [124] L. Risser, F.-X. Vialard, R. Wolz, M. Murgasova, D. D. Holm, and D. Rueckert. Simultaneous multi-scale registration using large deformation diffeomorphic metric mapping. IEEE transactions on medical imaging, 30(10):1746–1759, 2011.
- [125] N. Riviere, A. Amditis, A. Amiez, G. Athanasiou, J. Berggren, A. Boulch, N. Bozabaljian, D. Duarte, P.-E. Dupouy, P. Escalas, et al. 3d laser imaging techniques to improve usar operations for wide-area surveillance and monitoring of collapsed buildings. 2017.

- [126] N. Rivière, G. Anna, L. Hespel, B. Tanguy, M.-T. Velluet, and Y.-M. Frédéric. Modeling of an active burst illumination imaging system: comparison between experimental and modelled 3D scene. In G. W. Kamerman, O. Steinvall, K. L. Lewis, R. C. Hollins, T. J. Merlet, G. J. Bishop, and J. D. Gonglewski, editors, Electro-Optical Remote Sensing, Photonic Technologies, and Applications IV, volume 7835, page 783509. International Society for Optics and Photonics, SPIE, 2010.
- [127] N. Riviere, P.-É. Dupouy, A. Moussous, A. Schilling, and E. Viala. 3d lidar performance analysis for tactical-mapping applications. In 10th International Symposium on OPTRONICS IN DEFENCE AND SECURITY (OPTRO) 2022, 2022.
- [128] N. Riviere, P.-É. Dupouy, A. Moussous, A. Schilling, and E. Viala. Low-SWaP embedded 3D-LiDAR to detect non-cooperative targets. In M. A. Itzler, J. C. Bienfang, and K. A. McIntosh, editors, Advanced Photon Counting Techniques XVII, volume 12512, page 125120D. International Society for Optics and Photonics, SPIE, 2023.
- [129] B. Roman, A. Hansen, and B. Adcock. On asymptotic structure in compressed sensing, 2014.
- [130] J. Romberg. Imaging via compressive sampling. IEEE Signal Processing Magazine, 25(2):14–20, Mar 2008.
- [131] F. Rousset, N. Ducros, A. Farina, G. Valentini, C. d’Andrea, and F. Peyrin. Adaptive basis scan by wavelet prediction for single-pixel imaging. IEEE Transactions on Computational Imaging, 3(1):36–46, 2016.
- [132] R. Rubinstein, M. Zibulevsky, and M. Elad. Efficient implementation of the k-svd algorithm and the batch-omp method. Technical report, Technion - Computer Science Departement, 2008.
- [133] S. E. Schultz, N. P. Cervino, Z. D. Kurtz, and M. Z. Brown. Low-SWaP coincidence processing for Geiger-mode LIDAR video. In M. D. Turner, G. W. Kamerman, L. M. W. Thomas, and E. J. Spillar, editors, Laser Radar Technology and Applications XX; and Atmospheric Propagation XII, volume 9465, page 94650F. International Society for Optics and Photonics, SPIE, 2015.
- [134] B. E. Schutz, H. J. Zwally, C. A. Shuman, D. Hancock, and J. P. DiMarzio. Overview of the icesat mission. Geophysical Research Letters, 32(21), 2005.
- [135] C. E. Shannon. Communication in the presence of noise. Proc. Institute of Radio Engineers, 37(1):10–21, 1949.
- [136] Z. Shen, J. Feydy, P. Liu, A. H. Curiale, R. San Jose Estepar, R. San Jose Estepar, and M. Niethammer. Accurate point cloud registration with robust optimal transport. Advances in Neural Information Processing Systems, 34:5373–5389, 2021.
- [137] D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro. Photon-efficient computational 3-d and reflectivity imaging with single-photon detectors. IEEE Transactions on Computational Imaging, 1(2):112–125, Jun 2015.

- [138] D. Shin, F. Xu, D. Venkatraman, R. Lussana, F. Villa, F. Zappa, V. K. Goyal, F. N. C. Wong, and J. H. Shapiro. Photon-efficient imaging with a single-photon camera. Nature Communications, 7(1):12046, Sep 2016.
- [139] J. Son. UV detectors: status and prospects. In G. Léron del, S. Kawata, and Y.-H. Cho, editors, UV and Higher Energy Photonics: From Materials to Applications 2018, volume 10727, page 107270F. International Society for Optics and Photonics, SPIE, 2018.
- [140] D. Sprung, P. Grossmann, and E. Sucher. Investigation of seasonal and diurnal cycles on the height dependence of optical turbulence in the lower atmospheric boundary layer. In Laser Communication and Propagation through the Atmosphere and Oceans, volume 8517, pages 160–171. SPIE, 2012.
- [141] E. J. Stollnitz, T. D. DeRose, and D. H. Salesin. Wavelets for computer graphics: theory and applications. Morgan Kaufmann, 1996.
- [142] M.-J. Sun, M. Tong, M. Edgar, M. Padgett, and N. Radwell. A russian dolls ordering of the hadamard basis for compressive single-pixel imaging. Scientific Reports, 7(1):3464, Jun 2017.
- [143] M.-J. Sun and J.-M. Zhang. Single-pixel imaging and its application in three-dimensional reconstruction: A brief review. Sensors, 19(3):732, Feb 2019.
- [144] J. Tachella, Y. Altmann, N. Mellado, A. McCarthy, R. , G. S. Buller, J.-Y. Tourneret, and S. McLaughlin. Real-time 3d reconstruction from single-photon lidar data using plug-and-play point cloud denoisers. Nature Communications, 10(1):4989, Nov 2019.
- [145] J. Tachella, Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, S. McLaughlin, and J.-Y. Tourneret. Bayesian 3d reconstruction of complex scenes from single-photon lidar data. SIAM Journal on Imaging Sciences, 12(1):521–550, 2019.
- [146] A. M. Tillmann and M. E. Pfetsch. The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. IEEE Transactions on Information Theory, 60(2):1248–1259, 2013.
- [147] R. Tobin, A. Halimi, A. McCarthy, M. Laurenzis, F. Christnacher, and G. S. Buller. Three-dimensional single-photon imaging through obscurants. Opt. Express, 27(4):4590–4611, Feb 2019.
- [148] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. IEEE Transactions on Information Theory, 53(12):4655–4666, Dec 2007.
- [149] J. J. Van Zyl. The shuttle radar topography mission (srtm): a breakthrough in remote sensing of topography. Acta astronautica, 48(5-12):559–565, 2001.
- [150] N. Védrenne, M.-T. Velluet, V. Michau, G. Artaud, F. Lacoste, and E. Samain. Characterization of atmospheric turbulence for leo to ground laser beam propagation at low elevation angles. Proceedings of ICSOS, 2014.

- [151] W. B. Veldkamp and C. J. Kastner. Beam profile shaping for laser radars that use detector arrays. Applied Optics, 21(2):345–356, 1982.
- [152] S. V. Venkatakrisnan, C. A. Bouman, and B. Wohlberg. Plug-and-play priors for model based reconstruction. In 2013 IEEE global conference on signal and information processing, pages 945–948. IEEE, 2013.
- [153] E. Viala, P.-E. Dupouy, N. Riviere, and L. Risser. Increasing lateral resolution of 3d lidar gmapd using compressive sensing. In Applications of Lasers for Sensing and Free Space Communications, volume LsTu4C, page 2. Optica Publishing Group, 2022.
- [154] E. Viala, P.-E. Dupouy, N. Riviere, and L. Risser. Compressive sensing for 3d-lidar imaging: a pipeline to increase resolution of simulated single-photon camera. Journal of Quantitative Spectroscopy and Radiative Transfer, Special Issue(1):–, to be published 2023.
- [155] E. Viala, P.-E. Dupouy, N. Riviere, and L. Risser. Lateral-resolution enhancement of low-photon 3d-lidar by compressive sensing. In Advanced Photon Counting Techniques XVII, volume 12512, pages 62–67. SPIE, 2023.
- [156] F. Villa, R. Lussana, D. Bronzi, S. Tisa, A. Tosi, F. Zappa, A. Dalla Mora, D. Contini, D. Durini, S. Weyers, and W. Brockherde. Cmos imager with 1024 spads and tdc for single-photon timing and 3-d time-of-flight. IEEE Journal of Selected Topics in Quantum Electronics, 20(6):364–373, Nov 2014.
- [157] F. Villa, F. Severini, F. Madonini, and F. Zappa. Spads and sipms arrays for long-range high-speed light detection and ranging (lidar). Sensors, 21(11):3839, Jun 2021.
- [158] C. Villani et al. Optimal transport: old and new, volume 338. Springer, 2009.
- [159] J. G. Walker. Iterative correction for ‘pile-up’ in single-photon lifetime measurement. Optics Communications, 201(4):271–277, Jan 2002.
- [160] M. P. Wand and M. C. Jones. Kernel smoothing. CRC press, 1994.
- [161] J. Wen, Z. Zhou, J. Wang, X. Tang, and Q. Mo. A sharp condition for exact support recovery with orthogonal matching pursuit. IEEE Transactions on Signal Processing, 65(6):1370–1382, 2017.
- [162] N. Wiener, N. Wiener, C. Mathematician, N. Wiener, N. Wiener, and C. Mathématicien. Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications, volume 113. MIT press Cambridge, MA, 1949.
- [163] G. M. Williams and A. S. Huntington. Probabilistic analysis of linear mode vs. Geiger mode APD FPAs for advanced LADAR enabled interceptors. In R. T. Howard and R. D. Richards, editors, Spaceborne Sensors III, volume 6220, page 622008. International Society for Optics and Photonics, SPIE, 2006.

- [164] M. Yaghoobi, D. Wu, and M. E. Davies. Fast non-negative orthogonal matching pursuit. IEEE Signal Processing Letters, 22(9):1229–1233, 2015.
- [165] Y. X. Yang, X. X. Niu, and C. Q. Xu. Theory and applications of higher-dimensional hadamard matrices. (No Title), 2010.
- [166] Z. Zhang, X. Ma, and J. Zhong. Single-pixel imaging by means of fourier spectrum acquisition. Nature communications, 6(1):6225, 2015.
- [167] J. A. Zuclich, D. A. Gagliano, F. Cheney, B. E. Stuck, H. Zwick, P. R. Edsall, and D. J. Lund. Ocular effects of penetrating IR laser wavelengths. In S. L. Jacques, editor, Laser-Tissue Interaction VI, volume 2391, pages 112 – 125. International Society for Optics and Photonics, SPIE, 1995.