Compressive Sensing pour caméra LiDAR 3D en mouvement

CONSTANTIN DESOINDRE[1], ERWAN VIALA[1], PAUL-EDOUARD DUPOUY[1], LAURENT RISSER[2], NICOLAS RIVIERE[1]

[1] ONERA / DOTA, Université de Toulouse, F-31055, Toulouse, France
[2] Institut Mathématiques de Toulouse (UMR 5219), CNRS, Université de Toulouse, FR-31062, Toulouse, France

Résumé – L'utilisation de composants de modulation optique, comme le DMD (*Digital Micromirror Device*) permet d'améliorer la résolution latérale des acquisitions LiDAR 3D grâce à des algorithmes d'échantillonnage avancés. Parmi eux, le *Compressive Sensing* se distingue par sa capacité à reconstruire une scène avec moins de mesures que prédit par la limite de Nyquist-Shannon, en exploitant la compressibilité des scènes naturelles dans certaines bases. Cette approche est particulièrement adaptée aux contextes où l'acquisition des mesures est coûteuse en temps ou en ressource. Toutefois, lorsqu'une scène évolue au cours du temps, la prise en compte explicite de sa dynamique devient essentielle pour assurer la cohérence des reconstructions. Nous proposons ici d'intégrer le modèle LDDMM (*Large Deformation Diffeomorphic Metric Mapping*), afin de modéliser des transformations diffeomorphiques continues entre instants successifs. Ce cadre permet de régulariser temporellement les reconstructions en assurant la cohérence topologique des objets observés. La nouveauté de notre approche réside dans l'intégration explicite de ce modèle de dynamique dans un algorithme de *Compressive Sensing*, offrant ainsi une reconstruction spatio-temporelle plus fidèle des scènes dynamiques.

Abstract – The use of optical modulation components, such as the DMD can enhance the lateral resolution of 3D-LiDAR systems through advanced sampling algorithms. Among these, Compressive Sensing stands out for its ability to reconstruct a scene with fewer measurements than predicted by the Nyquist-Shannon limit by leveraging the compressibility of natural scenes in specific bases. This approach is particularly useful in contexts where data acquisition is time- or resource-intensive. However, when the observed scene evolves over time, explicitly modeling its temporal dynamics becomes essential to ensure consistent reconstruction. In this work, we propose to leverage the LDDMM framework to model continuous diffeomorphic transformations between successive time steps. This framework allows for temporally coherent reconstructions while preserving the topological structure of the observed objects. The novelty of our approach lies in the integration of this dynamic model into a Compressive Sensing scheme, enabling more accurate spatio-temporal reconstruction of dynamic scenes.

1 Introduction

Dans le contexte de la surveillance, le développement des capteurs de type GmAPD (*Geiger-mode Avalanche Photodiodes*) [1] a permis aux LiDAR 3D utilisant cette technologie d'atteindre de longues portées (> 10 km), grâce à leur capacité à détecter des photons uniques [2], [3]. Cependant, le nombre d'éléments photosensibles est limité, typiquement à 128×128 pixels² [4]. Pour pallier cette limitation, en résolution latérale, les dispositifs à micro-miroirs numériques (*Digital Micromirror Devices*, DMD) [5] permettent de moduler spatialement le signal lumineux en entrée de chaque pixel. En combinant cette modulation avec des acquisitions multiples, il devient possible d'augmenter la résolution latérale du système.

Comme les acquisitions peuvent être coûteuses en temps ou en ressources techniques, il est souvent souhaitable de réduire leur nombre. Les méthodes de *Compressive Sensing* (CS) [6], [7] exploitent la compressibilité des images dans certaines bases [8], [9] pour reconstruire, sous certaines conditions [10], un signal exact (sans bruit) ou approché (dont l'erreur est bornée par la norme du bruit) à partir d'un nombre réduit de mesures. Ce nombre est bien inférieur à celui requis pour un balayage classique, pixel par pixel.

Toujours dans le contexte de la surveillance, les objets de la scène peuvent être en mouvement relatif par rapport à la caméra. Une application directe des méthodes de CS devient insuffisante car le déplacement des objets entraîne des erreurs d'alignement entre les différentes modulations [11]. Il devient alors nécessaire d'estimer le flux optique [12], c'est-à-dire le déplacement des pixels de la scène au cours du temps.

Parmi les nombreuses méthodes d'estimation du flux optique, l'algorithme LDDMM (*Large Deformation Diffeomorphic Metric Mapping*) [13] se distingue par sa capacité à estimer des difféomorphismes, garantissant des transformations continues et inversibles. De plus, sa formulation variationnelle permet de contrôler le niveau de régularité spatiale et temporelle des déformations estimées. Elle ne nécessite pas de jeu de donnée d'entrainement et permet d'obtenir analytiquement des bornes d'erreurs, contrairement aux algorithmes d'apprentissage profond.

L'algorithme LDDMM a besoin de reconstructions initiales de la scène dans lesquelles les formes en mouvement apparaissent, même très bruitées, pour fonctionner. Il ne peut pas estimer le mouvement à partir de données modulées. La méthode présentée dans cet article adopte une approche multi-échelle de type *Coarse-to-Fine*, comparable à la méthode *CS multiscale video recovery* (CS-MUVI) [14]. Elle utilise LDDMM pour estimer les flux optiques entre deux scènes reconstruites par CS, à partir de motifs d'Hadamard

ordonnés [15], ce qui permet d'obtenir des estimations de plus en plus précises à chaque itération.

Ce cadre permet de gérer des scènes dynamiques malgré des reconstructions bruitées, tout en conservant une structure temporelle cohérente. Bien que préliminaire, cette approche vise à s'assurer que les flux optiques utilisés soient réguliers, continus, inversibles et d'inverses régulières tout en proposant des reconstructions intermédiaires dont la résolution augmente progressivement grâce à l'arrangement des motifs d'Hadamard.

Les résultats présentés dans cet article ont été simulés à partir de deux disques se déplaçant selon une translation linéaire horizontale à la matrice du capteur, dans des sens opposés, à une vitesse allant jusqu'à 1 pixel toutes les 16 images. En considérant, un système limité par une vitesse typique d'un DMD (10 kHz), le disque se déplacerait à une vitesse de 625 pixels par seconde. Soit 24 km.h^{-1} pour un disque à 10 m avec un champ de vue de 12,8 mrad et un capteur constitué de 128x128 pixels².

2 Compressive Sensing

2.1 Formulation matricielle du cas statique

Soit $x_0 \in \mathbb{R}^N$ notre scène observée. Soit $y \in \mathbb{R}^M$ le vecteur de nos M acquisitions. Chaque acquisition peut être écrite comme une combinaison linéaire des pixels de la scène :

$$\forall j \in [\![1, M]\!],\ y_j = \sum_{i=1}^{N} a_{ij}\, x_{0_i} = \langle a_j, x \rangle$$

où a_{ij} vaut 1 si le i-ème pixel du DMD renvoie la lumière vers le capteur, lors de la j-ième acquisition, et 0 sinon. En regroupant ces coefficients dans une matrice $A \in \{0,1\}^{M \times N}$, où chaque ligne correspond aux coefficients d'un filtre d'acquisition, le système s'écrit :

$$y = Ax_0 \quad (1)$$

L'objectif du CS est de reconstruire la scène x_0 avec un nombre réduit d'acquisitions, c'est-à-dire $M < N$. Dans ce cas, le système est sous-déterminé et admet une infinité de solutions.

Notons que cette formulation n'est plus valable dans le cas dynamique, puisque notre scène n'est plus constante selon le temps. Il est alors nécessaire de reformuler le problème selon $x(t)$, ce qui sera fait dans la section 4.2. Le cas statique correspondra alors au cas particulier : $\forall t \in [0, T], x(t) = x_0$.

2.2 Parcimonie

Les scènes naturelles présentent généralement des structures (surfaces, lignes, etc.) qui les rendent compressibles dans certaines bases, telles que les bases d'Ondelettes [9] ou de Fourier [8]. Cela signifie que leur représentation dans ces bases est parcimonieuse.

On définit la quasi-norme ℓ_0 pour mesurer la parcimonie d'un signal :

$$\forall s \in \mathbb{R}^N, \|s\|_0 = \#\{i \mid s_i \neq 0\}$$

Ainsi, pour un signal $x_0 \in \mathbb{R}^N$ et une base de compression $\Psi \in \mathbb{R}^{N \times N}$, on s'attend à ce que $\|\Psi x_0\|_0$ soit minimal. En reprenant le fait que $y = Ax_0$, on peut réécrire notre problème sous la forme d'un problème d'optimisation sous contrainte :

$$\hat{x} = \underset{x_0 : \|Ax_0 - y\|_2^2 \leq \varepsilon}{\operatorname{argmin}} \|\Psi x_0\|_0$$

avec $\varepsilon > 0$ le seuil de tolérance (dépendant du bruit).

La quasi-norme ℓ_0 n'étant pas convexe, le problème est NP-difficile [16]. Ce qui le rend impraticable pour des résolutions efficaces.

Pour contourner cette difficulté, deux grandes classes d'approches sont utilisées :

- Les algorithmes gloutons (ex. : *Matching Pursuit* [17], *Orthogonal Matching Pursuit* [18]) qui approchent la solution de manière itérative.
- La relaxation convexe par la norme ℓ_1 [19], consiste à résoudre le problème suivant :

$$\hat{x} = \underset{x : \|Ax_0 - y\|_2^2 \leq \varepsilon}{\operatorname{argmin}} \|\Psi x_0\|_1$$

Où $\|\Psi x_0\|_1 = \sum_{i=1}^{N} |(\Psi x_0)_i|$

Il a été démontré que si la matrice $\Theta = A\Psi$ respecte la *Restricted Isometry Property* (RIP), alors les différents algorithmes produisent une reconstruction stable, robuste et unique [10] et que les solutions du problème relaxé coïncident avec celles du problème initial [19].

Des travaux ont été menés pour lier la reconstruction par CS avec une reconstruction LiDAR 3D [11].

3 Estimation de mouvement par LDDMM

Soient $I_0,\ I_1 : \Omega \subset \mathbb{R}^3 \to \mathbb{R}$ une scène dynamique prise à deux temps différents. Ici, on pourrait typiquement prendre $I_0 = x(t = t_0)$ et $I_1 = x(t = t_1)$. Il existe de nombreuses méthodes [20] permettant de calculer les déplacements survenus entre ces deux acquisitions.

Parmi elles, l'algorithme Large *Deformation Diffeomorphic Metric Mapping* (LDDMM) [13] vise à calculer une transformation difféomorphique $\varphi : \Omega \to \Omega$ telle que son application réciproque à I_0 reconstitue I_1 :

$$I_0 \circ \varphi^{-1} = I_1$$

Afin de garantir que φ soit difféomorphique et qu'elle préserve la topologie des objets observées, nous introduisons une famille de transformations $\phi_t : \Omega \to \Omega$ pour tout $t \in [0,1]$ satisfaisant :

$$\phi_0 = Id,\ \phi_1 = \varphi$$

Cela permet de suivre un chemin continu de déformations entre l'identité et la transformation finale, garantissant l'inversibilité et la régularité de ϕ_t à chaque instant.

On associe à cette famille un champ de vitesses $v_t : \Omega \to \mathbb{R}^3, t \in [0,1]$ défini par l'équation différentielle ordinaire (EDO) suivante :

$$\dot{\phi}_t = v_t(\phi_t)$$

où $\dot{\phi}_t$ désigne la dérivée temporelle de ϕ_t.

Définissons alors deux énergies à minimiser. La première est une énergie mesurant à quel point la transformation $\phi_1 = \varphi$ permet bien de passer de I_0 à I_1 :

$$E_1(\phi_1) = \left\| I_1 - I_0 \circ \phi_1^{-1} \right\|_{L^2}^2$$

Plus cette énergie sera petite, plus φ^{-1} transformera bien I_0 en I_1. La deuxième énergie mesure la régularité spatiale de v_t, au sens de la norme L_2, afin de contrôler la déformation :

$$E_2(v_t) = \int_0^1 \|Lv_t\|_{L^2}^2 dt$$

où L est un opérateur différentiel. Typiquement $L = -\alpha\nabla^2 + \gamma Id$ avec $\alpha, \gamma > 0$.

En introduisant l'hyperparamètre σ afin de pondérer l'influence des deux énergies, nous pouvons alors écrire le problème de minimisation de l'énergie totale, sous la contrainte de l'EDO $\dot{\phi}_t = v_t(\phi_t)$:

$$\hat{v} = \underset{v:\dot{\phi}_t = v_t(\phi_t)}{\operatorname{argmin}} \left(E_1(\phi_1) + \frac{1}{\sigma^2} E_2(v_t) \right)$$

La dérivée de cette énergie totale selon v_t est calculable analytiquement [13], ce qui permet d'utiliser les algorithmes de descente de gradients classiques pour résoudre le problème.

4 Méthode *Coarse-to-Fine*

La méthode *Coarse-to-Fine* développée dans cet article consiste à estimer itérativement des scènes de plus en plus résolues ainsi qu'un flux optique de plus en plus précis en utilisant des motifs d'Hadamard ordonnés.

4.1 Motifs d'Hadamard Ordonnés

On choisit les matrices d'Hadamard construites par récurrence en utilisant le produit de Kronecker noté $\otimes$:

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$
$$\forall n > 1 : H_{2^{n+1}} = H_{2^n} \otimes H_2$$

Les lignes de cette matrice sont ensuite ordonnées de sorte à ce que pour tout $i \in [\![0, N]\!]$ H_i soit imbriqué dans H_{i+1} (l'ordre *Russian Doll* décrit par Sun et al. [15] et présenté dans la figure 1), il a été montré que les 2^i premiers motifs de notre matrice d'Hadamard de taille 2^n avec $i \leq n$ permettent une reconstruction de $2^i \times 2^i$ blocs de pixels² ayant des valeurs qui tendent à être celles de la moyenne des pixels de la scène réelle contenus dans le bloc.

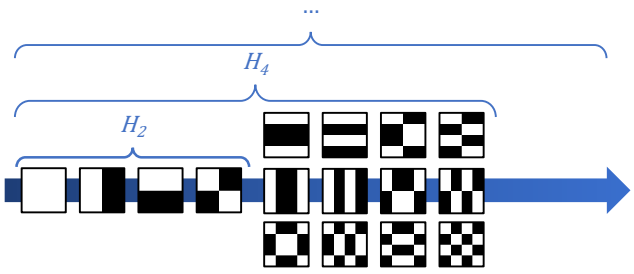

Figure 1 : Motifs d'Hadamard et leurs classements selon l'ordre des *Russian Doll*

4.2 Ecriture du problème matriciel

L'objectif est de réécrire l'équation (1) en prenant en compte la dépendance temporelle de la scène x. En écrivant $\forall i \in [\![1, N]\!], x_i = x(t = i), y_i = y(t = i)$ et en supposant que chaque mouvement $D_i \in \mathbb{R}^{N \times N}$ soit connu tel que : $\forall i \in [\![1, N]\!], x_i = D_i x_0$, on peut réécrire l'équation sous la forme :

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} < A_0, x_0 > \\ < A_1, x_1 > \\ \vdots \\ < A_n, x_n > \end{bmatrix} = \begin{bmatrix} < A_0, D_0 x_0 > \\ < A_1, D_1 x_0 > \\ \vdots \\ < A_n, D_n x_0 > \end{bmatrix} =$$
$$\begin{bmatrix} < D_0^T A_0, x_0 > \\ < D_1^T A_1, x_1 > \\ \vdots \\ < D_n^T A_n, x_n > \end{bmatrix} = \begin{bmatrix} D_0^T A_0 \\ D_1^T A_1 \\ \vdots \\ D_n^T A_n \end{bmatrix} x_0 = \breve{A} x_0 \quad (2)$$

où $\forall i \in [\![1, N]\!], A_i$ est la i-ème ligne de A. Comme x_0 est toujours une scène naturelle, et sous certaines conditions présentées par [21], on peut appliquer les algorithmes de CS.

4.3 Algorithme *Coarse-to-Fine*

L'algorithme suit les étapes suivantes :

(Initialisation) On initialise n par n_0 et on choisit un maillage temporel $(t_i)_{i \in [\![0,T]\!]}$. La matrice A est construire de sorte que, pour tout $j \in [\![1, T]\!]$, les motifs correspondant aux temps t_j à $t_j + 2^{n_{max}}$ soient les $2^{n_{max}}$ premiers motifs d'Hadamard dans l'ordre des *Russian Doll*. On initialise le mouvement par $D_i = Id$ (absence de mouvement).

(Compressive Sensing) On calcule $\breve{A}$ à partir de (2) pour estimer la scène à chaque pas de temps t_i, *via* un algorithme de ***Compressive Sensing***, ici FISTA [22].

(Estimation du mouvement) On utilise l'algorithme **LDDMM** appliqué entre chaque pas de temps t_j pour mettre à jour les mouvements D_i.

(Itération) En incrémentant n, et en répétant les étapes 2 et 3, on affine progressivement l'estimation de la scène et du mouvement.

(Estimation Finale) Une fois n_{max} atteint, on peut utiliser l'ensemble de la matrice A et des mouvements D_i pour construire notre dernière reconstruction par ***CS***. Une étape de post-traitement est appliquée pour régulariser par variation totale [23]. L'algorithme a été testé sur un disque subissant une translation linéaire et diagonale par rapport au capteur, comme illustré sur la Figure 2.

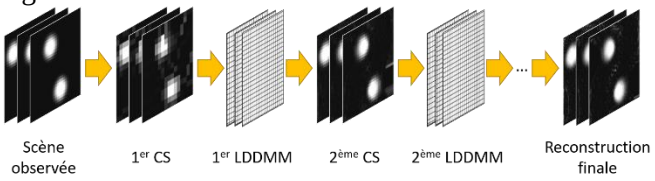

Figure 2 : Fonctionnement de l'algorithme

5 Résultats et discussions

Notre algorithme a donné des résultats significativement meilleurs qu'une application brute d'un algorithme de *CS* sur les données. La Figure 3 montre que notre algorithme a su aligner les différentes acquisitions sur la scène $x(t = 0)$. Ce n'est pas le cas pour une application direct du CS, puisque la reconstruction obtenue et une combinaison de l'ensemble des $x(t), t \in [0, T]$, ce qui empêche la reconnaissance du disque.

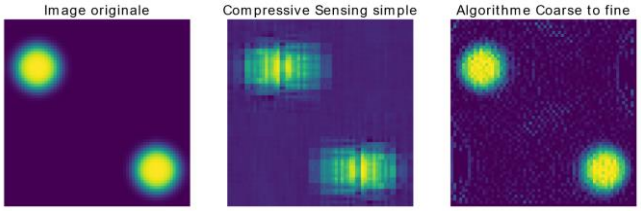

Figure 3 : De gauche à droite : image originale x_0, estimation renvoyée par application directe de FISTA, et estimation par notre algorithme *Coarse-to-Fine*

On peut alors suivre la variation de métriques (MSE, SSIM [24]) en fonction du taux d'échantillonnage (figure 4).

$$MSE(\hat{x}, x) = \frac{\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}{n}$$

$$SSIM(\hat{x}, x) = \frac{(2\mu_x\mu_{\hat{x}} + 6.5025)(2\sigma_{x\hat{x}} + 58.5225)}{(\mu_x^2 + \mu_{\hat{x}}^2 + 6.5025)(\sigma_x^2 + \sigma_{\hat{x}}^2 + 58.5225)}$$

Avec $\mu_x, \mu_{\hat{x}}$ les moyennes des pixels des images x et $\hat{x}$; $\sigma_x, \sigma_{\hat{x}}$ leurs variances et $\sigma_{x\hat{x}}$ leur covariance.

La MSE de l'estimation par la méthode *Coarse-to-Fine* diminue en fonction du taux d'échantillonnage, ce qui est un bon indicateur dans le cadre du CS. Parallèlement, le SSIM augmente et tend vers 1, indiquant une reconstruction de plus en plus fidèle à la scène réelle.

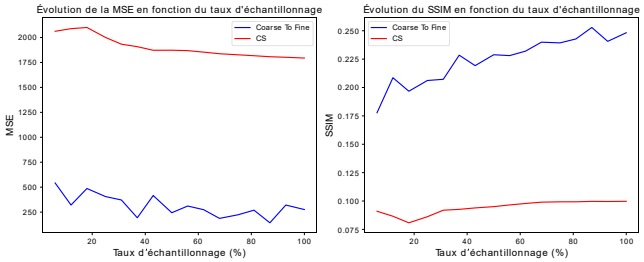

Figure 4 : À gauche évolution de la MSE comparant x_0 et $\widehat{x_0}$ en fonction du taux d'échantillonnage (M/N) ; à droite évolution du SSIM

L'application d'un algorithme de CS sans estimation du mouvement a donné des résultats nettement moins satisfaisants dans la majorité des cas. En effet, dans ce cas, la MSE obtenu est constamment supérieur à celui obtenu par notre méthode. Le SSIM obtenu par l'application directe du CS est lui inférieur (autrement dis l'image moins similaire à celle attendue).

6 Conclusion

Cet article a présenté une approche permettant d'améliorer la résolution latérale des imageurs LiDAR 3D basés sur des capteurs de type single pixel camera (ex : GmAPD), tout en tenant compte du mouvement des objets observés. En combinant les principes du *Compressive Sensing* avec une estimation robuste du flux optique *via* l'algorithme LDDMM, nous avons démontré qu'il est possible d'atténuer les erreurs d'alignement induites par le déplacement des cibles dans la scène.

Les simulations effectuées sur deux disques se déplaçant selon une translation linéaire horizontale à la matrice du capteur, dans des sens opposés ont montré que notre méthode permet une reconstruction précise de la scène, même à des vitesses élevées, grâce à une approche multi-échelle et à l'utilisation de motifs d'Hadamard ordonnés. Ces résultats ouvrent la voie à des applications concrètes en surveillance longue portée.

Cette méthode repose sur une estimation du flux optique à l'aide d'un algorithme permettant un contrôle explicite de la régularité temporelle, à la différence des approches classiques telles que celle de Horn et Schunck. Dans la suite de ce travail, nous envisageons d'exploiter la structure variationnelle de notre formulation pour développer une modélisation analytique des incertitudes, difficilement accessible dans le cadre des méthodes d'apprentissage profond. Ces premiers résultats constituent ainsi une première étape dans l'étude de l'apport des déformations difféomorphiques dans le cadre du compressive sensing dynamique, notamment pour l'analyse de scènes dynamiques à partir de données parcimonieuses.

Références

[1] B. Aull, *Geiger-Mode Avalanche Photodiode Arrays Integrated to All-Digital CMOS Circuits*, 2016

[2] P. A. Hiskett et al., *Long range 3D imaging with a 32×32 Geiger mode InGaAs/InP camera*, 2014

[3] M. Entwistle et al., *Geiger-mode APD camera system for single-photon 3D LADAR imaging*, 2012

[4] C. Bradley et al., *3D imaging with 128x128 eye safe InGaAs p-i-n lidar camera*, 2019

[5] B. Lee, *Introduction to ±12 degree orthogonal digital micromirror devices (dmds)*, 2008

[6] D. L. Donoho, *Compressed sensing*, 2006

[7] E. J. Candès, J. K. Romberg, T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, 2006

[8] R. N. Bracewell, *The Fourier transform*, 1989

[9] Y. Meyer, D. H. Salinger, *Wavelets and operators*, 1992

[10] E. J. Candes, T. Tao, *Decoding by Linear Programming*, 2005

[11] E. Viala, P.-E. Dupouy, N. Riviere, L. Risser, *Compressive sensing for 3D-LiDAR imaging: A pipeline to increase resolution of simulated single-photon camera*, 2024

[12] B. K. Horn, B. G. Schunck, *Determining optical flow*, 1981

[13] M. F. Beg et al. *Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms*, 2005

[14] A. C. Sankaranarayanan et al., *CS MUVI: Video compressive sensing for spatial-multiplexing cameras*, 2012

[15] M.-J. Sun et al. « A Russian Dolls ordering of the Hadamard basis for compressive single-pixel imaging », 2017

[16] B. K. Natarajan, *Sparse Approximate Solutions to Linear Systems*, 1995

[17] S. G. Mallat, Zhifeng Zhang, *Matching pursuits with time-frequency dictionaries*, 1993

[18] J. A. Trop, A. C. Gilbert, *Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit*, 2007

[19] D. L. Donoho, *For most large underdetermined systems of linear equations the minimal ℓ1-norm solution is also the sparsest solution*, 2006

[20] M. Zhai, et al. *Optical flow and scene flow estimation: A survey*, 2021

[21] S. Jalali, X. Yuan, *Snapshot Compressed Sensing: Performance Bounds and Algorithms*, 2019

[22] A. Beck, M. Teboulle, *A Fast-Iterative Shrinkage Thresholding Algorithm for Linear Inverse Problems*, 2009

[23] L. I. Rudin, S. Osher, E. Fatemi, *Nonlinear total variation based noise removal algorithms*, 1992

[24] Zhou Wanget al., *Image quality assessment: from error visibility to structural similarity*, 2004