

Projet de fin d'études
Compensation du mouvement pour la
reconstruction d'images avec une caméra
mono-pixel

Encadrants laboratoire :

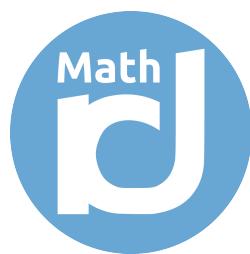
Auteur :
Thomas MAITRE

Elie BRETIN ¹
Nicolas DUCROS ²
Michaël SDIKA ²

Encadrant école :
Annabelle COLLIN

¹ Institut Camille Jordan (ICJ)
² Centre de Recherche en Acquisition et Traitement de l'Image pour la Santé (CREATIS)
21 Av. Jean Capelle, 69100 Villeurbanne

CREATIS



Mots-clés : Problème inverse, compensation de mouvement, imagerie mono-pixel, imagerie hyperspectrale, apprentissage profond

Table des matières

1	Introduction	3
2	Caméra mono-pixel	4
2.1	Un problème inverse	4
2.2	Choix des motifs projetés	5
2.3	Avantages de la caméra mono-pixel	6
3	Cas statique : reconstruction d'un objet immobile	7
3.1	Reconstruction sans bruit	7
3.2	Présence d'un bruit sur les mesures	7
3.3	Régularisation TV	8
4	Cas dynamique : reconstruction d'un objet en mouvement	10
4.1	Méthode de compensation du mouvement	11
4.1.1	Approche Lagrangienne	11
4.1.2	Approche Eulérienne	12
4.2	Reconstruction dynamique sur les données simulées	12
4.3	Erreur sur le modèle de mouvement	15
4.4	Sous-échantillonage	17
5	Estimation du mouvement	18
5.1	Champ constant	18
5.2	Algorithme de flot optique	20
5.3	Une méthode d'apprentissage plus rapide	21
5.4	Résultats	21
6	Reconstruction expérimentale	22
6.1	Présentation des données	23
6.2	Reconstruction statique	24
6.3	Estimation du mouvement	25
6.3.1	Calibration de la caméra	25
6.3.2	Estimation du mouvement en haute résolution	26
6.4	Reconstruction dynamique	27
7	Conclusion	30
8	Références	31
9	Annexes	34
9.1	Algorithme de régularisation TV en Python	34
9.2	Effet du sous-échantillonage sur le temps de calcul	34
9.3	Erreurs sur l'estimation du mouvement	34
9.4	Interpolation bilinéaire sur la tâche blanche	35
10	Résumé en anglais	36

Nomenclature

Tout au long de ce rapport, les variables en gras sont utilisées pour désigner la version discrétisée des variables considérées. De plus, des symboles minuscules et majuscules sont utilisés pour représenter respectivement des vecteurs et des matrices. Ainsi, \boldsymbol{x} représente un vecteur et \boldsymbol{X} représente une matrice.

Abréviations

ACP	Analyse en composantes principales
DLT	« Direct linear transform »
DMD	Matrice de micro-miroirs
FGS	Chirurgie guidée par fluorescence
IRM	Imagerie par résonance magnétique
PpIX	Protoporphyrin IX
PSNR	« Peak signal-to-noise ratio »
RGB	Rouge-vert-bleu : désigne les trois canaux d'une caméra
SP	« Single-pixel »
TV	Variation totale

Symboles

f, f_{ref}	Image et image de référence
$X = [0, 1]^2$	Domaine spatial de l'image
$N = n \times n$	Nombre de pixels de l'image f
K	Nombre de mesures
α	Coefficient représentant un nombre de photons qui sont effectivement captés par le détecteur
η	Coefficient de régularisation
τ	Pas de descente
λ	Longueur d'onde
T, T_{acq}	Période d'une pulsation, durée d'acquisition
u_t, v_t	Déformations directe et inverse à l'instant t
e^f, \bar{e}	Champ d'erreur à la frame f et champ d'erreur moyen sur l'ensemble des frames
$\mathcal{A}_{\text{stat}}, \mathcal{A}_{\text{dyn}}$	Opérateurs de reconstruction statique et dynamique
$\boldsymbol{H}, \boldsymbol{H}_d$	Matrice de Hadamard, matrice de Hadamard dynamique
\boldsymbol{I}_N	Matrice identité de taille $N \times N$
\boldsymbol{S}	Matrice de sous-échantillonage
$\boldsymbol{m} = (m_0, \dots m_{K-1})^T$	Vecteur des mesures

1 Introduction

S'il est conseillé de bouger le moins possible au cours d'une IRM, c'est parce que le mouvement d'un patient lors de cette opération crée des artefacts sur les images obtenues. Ce problème n'est pas spécifique à l'IRM mais commun à toutes les méthodes d'imagerie, en particulier pour celles ayant de longs temps d'acquisition. Bien souvent, ces artefacts de mouvement sont gênants et empêchent le clinicien d'évaluer correctement la situation du patient, ce qui peut engendrer de mauvaises prises de décisions dont les conséquences en chirurgie peuvent être particulièrement graves. L'objectif de ce stage est d'explorer des méthodes permettant de compenser le mouvement pendant la reconstruction de l'image avec un dispositif d'acquisition particulier : la caméra mono-pixel.

L'imagerie mono-pixel est un domaine qui a gagné en popularité grâce à l'article de Duarte et al. [1] qui en présente les avantages apportés par la théorie de l'acquisition comprimée. Le principe général de cette technique d'imagerie est d'utiliser une caméra mono-pixel pour acquérir les produits scalaires de l'image de la scène avec une série de fonctions de base définies au préalable [1] [2]. La connaissance de ces mesures permet ensuite de formuler un problème inverse qu'il est possible de résoudre grâce à des algorithmes de reconstruction [3].

L'imagerie mono-pixel trouve des applications dans de nombreux domaines dont une liste non exhaustive contient l'imagerie hyperspectrale [4], la microscopie par fluorescence [5] ou encore la chirurgie assistée par imagerie [6]. En particulier, elle peut-être utilisée pour la chirurgie guidée par fluorescence (FGS) [7], une technique qui facilite la chirurgie des tumeurs cérébrales en indiquant la position des gliomes en utilisant, par exemple, la protoporphyrin IX (PpIX) comme agent de contraste [8]. Le travail s'inscrit dans la continuation des travaux déjà effectués dans le but d'améliorer cette technique chirurgicale.

Au cours des années, l'IRM a bénéficié de plusieurs travaux proposant des stratégies de compensation de mouvement. Tandis que les premières méthodes proposées nécessitaient des hypothèses fortes sur le modèle de mouvement [9], des travaux proposant un cadre général ont depuis été réalisés [10] [11]. En revanche, ce n'est que beaucoup plus récemment que des articles d'imagerie mono-pixel proposent des méthodes de compensation de mouvement [12] [13].

Les travaux qui se veulent généraux (sans hypothèses préalables sur le modèle du mouvement) couplent leurs approches à des méthodes d'estimation de mouvement. Les méthodes traditionnelles pour estimer le mouvement à partir d'un flux vidéo sont basées sur la méthode du flot optique [14]. Afin d'estimer le mouvement en temps réel, une méthode plus rapide basée sur de l'apprentissage sera éventuellement utilisée [15]. Des méthodes d'apprentissage profond peuvent aussi être utilisées lors de la reconstruction pour réduire les temps de calcul [16] [17].

Dans le but d'améliorer la qualité des reconstructions hyperspectrales obtenues, plusieurs techniques ont été proposées dans la littérature [18]. Tout l'enjeu consiste à augmenter la résolution spatiale des images sans trop perdre de l'importante résolution spectrale des images hyperspectrales. On donne l'exemple des méthodes classiques de fusion par substitution de composant [19] [20] qui semblent faciles à implémenter (*e.g.* avec l'ACP) mais qui subissent une importante distorsion spectrale. Des méthodes prometteuses du domaine sont issues d'un cadre bayésien qui permettent d'avoir de bons résultats spectraux et spatiaux avec un faible coût computationnel [18].

Actuellement, l'acquisition d'un hypercube spectral de résolution $64 \times 64 \times 2048$ prend une dizaine de secondes [21]. Cela pose problème si l'on veut reconstruire une scène animée d'un mouvement physique plus rapide, *e.g.* la respiration ou le battement cardiaque du patient (environ un cycle par seconde). Il s'agit donc de développer une méthode permettant de compenser le mouvement durant la période d'acquisition.

Ce mémoire a été réalisé dans le cadre d'un projet de fin d'études de 6 mois au sein de l'équipe TOMORADIO du laboratoire CREATIS, affilié à l'INSA Lyon et de l'université Claude Bernard. Cette équipe travaille au quotidien dans le but pour contribuer aux progrès de différents domaines de la tomographie, notamment en étudiant les problèmes inverses en imagerie mais aussi en travaillant sur la simulation et la reconstruction en radiothérapie.

Ce travail vise à contribuer de plusieurs manières au domaine de l'imagerie mono-pixel. Tout d'abord, la section 2 présente la caméra mono-pixel. Ensuite, les éléments théoriques nécessaires à la reconstruction d'un objet immobile sont rappelés en section 3, accompagnés d'une présentation de la technique de régularisation retenue lors de ce projet. Le cas d'un objet animé d'un mouvement est ensuite étudié en section 4, où une stratégie de compensation de mouvement est proposée et validée à l'aide de simulations numériques. Des travaux portant sur l'estimation du mouvement sont ensuite menés en section 5. Enfin, on propose de confronter la méthode sur des données expérimentales en section 6.

2 Caméra mono-pixel

2.1 Un problème inverse

La caméra mono-pixel est un dispositif expérimental permettant l'acquisition d'une image f de taille $N = n \times n$ grâce à un seul détecteur là où les caméras classiques CCD et CMOS [22] en utilisent un grand nombre. Pour cela, on éclaire la scène à travers différents motifs lumineux à l'aide d'une matrice de micro-miroirs (DMD) pour sélectionner certains rayons de lumière, puis on utilise une lentille convergente afin de sommer ces rayons sur un unique pixel comme en Figure 1. En intégrant le signal obtenu pendant une période d'acquisition T_{acq} , on obtient un jeu de mesures $\{m_0, \dots, m_{K-1}\} \subset \mathbb{R}^K$. On considère qu'une image est une fonction bidimensionnelle dans $L^2(X)$ où $X = [0, 1]^2$ est l'espace des coordonnées de l'image. Cela permet de définir l'opérateur de reconstruction $\mathcal{A} : L^2(X) \rightarrow \mathbb{R}^K$ qui établit la relation entre les mesures m_k et l'image f . Pour $k \in [0, K - 1]$, on établit alors la relation :

$$\mathcal{A}[f]_k = m_k \quad (1)$$

où la notation $\mathcal{A}[f]_k$ est utilisée pour désigner la k -ème composante de l'opérateur de reconstruction \mathcal{A} appliquée à f .

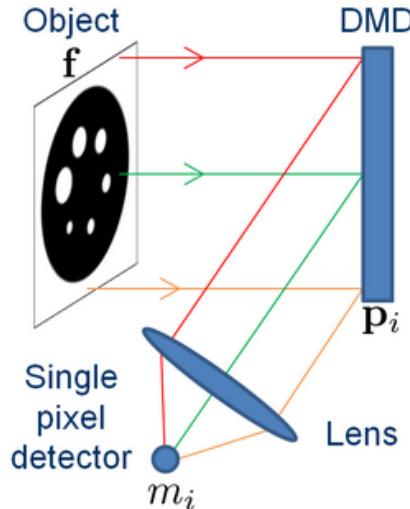


FIGURE 1 – Schéma d'une caméra mono-pixel

D'après le processus d'acquisition, chaque mesure m_k est obtenue par éclairage de la scène au travers d'un motif $a_k : X \rightarrow Z$ où $Z \subset \mathbb{R}$ est un espace discret qui dépend du système de représentation des motifs. En pratique, le DMD permet de rediriger une partie du faisceau lumineux vers le détecteur en inclinant une partie des petits miroirs qui le composent, tandis que l'autre partie est envoyée dans la direction opposée.

Par construction, l'opérateur \mathcal{A} de l'image f est supposé suivre une relation linéaire telle que, pour $k \in \llbracket 0, K - 1 \rrbracket$:

$$\mathcal{A}_{stat}[f]_k = \int_X a_k(x) f(x) dx = \langle a_k, f \rangle_{L^2(X)} \quad (2)$$

Cette relation est celle qui motive le choix de l'espace L^2 pour les images. En effet en faisant l'hypothèse raisonnable que le motif a_k est une fonction bornée, on voit tout de suite que (2) est bien définie. En considérant les équations (1) et (2), on peut écrire le processus d'acquisition comme un produit scalaire entre la scène et une série de motifs lumineux :

$$m_k = \langle a_k, f \rangle_{L^2(X)} \quad (3)$$

En discréétisant l'équation (3), on obtient :

$$\mathbf{m} = \mathbf{Af} \quad (4)$$

où $\mathbf{m} = (m_0, \dots, m_{K-1})^T$ est le vecteur contenant les mesures de la caméra mono-pixel, $\mathbf{f} \in \mathbb{R}^N$ est la version discrète de l'image mise sous forme d'un vecteur en rangeant ses éléments ligne par ligne et $\mathbf{A} \in \mathbb{R}^{K \times N}$ est la matrice déterminée par le choix des motifs a_k . On met ainsi en évidence un problème inverse, où il s'agit de déterminer \mathbf{f} étant données un jeu de mesures \mathbf{m} qui suivent la relation (3).

2.2 Choix des motifs projetés

Il y a plusieurs choix de motifs possibles. L'approche naturelle consisterait à acquérir chaque pixel indépendamment des autres, ce qui reviendrait à choisir $\mathbf{A} = \mathbf{I}_N$. En supposant que l'erreur des mesures est de variance σ^2 , on a alors que $Cov(f) = \sigma^2 \mathbf{I}_N$ [2]. En revanche, choisir la base discrète de Hadamard permettrait d'obtenir $Cov(f) = \frac{\sigma^2}{N} \mathbf{I}_N$, réduisant la variance du signal d'un facteur \sqrt{N} [2]. Dans ce rapport, c'est donc cette base qui est considérée bien qu'il est important de garder à l'esprit que d'autres choix sont possibles (e.g. base de Fourier [23]). La suite de cette partie présente avec plus de détails la transformée de Hadamard.

Une matrice de Hadamard est une matrice $\mathbf{H} \in \mathcal{M}(\{-1, 1\})$ dont les lignes sont orthogonales. Puisque la transformée est discrète, cela permet de facilement programmer le DMD sans les erreurs de discréétisations qu'il pourrait il y avoir en utilisant la transformée de Fourier. On considère donc $Z = \{-1, 1\}$, ce qui nécessite deux acquisitions pour le même motif : une première pour sa partie positive et une seconde pour sa partie négative. En suivant la démarche de [16], on considère ensuite que les mesures s'expriment comme la différence de leurs parties positives avec leurs parties négatives, *i.e.* $m_k = m_k^+ - m_k^-$.

Il est possible de construire des matrices de Hadamard de taille $2^p \times 2^p$ de manière itérative à l'aide de la méthode de Sylvester :

$$H_1 = (1) \text{ et } H_{2^p} = \begin{pmatrix} H_{2^{p-1}} & H_{2^{p-1}} \\ H_{2^{p-1}} & -H_{2^{p-1}} \end{pmatrix} \quad (5)$$

Pour la représentation d'un signal 2D à N pixels, on souhaitera construire la matrice de Hadamard de taille $N \times N$, qui vérifiera la relation $\mathbf{HH}^T = N\mathbf{I}_N$. On appliquera une permutation

sur les lignes de cette matrice selon l'ordre de Walsh afin d'avoir les motifs ordonnés selon l'ordre croissant des fréquences. Pour illustrer ce qui a été dit, on construit une matrice de Hadamard 2D pour une image 4×4 et on affiche les motifs associés en Figure 2.

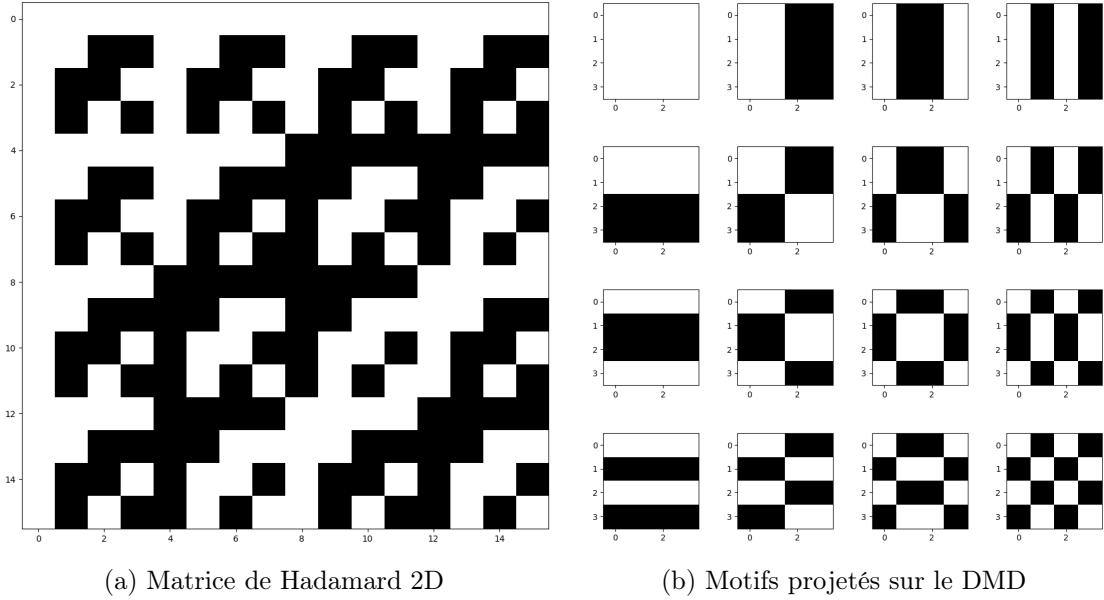


FIGURE 2 – Construction des motifs de Hadamard pour une image 4×4

Chaque motif correspond à une ligne de la matrice de Hadamard agencée en une matrice 4×4 en plaçant ses éléments ligne par ligne. Les pixels blancs sont égaux à 1, les pixels noirs à -1.

2.3 Avantages de la caméra mono-pixel

Alors que l'on parvient aujourd'hui à développer des caméras de plus en plus performantes, la question de l'utilisation de la caméra mono-pixel se pose. Son utilisation présente trois principaux avantages qui sont présentés dans cette partie.

Premièrement, la caméra mono-pixel permet de facilement ajouter une dimension spectrale en plus des dimensions spatiales lors de la reconstruction en branchant le récepteur à un spectromètre. Contrairement aux imageurs par snapshots qui doivent trouver un équilibre entre résolution spatiale et spectrale, l'imagerie mono-pixel permet d'avoir une très haute résolution spectrale [21] mais nécessite plus d'échantillons de mesures et donc prend naturellement plus de temps. Cependant, il est possible de fortement réduire ce nombre grâce la théorie de l'acquisition comprimée. C'est pour cette raison qu'il est intéressant d'utiliser le dispositif pour la chirurgie guidée par fluorescence.

De plus, l'imagerie mono-détecteur est facilement adaptable à d'autres domaines électromagnétiques que le visible en changeant simplement la source d'illumination utilisée. Typiquement il est possible de faire de l'imagerie infrarouge, à rayons X ou encore térahertz [24].

Enfin pour citer un aspect plus pratique, l'utilisation d'un unique détecteur permet de grandement réduire les coûts du dispositif, surtout pour de l'imagerie dans d'autres domaines spectraux que le visible où les capteurs peuvent être très onéreux.

Le dispositif d'acquisition ayant été présenté dans cette partie, il convient maintenant de proposer des méthodes de reconstruction d'images. Pour cela, on montre dans la prochaine partie comment reconstruire l'image d'un objet qui ne bouge pas au cours du temps quand on a le même nombre de mesures que de pixels, *i.e* $K = N$.

3 Cas statique : reconstruction d'un objet immobile

On considère dans cette partie un objet immobile. Le but est de reconstruire une image à l'aide de la caméra mono-pixel. Quand $K = N$, cela revient à résoudre le problème inverse :

$$\mathbf{H}\mathbf{f} = \mathbf{m} \quad (6)$$

Les motifs de Hadamard vérifient une condition d'orthogonalité, ce qui permet d'avoir une formule explicite pour reconstruire l'image :

$$\mathbf{f} = \frac{1}{N} \mathbf{H}^T \mathbf{m} \quad (7)$$

3.1 Reconstruction sans bruit

On simule des mesures sur le fantôme de Shepp-Logan en faisant le produit scalaire entre les motifs de Hadamard et l'image de référence. Dans un premier temps, aucun bruit lié au processus d'acquisition des mesures n'est considéré et on souhaite reconstruire une image 64×64 . On considère donc $N = 4096$ échantillons.

On obtient les résultats de la Figure 3, où l'image de référence 3a utilisée pour simuler les mesures 3b à partir desquelles on peut reconstruire le fantôme en 3c avec la formule (7). Le fantôme est parfaitement reconstruit, ce qui est un résultat attendu quand on se sert des même ingrédients pour acquérir les mesures et pour l'inversion du problème : on parle d'« inverse crime ». Ce cas n'étant pas représentatif de la réalité, on souhaite modéliser le bruit présent sur les mesures réelles.

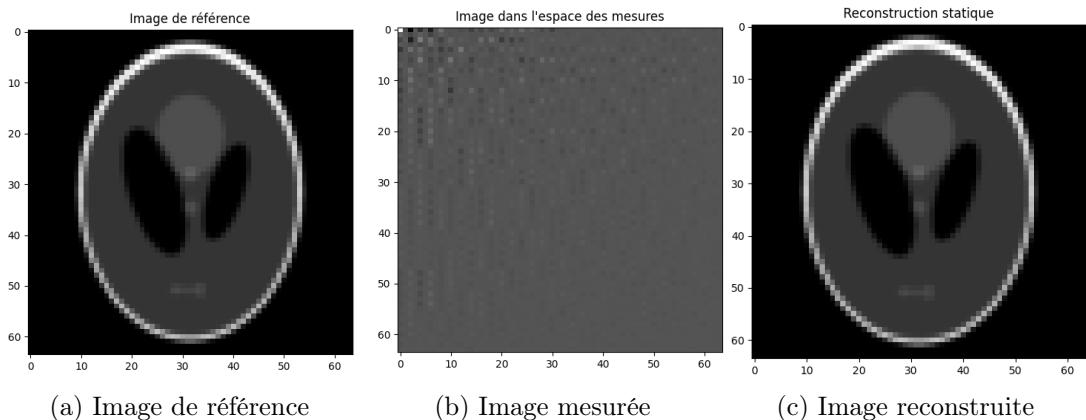


FIGURE 3 – Reconstruction du fantôme de Shepp-Logan immobile

3.2 Présence d'un bruit sur les mesures

Après s'être réfléchis sur le DMD, les rayons lumineux passent au travers d'une lentille convergente pour être sommés au niveau de la caméra mono-pixel. En réalité, seul un nombre aléatoire de photons sont captés par le détecteur. Pour compter le nombre de photons perçus par le détecteur pendant un certain intervalle de temps, une loi de Poisson est utilisée et on considère que les mesures brutes sont données par la distribution $\mathcal{P}(\alpha\mathbf{m})$ avec α un réel indiquant le niveau de bruit. Dans cette étude, il a été proposé [16] de travailler sur des mesures normalisées pour pouvoir faire varier le niveau de bruit en préservant l'intensité de l'image sous-jacente \mathbf{f} :

$$\hat{\mathbf{m}} = \frac{1}{\alpha} \mathcal{P}(\alpha\mathbf{m}) \quad (8)$$

où $\hat{\mathbf{m}}$ est le vecteur des mesures bruitées. En regardant de plus près les caractéristiques de cette loi, on voit que $\mathbb{E}(\hat{\mathbf{m}}) = \mathbb{E}(\mathbf{m})$ et $\mathbb{V}(\hat{\mathbf{m}}) = \frac{1}{\alpha}\mathbb{V}(\mathbf{m})$. Ainsi, plus la valeur de α augmente et plus la variance du signal diminue, représentant ainsi un niveau de bruit plus faible. Ce phénomène est illustré sur la Figure 4, où l'image reconstruite pour un α plus faible est la plus bruitée.

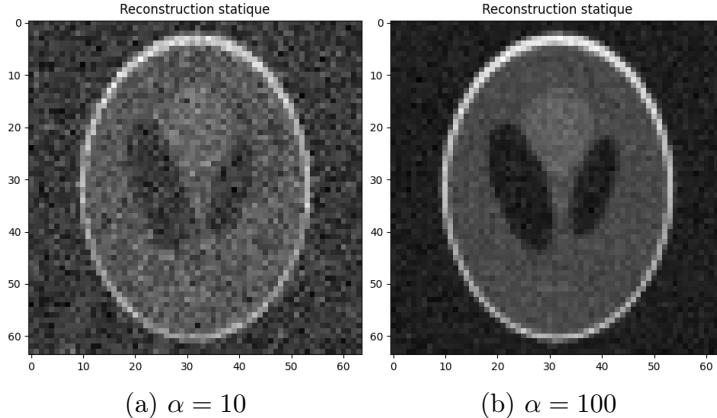


FIGURE 4 – Reconstructions avec différents niveaux de bruit sur les mesures

Les reconstructions de la Figure 4 sont mauvaises étant donné que la reconstruction du fantôme est parfaite quand il n'y a aucun bruit. On voudrait donc maintenant utiliser une technique qui permettrait de reconstruire l'image de référence tout en considérant la présence d'un bruit au niveau du détecteur lors de l'acquisition des mesures.

3.3 Régularisation TV

Dans le but d'éliminer le bruit lors du processus de reconstruction, une approche variationnelle du problème est adoptée. Cela permet d'ajouter un a priori sur la solution pour la contraindre à respecter certaines propriétés. Ainsi, on considère le problème :

$$\min_{\mathbf{f}} \frac{\eta}{2} \|\mathbf{H}\mathbf{f} - \hat{\mathbf{m}}\|_2^2 + \|\nabla \mathbf{f}\|_1 \quad (9)$$

où η est un paramètre à déterminer.

En examinant la fonctionnelle à minimiser de plus près, on peut remarquer qu'elle est composée de deux sous-problèmes bien connus. Le premier terme est le résidu du problème direct $\mathbf{H}\mathbf{f} = \hat{\mathbf{m}}$ et agit comme un terme d'attache aux données. Minimiser la norme L^2 du résidu constitue le problème des moindres carrés. Le deuxième terme est la norme L^1 du gradient de \mathbf{f} . Dans l'idée, l'utilisation de l'opérateur de gradient permet de minimiser les contrastes de l'image. Les approches utilisant un tel terme de régularisation portent le nom de régularisation TV (« total-variation »). Le modèle de débruitage TV a été d'abord introduit en 1992 par Rudin et al. [25] et est communément appelée ROF du nom de ses auteurs. Des algorithmes ont depuis été développés pour résoudre efficacement le ROF, comme celui de Chambolle [26] qu'on utilisera dans ce travail.

Les deux sous-problèmes étant beaucoup plus simples à résoudre que le problème initial, on propose dans ce travail un algorithme de minimisation alternée (une implémentation Python est disponible en Annexe 9.1) :

$$\mathbf{f}_{k+\frac{1}{2}} = \mathbf{f}_k - \tau \mathbf{H}^T (\mathbf{H}\mathbf{f}_k - \hat{\mathbf{m}}) \quad (10a)$$

$$\mathbf{f}_{k+1} = \min_{\mathbf{f}} \|\nabla \mathbf{f}\|_1 + \frac{\eta}{2\tau} \|\mathbf{f} - \mathbf{f}_{k+\frac{1}{2}}\|_2^2 \quad (10b)$$

Lors de la première étape de la minimisation (10a), on utilise une descente de gradient de pas constant τ pour le problème des moindres carrés. Une fois s'être un petit peu rapproché de la solution, on débruite l'image intermédiaire obtenue avec (10b) en utilisant l'algorithme de Chambolle.

Afin d'illustrer l'effet de la régularisation TV, on fait des simulations de reconstruction en faisant varier η en Figure 5. Pour assurer la convergence, il suffit que $\tau \leq \frac{1}{\|\mathbf{H}\mathbf{H}^T\|} = \frac{1}{N}$ [27]. On prend $\tau = \frac{1}{2N}$ pour les simulations. On évalue l'exactitude des reconstructions grâce au PSNR (« peak signal-to-noise ration »). Ce dernier est une fonction de l'erreur quadratique entre la reconstruction \mathbf{f} et l'image de référence \mathbf{f}_{ref} . Il se calcule avec la formule :

$$\text{PSNR}(\mathbf{f}, \mathbf{f}_{\text{ref}}) = 10 \log_{10} \frac{2^2}{\|\mathbf{f} - \mathbf{f}_{\text{ref}}\|^2} \quad (11)$$

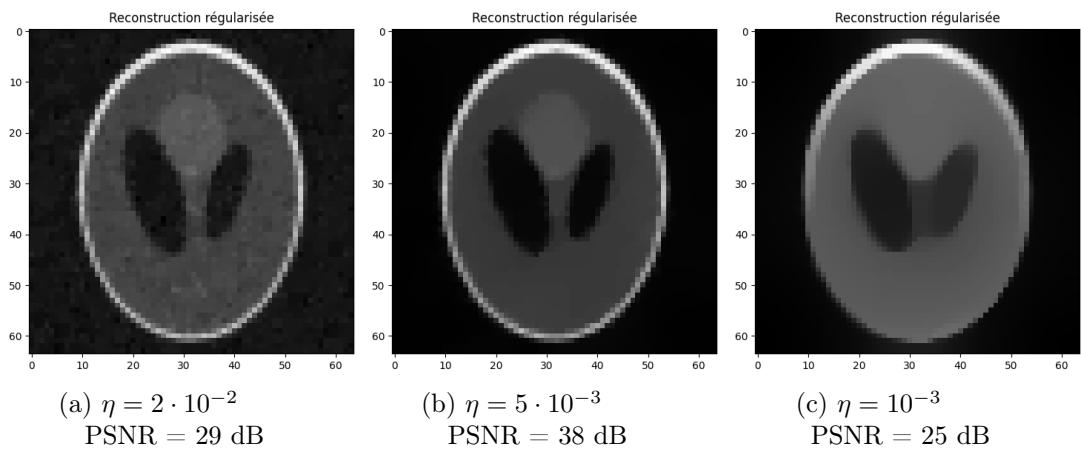


FIGURE 5 – Évolution de la reconstruction en fonction de η avec un bruit de Poisson $\alpha = 100$

On voit sur la Figure 5 l'effet du paramètre η sur la régularisation TV. Si η est trop grand comme en Figure 5a, le terme d'attache aux données est trop important et l'image reconstruite est encore bruitée. En revanche si η est trop petit comme en Figure 5c, l'image reconstruite est trop lissée et on perd de l'information (la tâche au dessus des deux alvéoles centrales devient invisible). Il est donc important de choisir une valeur du paramètre de régularisation qui ne soit ni trop grande, ni trop petite, afin de reconstruire des images débruitées contenant les informations importantes de l'image comme en Figure 5b.

Après avoir vu l'influence de η sur les reconstructions, il est intéressant de regarder l'effet de la méthode de régularisation proposée pour un niveau de bruit de Poisson plus important ($\alpha = 10$) en Figure 6 pour un η bien choisi. On observe une nette amélioration visuelle entre les Figures 6a et 6b qui se traduit par une hausse du PSNR.

Dans ce travail, le paramètre η est déterminé de manière empirique. Cependant, des techniques qui permettent de trouver ce paramètre automatiquement existent [28] et il pourrait être intéressant de poser cette question pour de futurs travaux. Une autre piste intéressante à explorer serait d'adapter le terme d'attache aux données pour qu'il puisse rendre compte du bruit de Poisson [29]. On notera que la technique de régularisation proposée n'en est qu'une parmi tant d'autres, qu'il serait impossible d'énumérer dans ce rapport. En restant dans le cadre variationnel, on pourrait ajouter différents a priori sur l'image comme celui proposé par Tikhonov.

L'étude du cas statique effectué dans cette partie a permis de voir comment reconstruire une image à partir de la série de mesures acquises par la caméra mono-pixel. Des simulations ont été

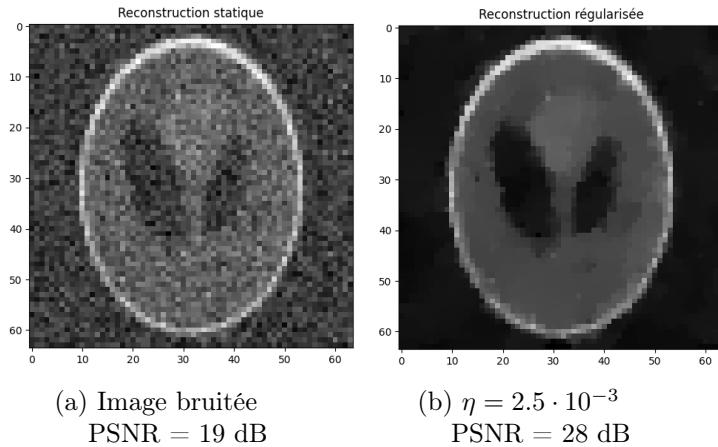


FIGURE 6 – Régularisation TV avec un bruit de Poisson $\alpha = 10$

réalisées pour valider les résultats sur des cas "parfaits" ainsi que sur des cas plus réalistes où l'on considère du bruit au niveau du détecteur. Enfin, une méthode de régularisation TV basée sur un principe de minimisation alternée a été proposée qui sera également utilisée dans la suite du travail. Maintenant que le cadre général est bien posé et qu'on a montré comment reconstruire une image d'une scène immobile, on s'intéresse dans la prochaine partie à l'élaboration d'une stratégie de compensation de mouvement dans le cas d'un objet qui bouge.

4 Cas dynamique : reconstruction d'un objet en mouvement

On considère à présent le cas d'un objet animé d'un mouvement. En se plaçant de nouveau dans un cadre continu, on veut reconstruire $f \in L^2([0, T_{acq}] \times X)$ où T_{acq} est le temps d'acquisition. La déformation $u_t : X \rightarrow \tilde{X}$ qui a lieu entre l'instant de référence et l'instant t est supposée connue, de telle sorte qu'il soit possible d'écrire :

$$f_t = f_{\text{ref}} \circ u_t \quad (12)$$

Pour $k \in \llbracket 0, K - 1 \rrbracket$, le processus d'acquisition reste le même et on peut écrire :

$$m_k = \langle h_k, f_{t_k} \rangle_{L^2(X)} \quad (13)$$

L'image f_{t_k} de l'équation (13) pouvant être différente à chaque instant t_k , on se doute qu'il va être difficile de reconstruire l'image de référence de la même façon que dans le cas d'une image immobile. Ceci est illustré à l'aide de la Figure 7 où les reconstructions obtenues avec la formule (7) sont montrées dans le cas d'un objet qui ne bouge pas et d'un objet animé d'un mouvement affine.

Pour les simulations sur le fantôme de Shepp-Logan, on commence par choisir un modèle de mouvement affine périodique. On a :

$$u_t(x) = c + \begin{pmatrix} s(t) & 0 \\ 0 & s(t)^{-1} \end{pmatrix} (x - c) \quad (14)$$

où $s(t) = 0.05 \cos(\frac{2\pi t}{T}) + 0.95$ et c est le centre de l'image.

Le mouvement ainsi défini est périodique, centré sur le centre du fantôme et d'amplitude 3.2 pixels. Une animation de la déformation est disponible sur un répertoire GitHub ¹.

1. url : https://github.com/Tommte/sp_moco_demo/blob/main/animations/deformations/deformation_1.gif

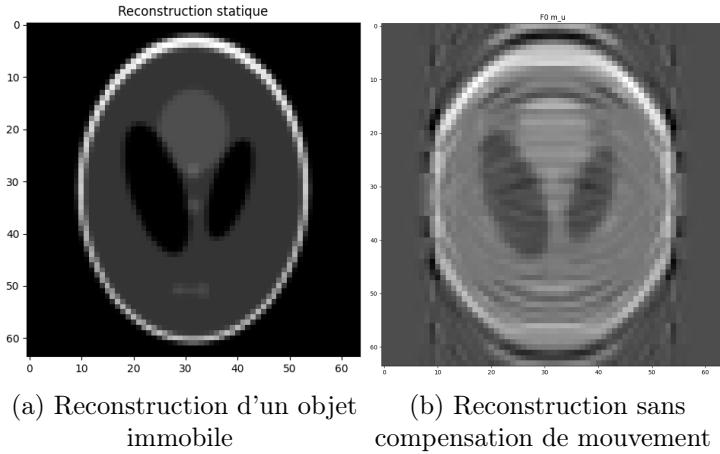


FIGURE 7 – Importance d'une méthode de compensation de mouvement

La Figure 7 montre la nécessité d'une stratégie de compensation de mouvement. Alors qu'il est possible de reconstruire parfaitement un objet qui ne bouge pas en Figure 7a, de forts artefacts de mouvement qui détériorent l'information du signal apparaissent sur la Figure 7b.

4.1 Méthode de compensation du mouvement

Deux approches sont adoptées pour effectuer la compensation du mouvement. La première suit le mouvement de l'image au cours de sa déformation tandis que la deuxième fixe l'image à un instant de référence et observe la déformation sur les motifs au fur et à mesure du temps. En employant des termes de la mécanique des fluides, on référera donc à ces deux approches en utilisant les termes d'approches Lagrangienne et Eulérienne respectivement. Le but de ces deux formalismes est d'arriver à construire des matrices englobant le caractère dynamique du problème, c'est-à-dire telles que :

$$\mathbf{m} = \mathbf{H}_{\text{dyn}} \mathbf{f}_{\text{ref}} \quad (15)$$

Une fois le problème mis sous la forme (15), l'accès à l'image de référence est rendu possible en "inversant" la matrice dynamique.

4.1.1 Approche Lagrangienne

D'après le processus d'acquisition,

$$m_k = \int_X h_k(x) f_{\text{ref}}(u_{t_k}(x)) dx \quad (16)$$

En considérant que l'interpolation de f_{ref} sur la grille déformée par u_{t_k} est une application linéaire, pour tout $k \in \llbracket 0, K - 1 \rrbracket$ on peut construire une matrice d'interpolation $\mathbf{U}_{t_k} \in \mathbb{R}^{N \times N}$ telle qu'on ait :

$$m_k = \mathbf{H}_k \mathbf{U}_{t_k} \mathbf{f}_{\text{ref}} = [\mathbf{H}_{\text{Lag}}]_k \mathbf{f}_{\text{ref}} \quad (17)$$

La relation (17) permet alors de construire une première matrice dynamique ligne par ligne en faisant le produit du k -ème motif de Hadamard avec la matrice d'interpolation \mathbf{U}_{t_k} . Afin de déterminer cette matrice d'interpolation, on écrit l'interpolation de la fonction de référence sous forme continue. En notant $x_j^c \in X$ le centre du pixel j , on a pour $x \in X$:

$$f_{\text{ref}}(x) = \sum_{j=0}^{N-1} f_{\text{ref}}^j \beta(x - x_j^c) \quad (18)$$

où f_{ref}^j est une représentation discrète de f_{ref} et β est la fonction d'interpolation utilisée. En évaluant l'équation (18) sur les coordonnées du centre du pixel i déformé par u_{t_k} , on obtient :

$$f_{t_k}(x_i^c) = \sum_{j=0}^{N-1} f_{\text{ref}}(x_j^c) \beta(u_{t_k}(x_i^c) - x_j^c) \quad (19)$$

On déduit de (19) la matrice d'interpolation :

$$\mathbf{U}_{t_k}^{i,j} = \beta(u_{t_k}(x_i^c) - x_j^c) \quad (20)$$

En pratique, la méthode d'interpolation bilinéaire est utilisée. Ce choix consiste à considérer des fonctions triangle t.q. pour $x = (x_1, x_2)^T \in X$, $\beta(x_1, x_2) = t(x_1)t(x_2)$ où $t : [0, 1] \rightarrow \mathbb{R}$ est définie par :

$$t(x) = \begin{cases} n(x + \frac{1}{n}) & \text{si } x \in [-\frac{1}{n}, 0] \\ n(\frac{1}{n} - x) & \text{si } x \in [0, \frac{1}{n}] \\ 0 & \text{sinon} \end{cases} \quad (21)$$

4.1.2 Approche Eulérienne

Une deuxième façon de procéder est d'utiliser $v_t = u_t^{-1}$. Pour cela, on effectue un changement de variable lors du processus d'acquisition. Pour $k \in \llbracket 0, K-1 \rrbracket$:

$$\begin{aligned} m_k &= \int_X h_k(x) f_{t_k}(x) dx \\ &= \int_X h_k(x) f_{\text{ref}}(u_{t_k}(x)) dx \\ &= \int_{\tilde{X}} h_k(v_{t_k}(\tilde{x})) f_{\text{ref}}(\tilde{x}) |\det \nabla v_{t_k}(\tilde{x})| d\tilde{x} \\ &= \langle |\det \nabla v_{t_k}| h_k \circ v_{t_k}, f_{\text{ref}} \rangle_{L^2} \end{aligned} \quad (22)$$

En discréteisant (22), on obtient une deuxième façon d'obtenir une matrice dynamique \mathbf{H}_{Eul} ligne par ligne :

$$m_k = [\mathbf{H}_{\text{Eul}}]_k f_{\text{ref}} \quad (23)$$

où $[\mathbf{H}_{\text{Eul}}]_{k,j} = |\det \nabla v_{t_k}(x_j)| h_k(v_{t_k}(x_j))$.

Avoir deux formalismes reposant sur les déformations directes et inverses est intéressant en particulier parce-que les algorithmes d'estimations du mouvement permettent d'obtenir l'une ou l'autre de ces déformations.

4.2 Reconstruction dynamique sur les données simulées

On implémente ces deux approches pour le problème de reconstruction dynamique sur des données issues de la simulation dans le but de valider la méthode. On utilise encore le fantôme de Shepp-Logan de taille 64×64 comme en section 3. On commence par reconstruire avec la méthode des moindres carrés après construction de la matrice dynamique \mathbf{H}_{Lag} obtenue avec l'approche Lagrangienne. Le résultat est montré en Figure 8.

On montre en Figure 8a que la reconstruction dynamique Lagrangienne par la méthode des moindres carrés fonctionne parfaitement dans le cas de l'« inverse crime », qui pour rappel n'arrive jamais en pratique. En revanche, cela permet déjà de valider la méthode de compensation de mouvement proposée puisque les forts artefacts présents sur la Figure 7b n'apparaissent plus.

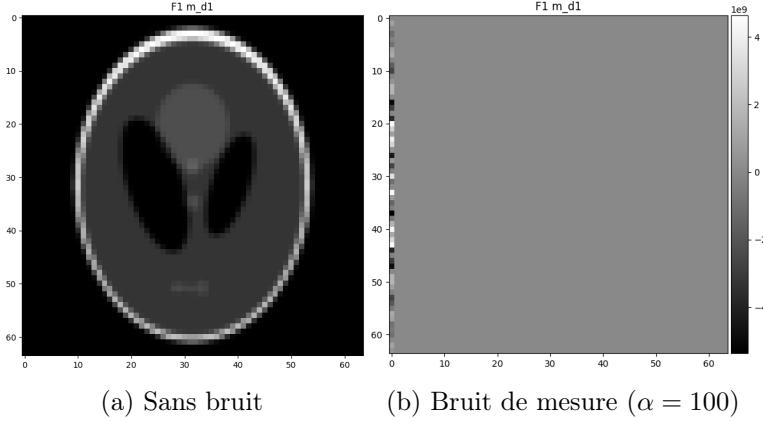


FIGURE 8 – Reconstruction dynamique Lagrangienne par les moindres carrés

On essaie d'effectuer la même reconstruction dans le cas d'un léger bruit de Poisson sur les mesures en Figure 8b, mais le résultat est décevant : on obtient un image dont les intensités des pixels ont explosés en certains points, rendant l'image totalement illisible. Ce phénomène est caractéristique des problèmes inverses mal posés introduits par Jacques Hadamard en 1902 [30]. Dans le cas présent, bien que les conditions d'existence et d'unicité semblent être vérifiées il est probable que la solution ne soit pas stable (*i.e.* ne dépende pas continûment des données). Il devient alors essentiel d'utiliser des méthodes de régularisation pour résoudre le problème de reconstruction dynamique. Une rapide étude du spectre de la matrice dynamique construite et une première façon de régulariser qui en résulte sont présentées en Figure 9.

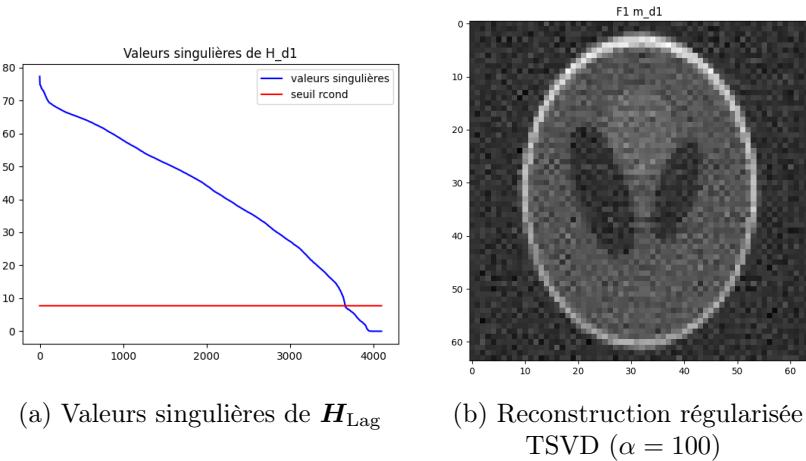


FIGURE 9 – Une première façon de régulariser

La Figure 9a montre les valeurs singulières de \mathbf{H}_{Lag} en bleu. Les faibles valeurs singulières posent problème : en notant la SVD de $\mathbf{H}_{\text{Lag}} = \mathbf{U}\Sigma\mathbf{V}^T$, la norme solution des moindres carrés $\|\mathbf{f}\|_{LS}^2 = \sum_{i=0}^{N-1} (\frac{\mathbf{u}_i^T \mathbf{m}}{\sigma_i})^2$ (avec \mathbf{u}_i la i -ème ligne de \mathbf{U} et σ_i la i -ème valeur singulière) explose quand σ_i est petit. La droite rouge, placée à une valeur seuil égale à 10 % de la valeur singulière maximale, sert de démarcation pour ces faibles valeurs singulières. Les valeurs singulières en dessous de ce seuil sont mises à zéro. Cette méthode de SVD tronquée (TSVD) [31] permet régulariser la solution de manière très similaire à la régularisation de Tikhonov. Le résultat de cette première méthode de régularisation est montré en Figure 9b. On voit qu'il est alors possible de retrouver le fantôme de Shepp-Loggan, cependant la méthode est assez décevante car elle ne propose pas assez de lissage pour la solution.

On adapte donc la méthode de régularisation TV vue en partie 3.3 en cherchant maintenant à minimiser $\mathcal{L}(f) = \|\nabla f\|_1 + \frac{\eta}{2} \|H_{d_i} f - \hat{m}\|_2^2$ pour $i \in \{1, 2\}$. En déterminant empiriquement le paramètre η , les résultats de la Figure 10 sont obtenus.

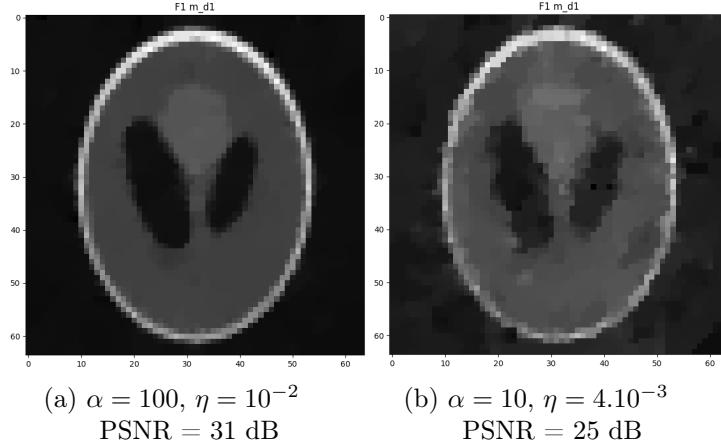


FIGURE 10 – Reconstruction dynamique Lagrangienne avec régularisation TV

La Figure 10 montre des exemples de reconstruction avec compensation de mouvement pour deux niveaux de bruit sur les mesures. La première chose à remarquer est que la méthode employée permet toujours de compenser le mouvement même en présence de bruit de Poisson. Ensuite, il est intéressant de comparer ces résultats avec ceux obtenus dans le cas statique en Figures 5b et 6b. On peut dire que la technique de compensation permet d'obtenir des reconstructions décentes à l'oeil nu, bien qu'on puisse relever une baisse du PSNR quand l'objet à imager bouge.

La convergence de l'algorithme de reconstruction avec régularisation est atteinte très rapidement. En effet, on voit qu'une dizaine d'itérations suffisent pour que la fonction de perte arrive à un palier en Figure 11. On peut aussi améliorer la convergence de la descente de gradient (ISTA) avec la variante FISTA [27].

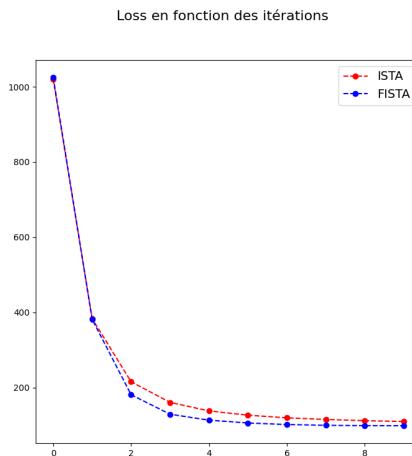


FIGURE 11 – Fonction de perte \mathcal{L} en fonction du nombre d'itérations

Après avoir confirmé que l'approche Lagrangienne de la méthode de reconstruction dynamique marchait avec des données simulées, c'est au tour de l'approche Eulérienne d'être mise à

l'épreuve. On obtient les résultats présentés en Figure 12.

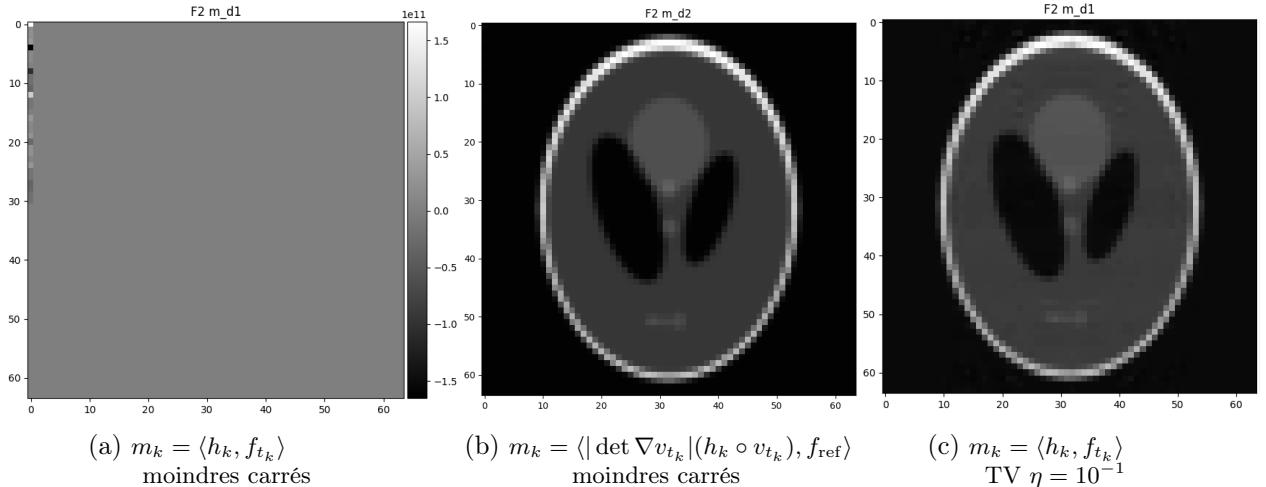


FIGURE 12 – Reconstruction dynamique Eulérienne sans bruit ajouté

On voit en Figure 12a que la reconstruction dynamique Eulérienne par la méthode des moindres carrés sans bruit de Poisson sur les mesures semble être mal posée. Ce résultat étonnant est dû à une légère différence entre les discrétisations des deux approches du problème dynamique. Autrement dit, les mesures utilisées pour la méthode de reconstruction Eulérienne sans bruit sont déjà légèrement bruitées par rapport à celles qu'il faudrait utiliser pour être dans la configuration de l'« inverse crime ». Ce phénomène est illustré à l'aide de la Figure 12b où la reconstruction faite à partir des mesures simulées en utilisant les motifs de Hadamard déformés par la déformation inverse (équation (22)) est parfaite. On voit alors que cette technique de compensation de mouvement fonctionne dans le sens où elle permet d'éradiquer les artefacts de la Figure 7b lors du processus de reconstruction.

Cela étant dit, les mesures déformées par v_t n'ont pas de réel sens physique puisque l'on ne récupère en pratique que les produits scalaires entre les motifs de Hadamard h_k et la scène à l'instant t_k . Considérer le cas parfait de la Figure 12b n'a donc pas de réel intérêt d'un point de vue physique et ne sert qu'à valider l'implémentation de l'algorithme. Toutefois, on peut voir avec la Figure 12c que le bruit causé par la différence entre les discrétisations est en fait négligeable et peut être corrigé avec un tout petit peu de régularisation. Dans le cas plus réaliste où le détecteur ne capte qu'un nombre aléatoire de photons et où on est naturellement amené à utiliser la méthode de régularisation TV, considérer les mesures directes n'aura donc aucun effet notable. Des résultats pour deux niveaux de bruit de mesure sont montrés en Figure 13.

La Figure 13 permet de valider les résultats en présence de bruit au niveau du détecteur. On obtient des résultats tout à fait similaires à ceux obtenus en Figure 10 et des scores de PSNR assez proches pour pouvoir être considérés identiques. Les deux formalismes peuvent donc être considérés équivalents en terme de précision des résultats. On utilisera l'un ou l'autre selon la connaissance de la déformation directe ou inverse par les méthodes d'estimation de mouvement. Pour la suite des simulations, sauf mention du contraire seuls les résultats de la reconstruction dynamique Lagrangienne seront présentés pour plus de concision.

4.3 Erreur sur le modèle de mouvement

Le modèle exact du mouvement n'est pas connu en pratique, mais il peut être estimé. Puisqu'il s'agit d'une estimation, il y a une erreur entre le modèle exact et le modèle estimé. On simule cette

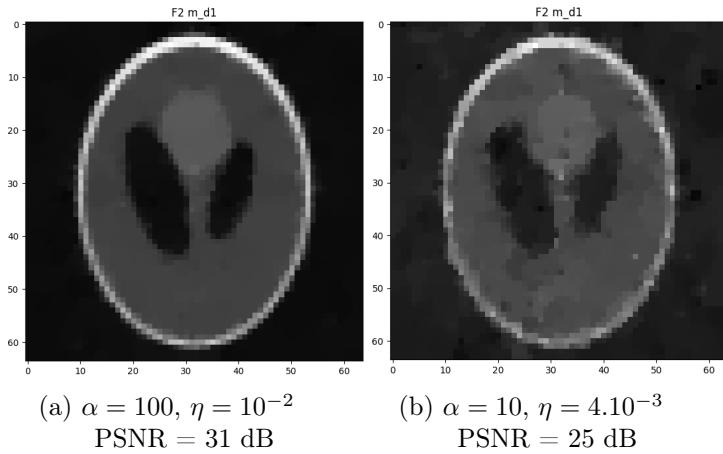


FIGURE 13 – Reconstruction dynamique Eulérienne avec régularisation TV

erreur en jouant directement sur des paramètres spécifiques du modèle de mouvement utilisé. Ainsi, on va considérer que le modèle estimé est donné par (14) où $s(t) = \mathbf{0.03} \cos(\frac{2\pi t}{T}) + \mathbf{0.97}$. L'erreur de mouvement est ensuite obtenue en utilisant les mesures obtenues par le premier modèle et les matrices dynamiques obtenues avec le second. Les résultats obtenus sont affichés sur la Figure 14.

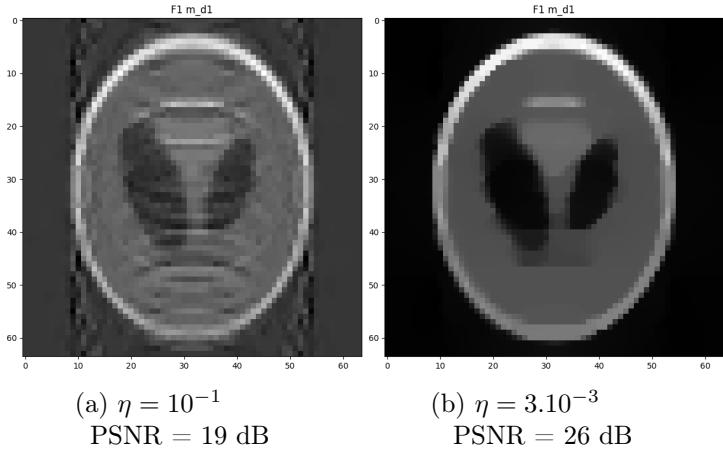


FIGURE 14 – Reconstruction dynamique TV avec un erreur sur le modèle de mouvement

La Figure 14 montre des reconstructions dynamiques pour deux niveaux de régularisation TV. Sur la Figure 14a, on choisit de peu régulariser la reconstruction. Cela permet d'observer d'artefacts de mouvement similaires, bien que d'intensités moindres, à ceux qui apparaissent quand l'image est reconstruite sans compensation du mouvement en Figure 7b. La Figure 14b prouve qu'il est possible d'effacer ces artefacts avec la méthode de régularisation proposée en diminuant la valeur du paramètre η pour obtenir une reconstruction passable. Il est important de remarquer que l'erreur sur le modèle de mouvement peut tout de même entraîner des pertes d'informations. Dans le cas présent, cela peut se voir avec un faible score PSNR ou même à l'oeil nu où le cercle central au dessus des deux alvéoles se retrouve coupé horizontalement en deux. Il est donc important de bien estimer le mouvement afin d'avoir le moins d'erreur possible.

4.4 Sous-échantillonage

La caméra mono-pixel permet de se placer très simplement dans le cadre de l'acquisition comprimée [1]. Le processus d'acquisition étant séquentiel, cela revient à retenir les $K \leq N$ premières mesures acquises pour reconstruire le signal. Il est ainsi possible de réduire les temps d'acquisition et de reconstruction, accélérant l'imagerie de l'objet de manière substantielle. De plus, dans certains cas l'acquisition comprimée permet également de réduire les bruits en éliminant certaines fréquences (*e.g.* les hautes fréquences) où les bruits sont plus présents.

Dans un cadre plus formel, on introduit une matrice de sous-échantillonage $S \in \mathbb{R}^{K \times N}$. On pourra prendre par exemple $S = [\mathbf{I}_K \ \mathbf{0}]$ pour sélectionner les K premiers motifs de Hadamard. Il s'agit maintenant de résoudre le système sous déterminé :

$$\mathbf{m} = S\mathbf{H}\mathbf{f} \quad (24)$$

On continue de résoudre ce système par la méthode de régularisation TV présentée. Cette méthode permet de résoudre le problème statique, et il suffit de changer la matrice de Hadamard par une des matrices de Hadamard dynamiques présentées en partie 4.1 pour résoudre le problème de sous-échantillonage dans le cas dynamique. L'effet du sous-échantillonage sur la reconstruction dynamique est montré en Figure 15 avec $\eta = 10^{-2}$.

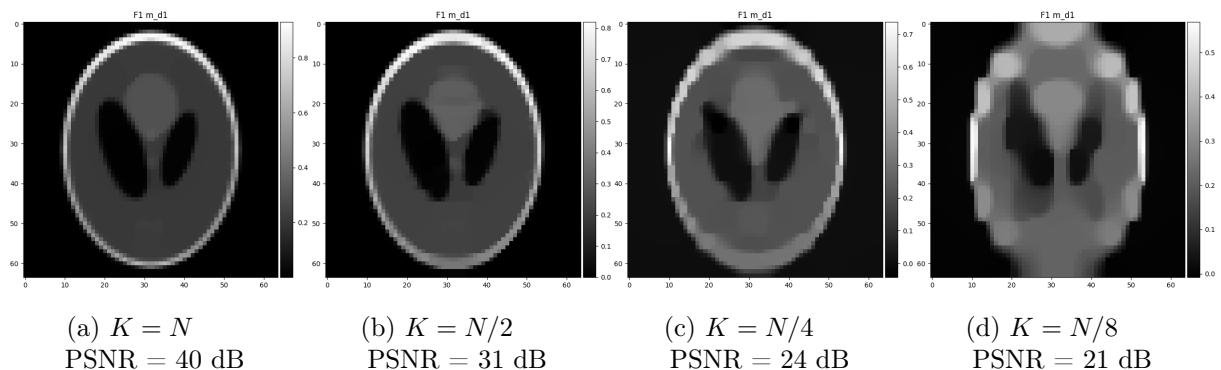


FIGURE 15 – Reconstruction dynamique TV sous-échantillonée

La Figure 15 montre l'évolution de la reconstruction dynamique en fonction du nombre d'échantillons retenus. Comme attendu, la qualité visuelle de la reconstruction et le PSNR baissent au fur et à mesure que K diminue. Puisque le choix de la base de Walsh a été fait dans cette partie, retenir les premières mesures revient à sélectionner les plus basses fréquences du signal. Comme les hautes fréquences sont responsables des détails d'une image, il est normal d'observer de bonnes, bien que plus grossières, reconstructions du fantôme de Shepp-Logan en Figures 15b et 15c. En revanche il est de plus en plus difficile de reconnaître le fantôme quand on commence à enlever les basses fréquences du signal comme en Figure 15d. On observe par ailleurs une diminution linéaire du temps de calcul (voir Annexe 9.2), un résultat attendu par la nature séquentielle de l'acquisition.

Dans cette partie, on a pu se plonger dans le coeur du sujet en considérant un objet animé d'un mouvement. Deux approches, qui se distinguent par l'usage de la déformation directe ou inverse, ont été utilisées pour permettre de reconstruire une image en compensant le mouvement. La méthode proposée a été validée par la simulation en prenant en compte un bruit au niveau du détecteur ainsi qu'une erreur sur l'estimation du mouvement. Enfin, les principes de l'acquisition comprimée ont pu être appliqués avec la méthode de compensation de mouvement. Dans la prochaine partie, l'accent est mis sur les méthodes visant à estimer le mouvement tout en cherchant à évaluer avec précision l'erreur associée à ces estimations.

5 Estimation du mouvement

Cette partie se concentre sur l'estimation du mouvement de la scène étant donné un flux vidéo. Pour cela, on considère des données simulées en construisant une tâche blanche sur une grille 64×64 ayant 5 pixels de rayon et placée au centre de l'image. Cette tâche est ensuite déplacée selon un mouvement de translation donné par :

$$u_t(x) = x - d \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} \frac{t}{T_{acq}} \quad (25)$$

où $\theta \in \{\frac{\pi}{4}, \frac{3\pi}{4}\}$ est l'angle selon lequel s'effectue la translation et d représente la distance parcourue par l'objet pendant la période d'acquisition. Pour plus de similarité avec la partie expérimentale, le flux vidéo crée contient 586 frames. Le mouvement est alors connu tout les 6 ou 7 motifs de Hadamard et des formules d'interpolations sont nécessaires pour résoudre le problème. L'utilisation d'interpolation constante par morceaux ou linéaire à ce niveau là n'a pas beaucoup influé sur les résultats.

Puisque la vérité terrain du mouvement est connue, il est possible de calculer des erreurs sur l'estimation du mouvement. A chaque frame f , on commence par calculer le champ d'erreur tel que pour $x \in X$,

$$e^f(x) = \sqrt{e_{x_1}^f(x)^2 + e_{x_2}^f(x)^2} \quad (26)$$

où $e_{x_i}^f(x) = \tilde{v}_{x_i}^f(x) - v_{x_i}^f(x)$ avec $\tilde{v}_{x_i}^f$ une approximation de $v_{x_i}^f$ donnée par la méthode d'estimation du mouvement. Des scores L^2 et L^∞ sont ensuite calculés sur ce champ pour chaque frame :

$$\begin{cases} e_{L^2}^f = \frac{1}{\sqrt{N(N+1)}} \sqrt{\sum_{j=0}^{N-1} e^f(x_j)^2} \\ e_{L^\infty}^f = \max_{j \in [0, N-1]} (e^f(x_j)) \end{cases} \quad (27)$$

Cette première façon de faire permet de visualiser l'erreur au cours du temps. Par construction, elle vaut 0 dans le meilleur des cas et est supérieure à 1 dans le pire des cas (voir Annexe 9.3 pour plus de détails). Une indication globale sur l'erreur de l'estimation du mouvement est également proposée en utilisant les mêmes métriques sur le champ moyen d'erreur calculé par :

$$\bar{e}(x) = \frac{1}{n_{frames}} \sum_{f=0}^{n_{frames}-1} e^f(x) \quad (28)$$

Il est important de noter que le problème d'estimation du mouvement est mal posé, notamment car plusieurs champs de déformations peuvent donner lieu à une même déformation. Par exemple dans le cas de la tâche blanche soumise au mouvement de translation, il suffit de considérer les champs de mouvement résultants de la composition d'une rotation ayant pour origine le centre de la tâche et de la translation vérité-terrain. Bien qu'ils soient totalement différents, tous ces champs donnent des solutions correctes au problème d'estimation du mouvement. C'est pour cette raison qu'en général, les erreurs calculées à partir d'une vérité-terrain sont des mauvaises mesures pour évaluer l'estimation du mouvement. Cependant ces métriques sont valides dans ce travail car on prend soin d'utiliser des méthodes ayant un *a priori* permettant de trouver une solution s'approchant du mouvement de translation idéal donnée par le modèle (25).

5.1 Champ constant

Une première façon d'aborder le problème consiste à considérer le mouvement de l'ensemble de l'image comme étant rigide. On considère donc que le champ est constant dans l'espace à

chaque instant sur toute l'image. Ainsi, il suffit de repérer un point matériel de l'objet et de le suivre au cours du temps. En particulier pour l'objet considéré (une tâche blanche sur fond noir), on choisit de suivre la position du centre de masse de la tâche en projetant l'image sur ses deux axes et en repérant les pics d'intensités comme en Figure 16. Puisque la position d'un seul point est suivie, le problème est beaucoup mieux posé et la solution obtenue est proche de la vérité-terrain.

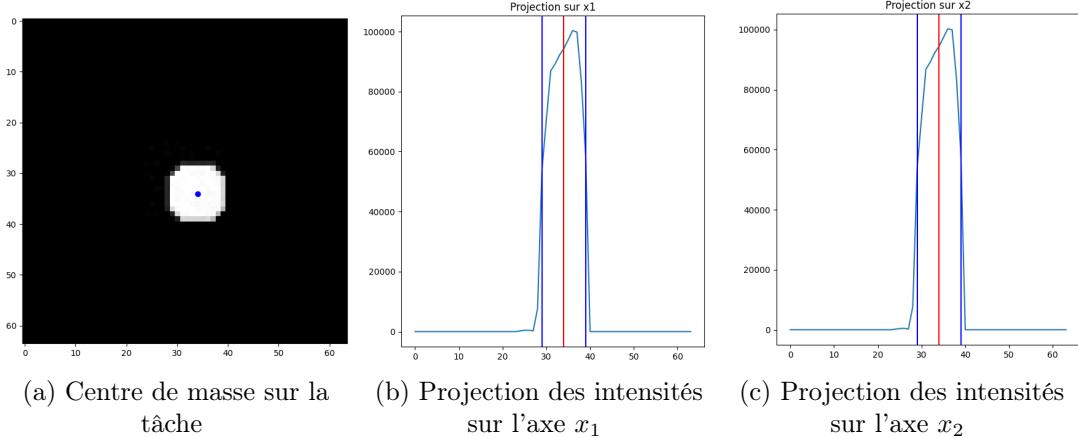


FIGURE 16 – Détermination du centre de masse sur un flux vidéo (snapshot sur la frame 60)

Sur la Figure 16a, on représente le centre de masse par un point bleu dont les coordonnées sont obtenues avec les Figures 16b et 16c. Plus précisément, pour $i \in \{1, 2\}$, on cherche la position moyenne dans l'intervalle où les intensités I_{x_i} sont supérieures à $0.5 \times \max_x(I_{x_i}(x))$ représenté ici par des lignes verticales bleues. Ensuite, l'erreur commise avec cette méthode d'estimation du mouvement est évaluée en Figure 17.

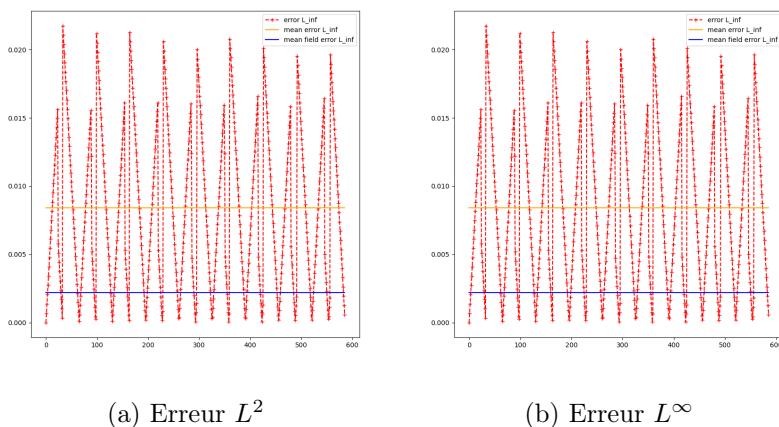


FIGURE 17 – Erreurs sur l'estimation par champ constant en fonction des frames

La première chose à remarquer sur la Figure 17 est que les erreurs obtenues sont assez faibles ($e_{L^2} \in [0, 3 \cdot 10^{-4}]$ et $e_{L^\infty} \in [0, 2 \cdot 10^{-2}]$), ce qui indique une bonne estimation du mouvement. Ensuite, on peut observer de fortes oscillations des erreurs au cours des frames. Ce phénomène résulte de l'utilisation d'une méthode d'interpolation permettant de déplacer l'objet à une échelle sub-pixel, tandis que le centre de masse est quantifié au niveau du pixel le plus proche. Une illustration de l'interpolation utilisée est donnée en Annexe 9.4. On constate aussi que l'erreur sur le champ moyen (ligne bleue) est bien plus faible que la moyenne des erreurs sur chaque frame (ligne orange), ce qui montre que la méthode sous-estime autant qu'elle sur-estime le mouvement.

La méthode proposée est donc appropriée pour donner une première approximation du mouvement de la tâche blanche. Cependant de manière générale, les objets considérés ne sont pas aussi simples et bien souvent soumis à des mouvements plus complexes. Il est donc obligatoire de considérer une méthode d'estimation de mouvement plus générale.

5.2 Algorithme de flot optique

Les méthodes de flot optiques reposent sur le fait que l'intensité d'un point matériel reste constante au cours du temps. Cela s'écrit :

$$I(x_1 + dx_1, x_2 + dx_2, t + dt) = I(x_1, x_2, t) \quad (29)$$

Il est alors possible d'établir l'équation du flot optique après un développement de Taylor à l'ordre 1 sur l'équation (29). On obtient :

$$\nabla_{x_1, x_2} I \cdot \mathbf{u} + \partial_t I = 0 \quad (30)$$

avec $\mathbf{u} = (u_1, u_2)^T = (\frac{dx_1}{dt}, \frac{dx_2}{dt})^T$. Pour chaque pixel, on est donc ramené à une équation à deux inconnues qu'on ne peut pas résoudre sans hypothèses supplémentaires.

Une première approche mène à la méthode de Lucas-Kanade [32] en faisant l'hypothèse que les intensités sur un voisinage du pixel sont égales. Bien souvent, cette méthode est utilisée sur des points clés bien choisis comme ceux de Shi-Tomasi [33]. Pour les besoins de l'étude, il est préférable d'utiliser des méthodes de flot optiques denses comme celle de Gunnar-Farnebäck [34] qui considèrent que le signal est quadratique sur un voisinage du pixel considéré.

Une deuxième manière de faire repose sur une approche variationnelle du problème de flot optique. Dans leur article séminal [14], Horn et Schunck proposent un terme de d'attache aux données L^2 et un terme de régularisation $\mathcal{R}(\mathbf{u}) = |u_1|^2 + |u_2|^2$ promouvant la régularité de la solution. D'autres méthodes ont depuis été développées, et en particulier la méthode TV- L^1 [35] avec un terme d'attache aux données L^1 et un terme de régularisation $\mathcal{R}(\mathbf{u}) = |u_1| + |u_2|$ permettant de mieux capter les discontinuités dans les mouvements de grande amplitude.

Pour la tâche blanche, la méthode TV- L^1 est adoptée pour capter le grand déplacement de l'objet. De plus, le terme de régularisation choisi permet de chercher des solutions ayant de faibles gradients, ce qui est adapté à un simple mouvement de translation. De la même façon que pour la méthode du champ constant, on évalue les erreurs d'estimation du mouvement en Figure 18.

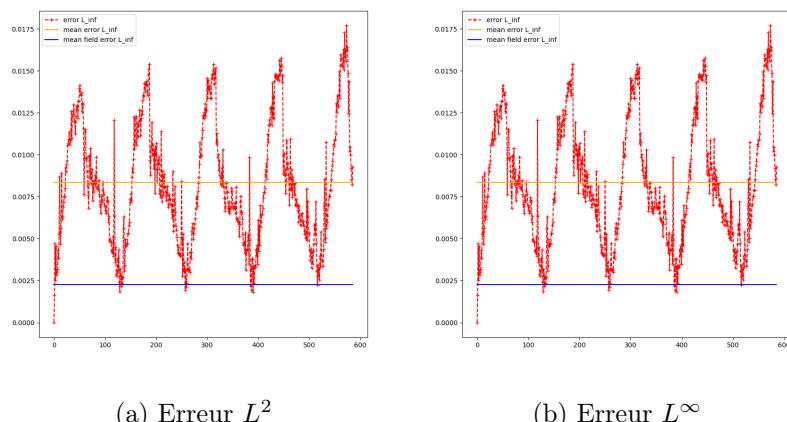


FIGURE 18 – Erreurs sur l'estimation par la méthode de flot optique TV- L^1 en fonction des frames

Un premier constat après avoir regardé la Figure 18 est que les erreurs obtenues sont plus faibles que celles données par la méthode du champ constant. En effet, $e_{L^2} \in [0, 1.6 \cdot 10^{-4}]$ et $e_{L^\infty} \in [0, 1.8 \cdot 10^{-2}]$, ce qui indique que l'estimation du mouvement est globalement meilleure avec un algorithme de flot optique qu'avec la méthode précédente. En comparant l'allure des erreurs de la Figure 18 avec celles de la Figure 17, on retrouve le même genre de pics qui sont sûrement dûs à la méthode d'interpolation. Ces pics sont plus espacés ce qui laisse supposer que la méthode $TV-L^1$ parvient à compenser l'effet de l'interpolation dans certains cas. De plus, on relève une erreur sur le champ moyen de $\bar{e}_{L^2} = 2.8 \cdot 10^{-5}$, ce qui est également inférieur à l'erreur sur le champ moyen relevée précédemment de $\bar{e}_{L^2} = 3.4 \cdot 10^{-5}$.

5.3 Une méthode d'apprentissage plus rapide

Bien souvent, les méthodes d'estimation du mouvement par flot optique nécessitent un temps de calcul plus grand que celui de la vidéo donnée en entrée. Une méthode basée sur des techniques d'apprentissage issue de l'article [15] est alors proposée pour estimer des mouvements répétitifs en temps réel pour la neurochirurgie. Plus précisément, la méthode suppose que le mouvement du cerveau peut être décrit dans un espace de faible dimension qu'on détermine en utilisant une ACP sur les premières frames de la vidéo. Il est possible de choisir l'une ou l'autre des méthodes de flot optique de Gunnar-Farnebäck ou $TV-L^1$ pendant la période d'apprentissage.

Pour le moment, l'application de cette méthode ne semble pas concluante sur la vidéo synthétique 64×64 de la tâche blanche. Cependant, l'algorithme a montré des résultats prometteurs sur les vidéos expérimentales de plus haute résolution en utilisant une grande taille de fenêtre, c'est à dire qu'on considère un grand voisinage autour de chaque pixel de l'image. Parvenir à faire fonctionner l'algorithme sera l'objet de futur travaux.

5.4 Résultats

Des simulations avec du bruit de Poisson ($\alpha = 100$) sur les mesures ont été effectuées. Le coefficient de régularisation est fixé à $\eta = 10^{-3}$ pour les reconstructions. Avec la connaissance exacte du mouvement de l'équation (25), on obtient une vérité terrain sur les reconstructions pour la tâche blanche. On montre les résultats obtenus en Figure 19.

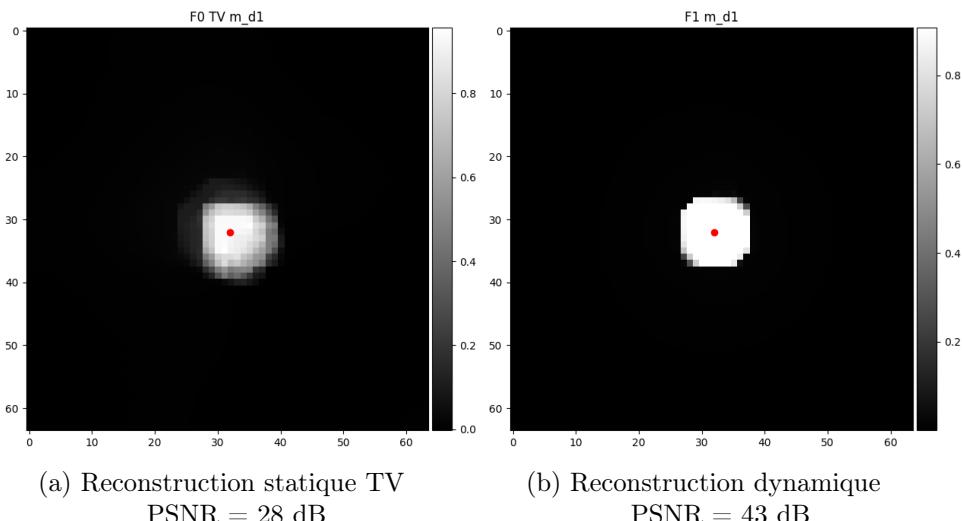


FIGURE 19 – Reconstructions de la tâche blanche connaissant le modèle de mouvement

Un point rouge est placé au centre de la tâche blanche à l'instant initial. Puisque ce point

n'est pas au centre de la tâche reconstruite sur la Figure 19a, il est une fois de plus clair que la méthode de reconstruction sans compensation du mouvement n'est pas bonne. En revanche, la reconstruction dynamique connaissant le modèle de mouvement exact montrée en Figure 19b montre que le mouvement est bien compensé car le point rouge est bien au centre de la tâche blanche. A présent, on utilise les méthodes d'estimation du mouvement vues dans cette partie pour reconstruire les images de la Figure 20.

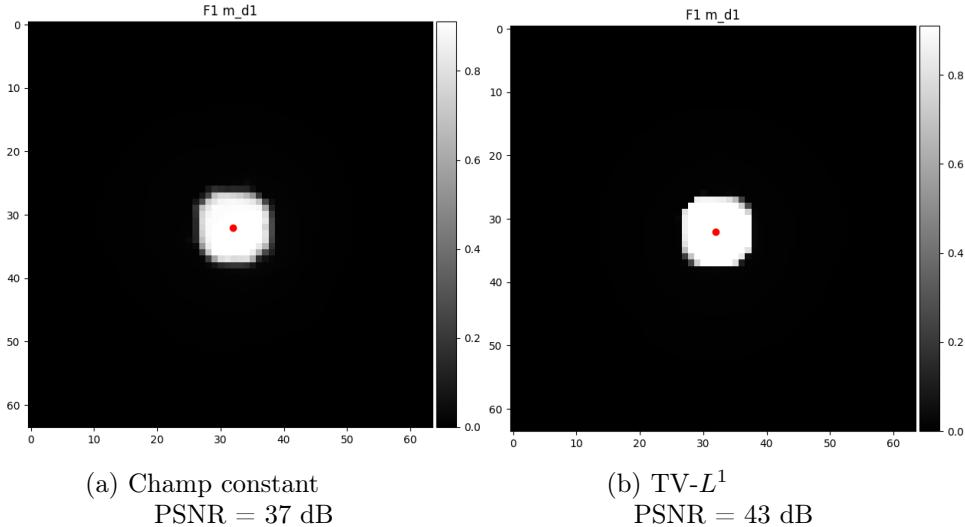


FIGURE 20 – Reconstruction dynamique de la tâche blanche avec deux méthodes d'estimation du mouvement

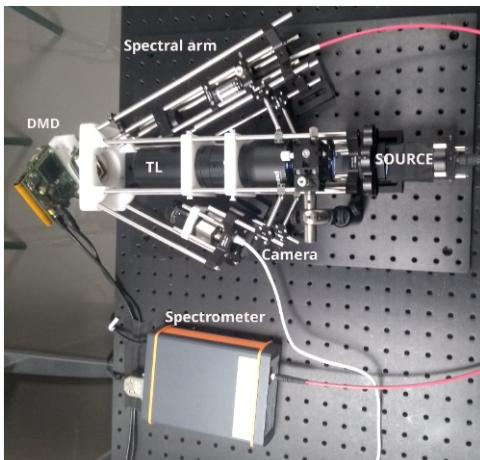
On affiche en Figure 20a la reconstruction dynamique obtenue avec la méthode de champ constant et en Figure 20b celle obtenue avec l'estimation $TV-L^1$. En vérifiant la position du point rouge (toujours initialement placé au centre de la tâche blanche) on voit que la compensation du mouvement marche bien dans les deux cas. Cependant, les bords de la tâche de la Figure 20a sont plus flous que ceux obtenus avec la méthode de flot optique, ce qui montre encore une fois l'intérêt d'utiliser cette méthode. Cela se voit de manière plus quantitative où en observant une hausse de PSNR de 6 dB entre les deux reconstructions.

Plusieurs méthodes ont été étudiées pour l'estimation du mouvement. Parmi elles, considérer le champ de mouvement comme étant constant dans l'espace et utiliser une méthode de flot optique $TV-L^1$ sont particulièrement efficaces pour estimer le mouvement d'une tâche blanche soumise à une translation. Bien que l'estimation obtenue soit meilleure avec la technique de flot optique, le gros avantage de la méthode est qu'elle permet d'estimer un champ de mouvement pour n'importe quel type d'objet et de mouvement. Maintenant que la méthode a été validée grâce à des simulations, on va s'intéresser à valider la méthode sur des données expérimentales.

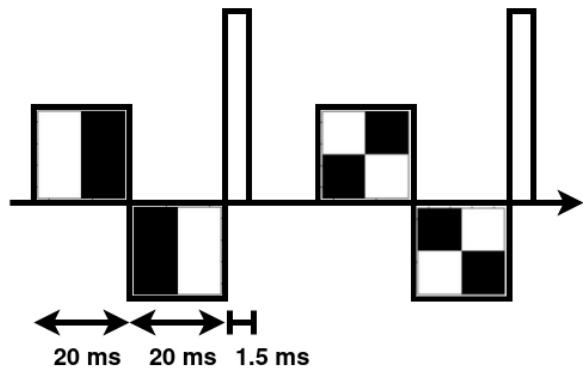
6 Reconstruction expérimentale

On souhaite maintenant appliquer la méthode de compensation de mouvement sur des données expérimentales. On réalise l'acquisition de ces données à l'aide du dispositif présenté en Figure 21a. On retrouve au centre du dispositif une source de lumière blanche LED (Thorlabs LIUCWHA) éclairant une matrice de micro-miroirs 1024×768 (ViALUX GmbH DLP V-700). D'un côté du DMD, on place la caméra mono-pixel qu'on relie à un spectromètre (Avantes AvaSpec-ULS2048CL-EVO) permettant de décomposer le signal reçu en son spectre. De l'autre

côté du DMD, on capture un flux vidéo de la scène à l'aide d'une caméra RGB. L'analyse de ce flux vidéo permet ensuite d'estimer le mouvement comme vu en section 5. Puisque le processus d'acquisition de la caméra mono-pixel est séquentiel, on se voit obligés d'intercaler des motifs « blancs », c'est à dire qui réfléchissent l'intégralité des rayons lumineux, entre les motifs de Hadamard comme en Figure 21b. De plus, on projette maintenant les motifs de Hadamard selon l'ordre de leurs variances décroissante.



(a) Dispositif expérimental



(b) Processus d'acquisition à deux caméras

FIGURE 21 – Présentation de l'acquisition expérimentale

Les données acquises par des membres de l'équipe sont ensuite enregistrées dans la collection d'hypercubes SPIHIM [21] et rendues disponibles sur la plateforme PILoT¹.

6.1 Présentation des données

On s'intéresse à reconstruire les images de deux objets filmés par la caméra mono-pixel en compensant le mouvement. Le premier objet est une simple tâche lumineuse blanche mesurant environ 2.2 cm de diamètre et qui sert à valider la méthode expérimentale sur un cas simple. Une fois ceci fait, on voudrait appliquer l'algorithme sur un cas plus réel en utilisant un cerveau d'agneau.

On dispose de plusieurs acquisitions de la tâche blanche bougeant de différentes manière. En particulier, on retient trois acquisitions faisant bouger la tâche respectivement sur la diagonale du coin supérieur gauche de l'image au coin inférieur droit, sur la diagonale du coin supérieur droit au coin inférieur gauche, et enfin suivant un carré. Des vidéos permettant de visualiser le mouvement sont disponibles sur la plateforme PILoT. Une autre acquisition sans mouvement de la tâche est aussi disponible. On montre un snapshot de la tâche à l'instant initial en Figure 22a.

Le deuxième objet considéré est un cerveau d'agneau pris sous trois angles de vue différents. De la même manière que pour la tâche blanche, on a pour chaque angle de vue une acquisition du cerveau immobile et une autre avec mouvement. On montre également un snapshot de l'objet par point de vue en Figures 22b, 22c et 22d. La pulsation *in-vivo* du cerveau est réalisée à l'aide d'un piston faisant des mouvements de va-et-vien avec une amplitude de 2cm selon le cycle affine décris en Figure 23. Ce cycle est choisi pour modéliser la pression sanguine dans le cerveau et est

1. url PILoT : <https://pilot-warehouse.creatis.insa-lyon.fr/>

ainsi composé d'un pic diastolique, d'un pic systolique puis d'une autre pulsation de plus faible amplitude.

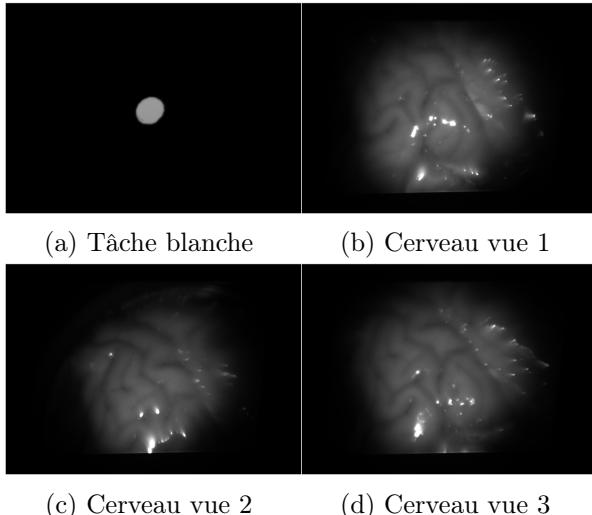


FIGURE 22 – Présentation des objets expérimentaux

Mouvement du piston sur une période

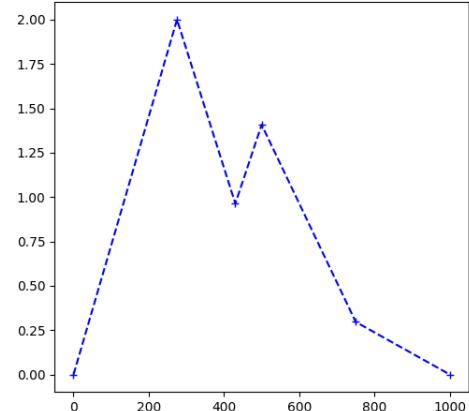


FIGURE 23 – Evolution de l'amplitude (cm) du piston en fonction du temps (ms) sur une période

6.2 Reconstruction statique

Dans un premier temps, on peut s'intéresser à reconstruire les objets quand ils ne bougent pas. Cela permet non seulement de repérer les difficultés liées à la reconstruction sur les données expérimentales mais aussi de jauger le niveau de bruit au niveau du détecteur présent sur les acquisitions.

On constate dès les premières reconstructions que la caméra mono-pixel reconstruit l'image à l'envers. Cela est possible car le détecteur étant unique, il ne prend pas en compte la distribution spatiale des intensités qu'il reçoit. On pivotera les images présentées de 180 degrés.

Bien que cet aspect n'ait pas été considéré jusqu'à présent, il est important de prendre en compte le caractère hyperspectral du problème dans le cas réel. On dispose d'hypercubes de données de taille $2N \times N_\lambda$, où $N = 4096 = 64 \times 64$ est la taille de l'image et $N_\lambda = 2048$ est le nombre de longueurs d'ondes isolées par le spectromètre. Le problème reste le même en pratique et pour une longueur d'onde λ donnée, on a :

$$\mathbf{f}_\lambda = \mathbf{H} \mathbf{m}_\lambda \quad (31)$$

Toutefois, les reconstructions sur une unique longueur d'onde sont soumises à un bruit de Poisson important. On peut alors augmenter la résolution spatiale de l'image en réduisant la dimension spectrale du problème en considérant des tranches du spectre expérimental b_i tels que :

$$\mathbf{f}_{b_i} = \sum_{\lambda \in b_i} \mathbf{f}_\lambda = \mathbf{H} \sum_{\lambda \in b_i} \mathbf{m}_\lambda = \mathbf{H} \mathbf{m}_{b_i} \quad (32)$$

Ce travail étant plus porté sur la compensation de mouvement que sur l'aspect spectral du problème, on considère une seule tranche du spectre expérimental $b_0 = [515, 750]$ nm. On donne les reconstructions statiques des différents objets considérés en Figure 24 sans aucune régularisation.

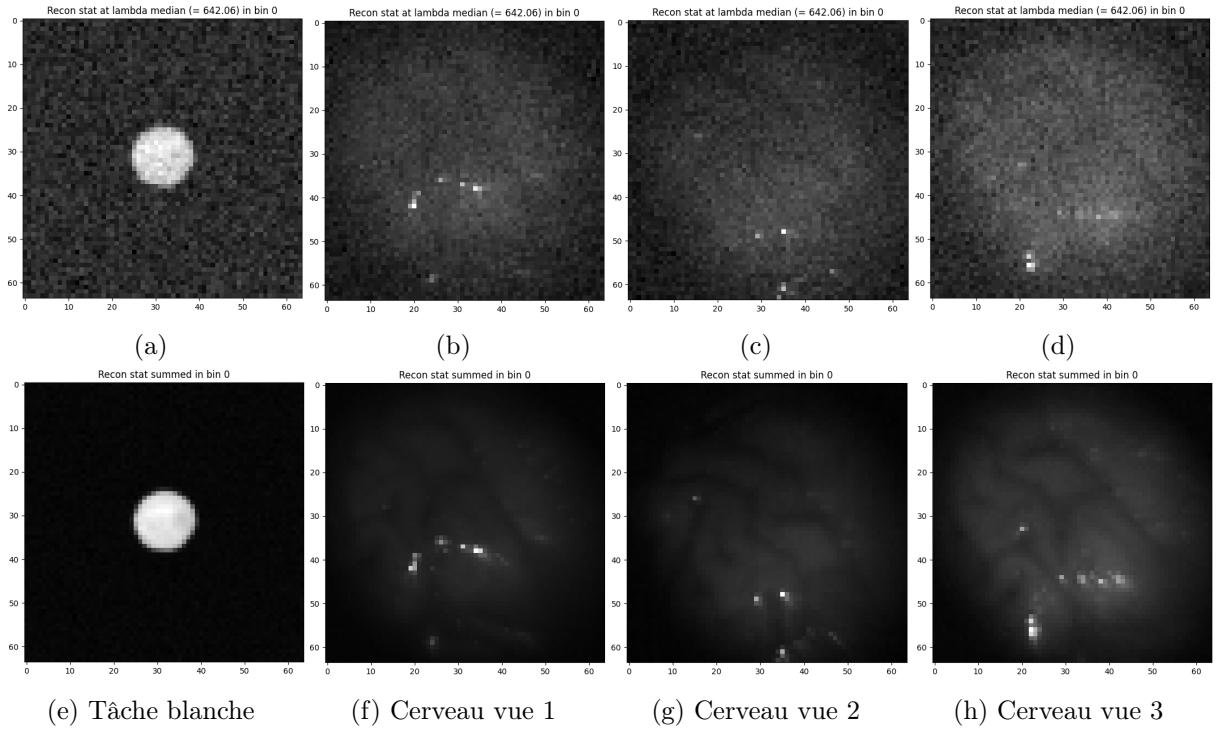


FIGURE 24 – Reconstructions statique de la tâche blanche et du cerveau immobiles. Les images (a)-(d) sont reconstruites sur la longueur d'onde médiane de b_0 . Celles notées (e)-(h) présentent des reconstructions sommées sur l'entièreté de b_0 .

On observe sur la première ligne de la Figure 24 la présence d'un bruit important au niveau du détecteur mono-pixel. Cependant on peut aussi voir sur la deuxième ligne de cette Figure que ce bruit est considérablement réduit en sommant tous les canaux spectraux dans le cas d'un objet immobile. Les reconstructions statiques d'un objet expérimental immobile sont assez satisfaisantes.

6.3 Estimation du mouvement

6.3.1 Calibration de la caméra

La manipulation des données réelles ne permet pas d'estimer directement le mouvement comme en partie 5. En effet, les caméras RGB et mono-pixel (abrégée "SP" dans la suite) ayant des champs de vue différents, il est nécessaire de bien calibrer les caméras pour prendre en compte d'éventuels facteurs d'échelles ou autres transformations linéaires de la scène acquise par les deux dispositifs. On cherche donc une matrice A de taille 3×3 telle que :

$$x^{SP} = Ax^{RGB} \quad (33)$$

où x^{SP} et x^{RGB} sont respectivement les coordonnées homogènes d'un point de l'image dans l'espace de la caméra mono-pixel et dans celui de la caméra RGB.

On propose dans ce travail l'utilisation de la méthode DLT (« Direct Linear Transform ») [36] qui permet de déterminer l'homographie A à partir d'un jeu de points clés entre l'image RGB et celle SP. On détermine les points clés en les plaçant manuellement, cependant on note qu'il est possible de les placer automatiquement en utilisant par exemple les points clés SIFT [37] ou encore ceux de Shi-Tomasi [33].

Pour la tâche blanche, le manque d'attributs distinctifs de l'image permet difficilement d'obtenir une homographie convenable. On utilise donc une acquisition effectuée le même jour sur une image statique d'un chat dont les nombreux détails permettent de bien placer les points clés. On montre le processus de calibration en Figure 25. Après avoir placé les points clés à la main sur l'image RGB et sur la reconstruction mono-pixel en Figures 25a et 25b, on calibre l'image RGB dans l'espace de la caméra mono-pixel en Figure 25c. On peut voir sur la Figure 25d que l'erreur entre l'image RGB et celle mono-pixel est assez faible dans la zone où l'image est effectivement reconstruite

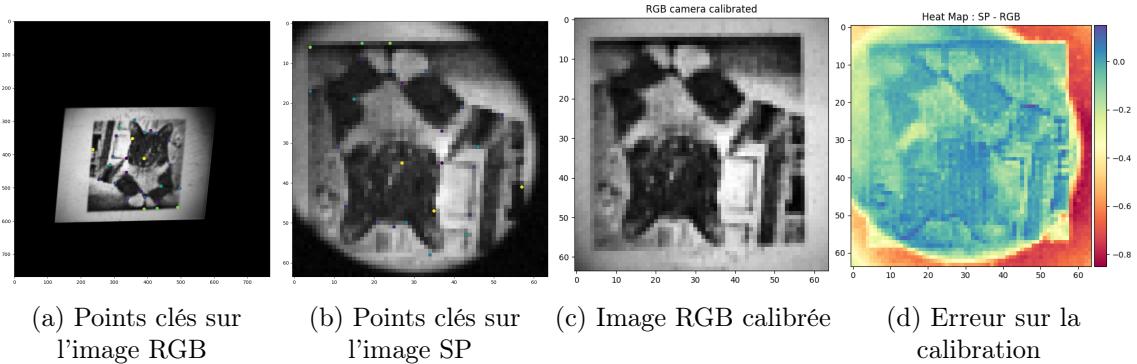


FIGURE 25 – Processus de calibration pour la tâche blanche

L'utilisation de l'homographie apprise permet en principe de calibrer la tâche blanche acquise par la caméra RGB à la résolution de la caméra mono-pixel. En pratique, on observe un décalage de $(1, -8)$ pixels entre la reconstruction statique de la tâche et l'image RGB de la frame initiale de la vidéo. Cette différence ne s'explique que par un léger décalage de la caméra mono-pixel due à une erreur de manipulation entre les acquisitions du chat et celles de la tâche blanche.

Les acquisition sur le cerveau d'agneau ayant été faites 2 mois plus tard, l'homographie apprise sur l'acquisition du chat n'est plus du tout valable. Il faut donc en déterminer une nouvelle à partir des données disponibles, mais cela pose un challenge car les points clés ne sont pas forcément évidents à placer.

Il est maintenant possible de ramener les images de la caméra RGB sur les mêmes champs de vue et résolutions spatiales que la caméra SP. Cependant, procéder ainsi fait perdre énormément d'information au niveau de la résolution spatiale, et bien que ça ne soit pas forcément important pour la tâche blanche, ça le sera pour des objets plus complexes possédant de fins détails.

6.3.2 Estimation du mouvement en haute résolution

Dans cette partie, on utilise la l'homographie $A = (a_{ij})_{(i,j) \in [1,3]^2}$ déterminée par la DLT pour ramener un champ de déformation déterminé sur l'image RGB à l'échelle de l'image SP. En se ramenant à un espace 2D, on peut écrire A comme l'application :

$$\mathcal{A} : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} \frac{a_{11}x_1 + a_{12}x_2 + a_{13}}{a_{31}x_1 + a_{32}x_2 + a_{33}} \\ \frac{a_{21}x_1 + a_{22}x_2 + a_{23}}{a_{31}x_1 + a_{32}x_2 + a_{33}} \end{pmatrix} \quad (34)$$

On note u^{RGB} et u^{SP} les déplacements qui ont lieu respectivement dans les espaces RGB et SP . On propose de trouver la relation entre ces deux déplacements en suivant un point matériel x^{RGB} . Après s'être déformé, ce point matériel se trouve aux coordonnées $\tilde{x}^{RGB} = u^{RGB}(x^{RGB})$. On applique l'homographie sur ce point pour déterminer la position de ce même point dans

l'espace mono-pixel et on trouve $\tilde{x}^{SP} = \mathcal{A}(\tilde{x}^{RGB}) = \mathcal{A}(u^{RGB}(x^{RGB})) = \mathcal{A}(u^{RGB}(\mathcal{A}^{-1}(x^{SP})))$. On en déduit la relation :

$$u^{SP} = \mathcal{A} \circ u^{RGB} \circ \mathcal{A}^{-1} \quad (35)$$

qui permet d'estimer le mouvement directement sur la vidéo haute résolution de la caméra RGB.

On compare un snapshot du champ de déformation ainsi estimé avec ceux estimés sur les vidéos calibrées pour la tâche blanche bougeant sur la deuxième diagonale en Figure 26. On voit que l'allure générale du champ déterminé par flot optique TV- L^1 en haute résolution est la même que pour les autres algorithmes, cependant l'amplitude est plus faible plus on s'éloigne de la position d'origine de la tâche. Par ailleurs, on voit que les flèches autour de la tâche semblent partir dans deux directions symétriques par rapport à l'axe diagonal du mouvement. Tout cela est rendu possible par la nature mal posée du problème d'estimation de mouvement, et il va falloir confirmer que l'estimation est bonne en regardant directement la reconstruction dynamique.

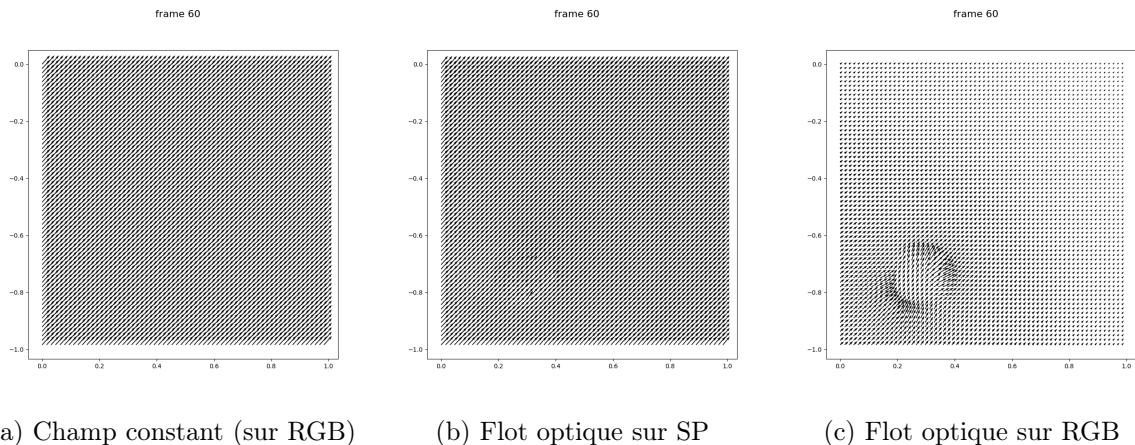


FIGURE 26 – Champs de déformations à un instant donné avec différentes méthodes d'estimation du mouvement

6.4 Reconstruction dynamique

Une fois le mouvement de l'objet expérimental estimé, on peut utiliser la méthode de reconstruction avec compensation de mouvement pour obtenir des résultats. On commence par reconstruire la tâche blanche en Figure 27, le paramètre de régularisation étant choisi constant ($\eta = 5 \cdot 10^{-4}$) pour toutes les reconstructions TV.

Les trois lignes de la Figure 27 présentent les résultats obtenus pour le mouvement de translation sur la première diagonale, sur la seconde diagonale et sur le carré. On montre d'abord sur la première colonne les reconstructions statiques de la tâche blanche, sur lesquelles on peut observer de forts artefacts de mouvement. Ces artefacts sont réduits sur la deuxième colonne avec l'utilisation de la régularisation TV, mais un effet de flou est toujours présent au niveau de la tâche. Sur les trois colonnes qui suivent, on montre les résultats de la compensation de mouvement en utilisant respectivement le champ de déformation constant sur la vidéo haute résolution RGB, l'algorithme de flot optique sur la vidéo calibrée et l'algorithme de flot optique sur la vidéo haute résolution RGB. Dans tous les cas, la reconstruction est bien meilleure que celles sans compensation et on reconnaît la tâche blanche sans problèmes. En regardant de plus près les résultats, les reconstructions utilisant le champ de mouvement constant et celui estimé par flot optique à faible résolution spatiale sont meilleures que celle utilisant le flot optique à haute résolution spatiale. Ceci peut s'expliquer par une mauvaise détermination de la DLT lors

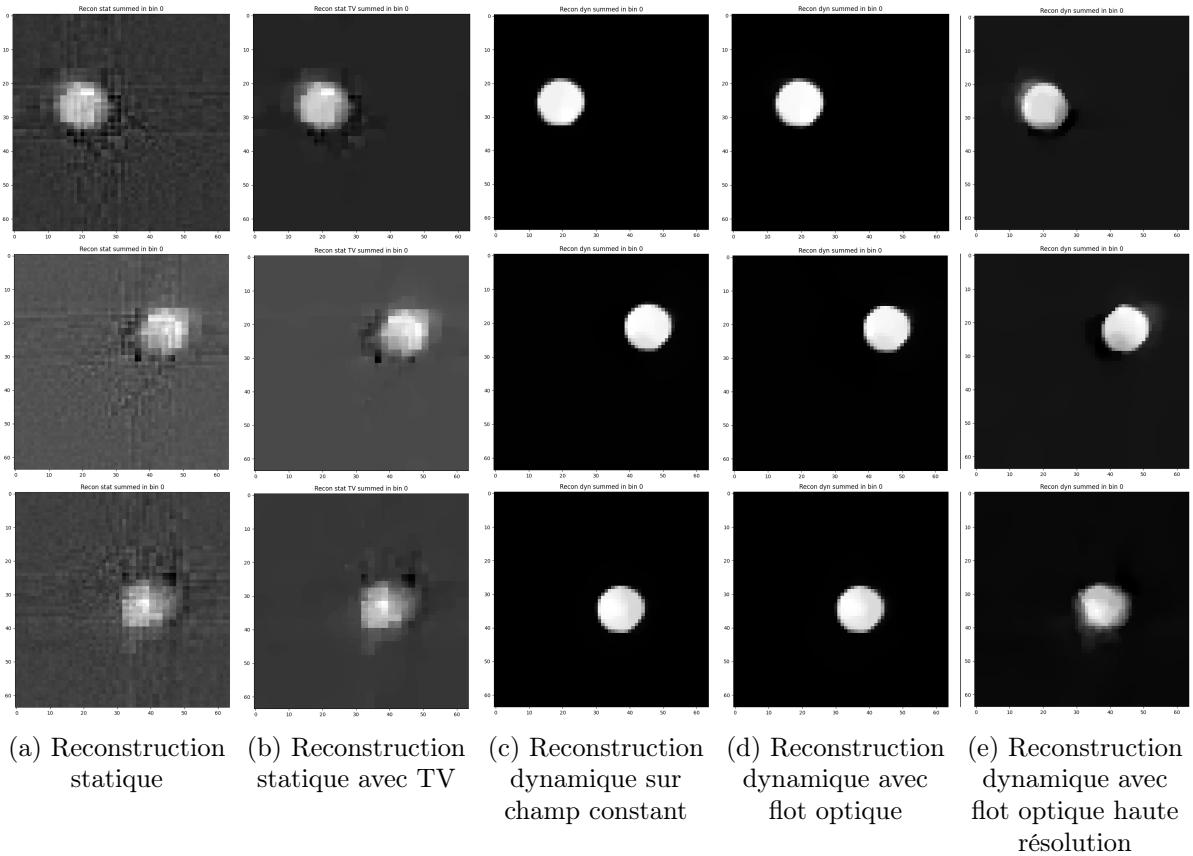


FIGURE 27 – Reconstructions dynamiques sur la tâche blanche

du placement des points clés à la main, où un grand intervalle de points est possible sur l'image RGB pour un point fixé sur l'image mono-pixel. L'intérêt de cette méthode est donc discutable pour un objet aussi grossier que la tâche blanche, cependant il pourrait donner de meilleurs résultats que les autres méthodes dans le cas d'un objet avec de fins détails. Pour une évaluation un peu plus quantitative des résultats, on compare les scores PSNR des différentes méthodes d'estimation du mouvement sur la Table 1. N'ayant pas accès à de vérité terrain dans le cas réel, on considère que l'image de référence est la première frame du flux vidéo RGB.

Mouvement	Champ constant	Flot optique	Flot optique haute résolution
Diagonale 1	37	39	27
Diagonale 2	40	40	27
Carré	39	39	29

TABLE 1 – Scores PSNR (dB) expérimentaux pour différentes méthodes d'estimation du mouvement sur trois modèles de mouvement pour la tâche blanche

La Table 1 montre que considérer le champ constant est une bonne approximation du modèle de mouvement ici puisque les scores PSNR obtenus sont très proches de la méthode standarde de flot optique. En revanche, une diminution du PSNR est observée quand on estime le mouvement sur les vidéos en haute résolution spatiale, sûrement dû à une mauvaise DLT. On peut également vérifier la compensation de mouvement visuellement sur la Figure 28 où l'on affiche la différence entre la reconstruction dynamique et la frame initiale.

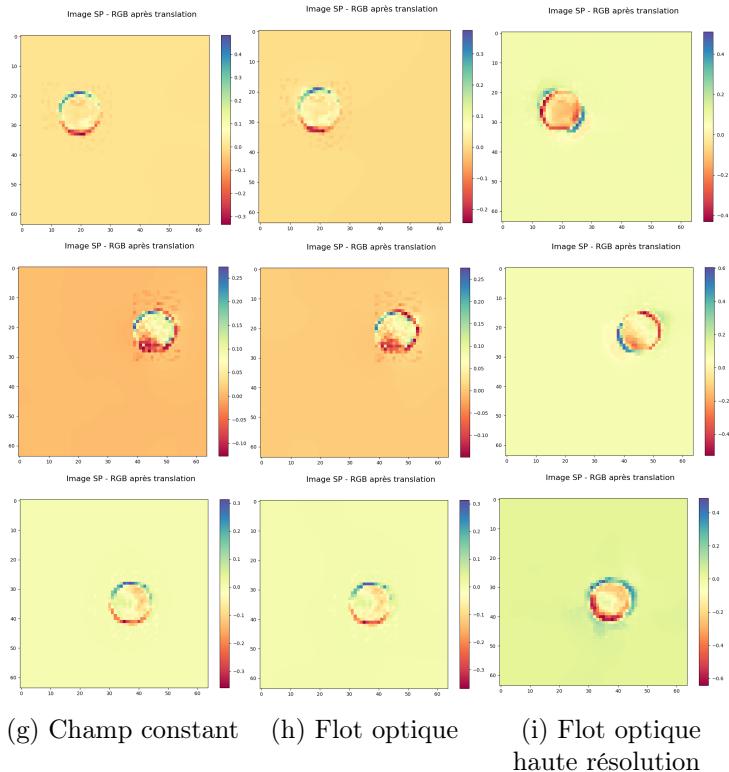


FIGURE 28 – Vérification de la compensation de mouvement

Globalement, on constate à l'aide de la Figure 28 que la méthode de compensation de mouvement marche plutôt bien. Un léger décalage dans le sens du mouvement est cependant systématiquement observé. Ceci s'explique par le mode d'acquisition séquentiel qui crée un décalage temporel entre la frame initiale et les premières acquisitions de la caméra mono-pixel.

Après avoir validé expérimentalement la méthode de compensation de mouvement sur un objet grossier soumis à un déplacement simple, on voudrait confronter la méthode de compensation sur le cerveau d'agneau en mouvement. Une DLT de moindre qualité à d'abord été déterminée de la même façon que précédemment. L'étude a été menée sur les trois acquisitions disponibles mais les résultats étant très similaires, on ne présente ici que ceux pris avec le point de vue 3 du cerveau. La Figure 29 est obtenue avec un coefficient de régularisation $\eta = 1.75 \cdot 10^{-3}$.

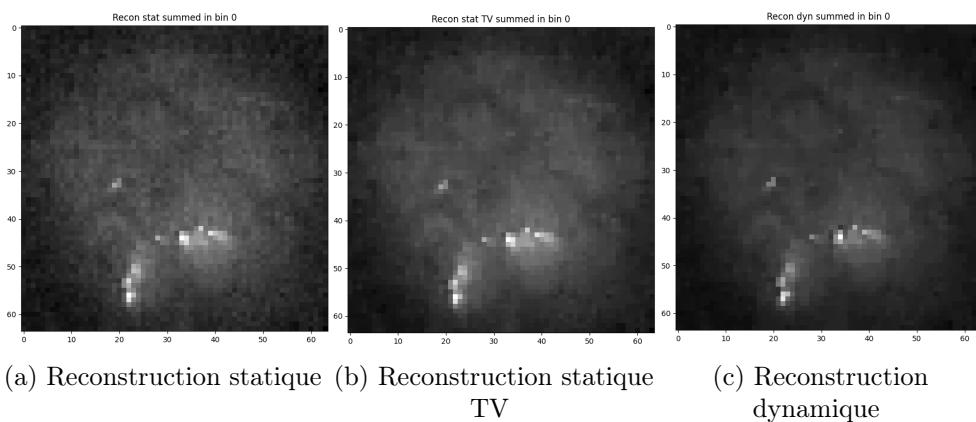


FIGURE 29 – Reconstruction dynamique sur le cerveau d'agneau

Les résultats montrés en Figure 29 sont malheureusement peu satisfaisants. Le fort bruit dû au mouvement visible sur la Figure 29a est effectivement lissé par l'algorithme de régularisation TV en Figure 29b. Cependant, bien que l'on puisse observer une diminution du bruit et une légère amélioration de la reconstruction en Figure 29c, la méthode de compensation ne permet pas de retrouver l'image nette de la Figure 24h dans le cas du cerveau. Il y a plusieurs causes possibles qui expliquent cet échec. Premièrement, il est probable qu'il y ait eu une erreur au niveau de la détermination de la DLT par manque de points clés sur le cerveau. Deuxièmement, la qualité des vidéos RGB (mise au point, reflets sur le cerveau...) n'étaient pas idéales, ce qui a posé des difficultés lors de l'estimation du mouvement. Ces deux pistes sont les causes d'erreur les plus probables, toutefois il est aussi possible que l'augmentation des temps d'intégration des canaux hyperspectral et RGB par rapport à l'expérience de la tâche blanche ait aussi quelque chose à voir avec la différence observée.

7 Conclusion

La caméra mono-pixel est un dispositif d'acquisition particulier qui tient son nom de son unique détecteur. La formulation d'un problème inverse basé sur la transformée discrète d'Hadamard en section 2 a rendu possible la reconstruction d'un objet qui ne bouge pas en section 3 en introduisant la technique de régularisation TV principalement utilisée dans ce travail. Cependant, à cause de son mode d'acquisition séquentiel d'une dizaine de secondes, de forts artefacts sont visibles lors de l'imagerie d'un objet en mouvement. Dans le but de développer cette technologie pour qu'elle puisse être utilisée *in-vivo* pour assister le chirurgien au bloc opératoire, la section 4 étudie le cas précis où l'objet à imager bouge selon un modèle de mouvement connu. Après avoir proposé deux approches pour compenser le mouvement, des simulations numériques sont menées sous différentes conditions pour mettre à l'épreuve la méthode. Pour généraliser le travail mené jusqu'ici, on étudie en section 5 des algorithmes permettant d'estimer le mouvement d'un objet à partir d'un flux vidéo. Ce travail est effectué sur des données simulées pour avoir de bonnes façons de mesurer les résultats obtenus. Enfin, la méthode est confrontée à des données expérimentales en section 6. Des acquisitions d'une tâches blanche sont d'abord utilisées et permettent de valider le travail effectué, puis un essai de compensation du mouvement sur un cerveau d'agneau est effectué.

Les résultats obtenus pour le cerveau d'agneau n'étant pas très satisfaisant, il faudra essayer d'améliorer ces résultats à l'avenir en cherchant à améliorer la technique d'estimation du mouvement sur le flux vidéo haute résolution. Pour cela, il faudra déterminer des DLT plus précises et éventuellement faire de nouvelles acquisitions. Si les qualités spatiales ne sont toujours pas suffisantes, il sera intéressant de se pencher plus en détail sur les techniques de pan-sharpening mentionnées en introduction. Une fois ce travail fait, on pourra utiliser des techniques d'apprentissage profond lors de la reconstruction pour améliorer le temps d'exécution du programme. La méthode étant naturellement basée sur des techniques itératives, il faudra s'intéresser à des techniques d'« unrolling ».

Ce stage à CREATIS a pu confirmer mon intérêt pour les sciences appliquées au domaine médical. J'aimerais à ce propos remercier mes trois tuteurs de stage Elie BRETIN, Nicolas DUCROS et Michaël SDIKA pour m'avoir introduit au domaine du traitement du signal et pour m'avoir patiemment accompagné tout au long de mon stage. J'ai réellement apprécié les moments passés ensemble, que ce soit en réunion ou non, et je me réjouis à l'idée de pouvoir continuer à travailler avec vous prochainement. Je voudrais également remercier Laurent MAHIEU-WILLIAME pour avoir été très disponible sur les petites questions que j'ai pu me poser sur la partie expérimentale, ainsi que Jérémy COHEN pour les conseils apportés au cours du stage.

8 Références

- [1] Marco F Duarte, Mark A Davenport, Dharmpal Takhar, Jason N Laska, Ting Sun, Kevin F Kelly, and Richard G Baraniuk. Single-pixel imaging via compressive sampling. *IEEE signal processing magazine*, 25(2) :83–91, 2008.
- [2] Nicolas Ducros. Une introduction à l'imagerie computationnelle monodétecteur. 2022.
- [3] Matthew P Edgar, Graham M Gibson, and Miles J Padgett. Principles and prospects for single-pixel imaging. *Nature photonics*, 13(1) :13–20, 2019.
- [4] Florian Rousset, Nicolas Ducros, Françoise Peyrin, Gianluca Valentini, Cosimo D’Andrea, and Andrea Farina. Time-resolved multispectral imaging based on an adaptive single-pixel camera. *Opt. Express*, 26(8) :10550–10558, Apr 2018.
- [5] Vincent Studer, Jérôme Bobin, Makhla Chahid, Hamed Shams Mousavi, Emmanuel Candes, and Maxime Dahan. Compressive fluorescence microscopy for biological and hyperspectral imaging. *Proceedings of the National Academy of Sciences*, 109(26) :E1679–E1687, 2012.
- [6] Enagnon Aguénounon, Foudil Dadouche, Wilfried Uhring, Nicolas Ducros, and Sylvain Gioux. Single snapshot imaging of optical properties using a single-pixel camera : a simulation study. *Journal of Biomedical Optics*, 24(7) :071612–071612, 2019.
- [7] Tadanobu Nagaya, Yu A Nakamura, Peter L Choyke, and Hisataka Kobayashi. Fluorescence-guided surgery. *Frontiers in oncology*, 7 :314, 2017.
- [8] Jaime J Bravo, Jonathan D Olson, Scott C Davis, David W Roberts, Keith D Paulsen, and Stephen C Kanick. Hyperspectral data processing improves ppix contrast during fluorescence guided surgery of human brain tumors. *Scientific reports*, 7(1) :9455, 2017.
- [9] David Atkinson, Derek LG Hill, Peter NR Stoyle, Paul E Summers, Stuart Clare, Richard Bowtell, and Stephen F Keevil. Automatic compensation of motion artifacts in mri. *Magnetic Resonance in Medicine : An Official Journal of the International Society for Magnetic Resonance in Medicine*, 41(1) :163–170, 1999.
- [10] PG Batchelor, D Atkinson, P Irarrazaval, DLG Hill, J Hajnal, and D Larkman. Matrix description of general motion correction applied to multishot images. *Magnetic Resonance in Medicine : An Official Journal of the International Society for Magnetic Resonance in Medicine*, 54(5) :1273–1280, 2005.
- [11] Aurélien Bustin, Anne Menini, Martin A Janich, Darius Burschka, Jacques Felblinger, Anja CS Brau, and Freddy Odille. Motion estimated-compensated reconstruction with preserved-features in free-breathing cardiac mri. In *Reconstruction, Segmentation, and Analysis of Medical Images : First International Workshops, RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers 1*, pages 70–80. Springer, 2017.
- [12] Zijun Guo, Wenwen Meng, Dongfeng Shi, Linbin Zha, Wei Yang, Jian Huang, Yafeng Chen, and Yingjian Wang. Fast localization and single-pixel imaging of the moving object using time-division multiplexing. *arXiv preprint arXiv* :2208.07371, 2022.
- [13] Shuming Jiao, Mingjie Sun, Yang Gao, Ting Lei, Zhenwei Xie, and Xiaocong Yuan. Motion estimation and quality enhancement for a single image in dynamic single-pixel imaging. *Optics express*, 27(9) :12841–12854, 2019.
- [14] Berthold KP Horn and Brian G Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3) :185–203, 1981.
- [15] Michaël Sdika, Laure Alston, David Rousseau, Jacques Guyotat, Laurent Mahieu-Williame, and Bruno Montcel. Repetitive motion compensation for real time intraoperative video processing. *Medical Image Analysis*, 53 :1–10, 2019.

- [16] Antonio Lorente Mur, Pierre Leclerc, Fran oise Peyrin, and Nicolas Ducros. Single-pixel image reconstruction from experimental data using neural networks. *Optics Express*, 29(11) :17097–17110, 2021.
- [17] Antonio Lorente Mur, Paul Bataille, Fran oise Peyrin, and Nicolas Ducros. Deep expectation-maximization for image reconstruction from under-sampled poisson data. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1535–1539. IEEE, 2021.
- [18] Laetitia Loncan, Luis B De Almeida, Jos  M Bioucas-Dias, Xavier Briottet, Jocelyn Chanussot, Nicolas Dobigeon, Sophie Fabre, Wenzhi Liao, Giorgio A Licciardi, Miguel Simoes, et al. Hyperspectral pansharpening : A review. *IEEE Geoscience and remote sensing magazine*, 3(3) :27–46, 2015.
- [19] Te-Ming Tu, Shun-Chi Su, Hsuen-Chyun Shyu, and Ping S Huang. A new look at ihs-like image fusion methods. *Information fusion*, 2(3) :177–186, 2001.
- [20] Wen Dou, Yunhao Chen, Xiaobing Li, and Daniel Z Sui. A general framework for component substitution image fusion : An implementation using the fast image fusion method. *Computers & Geosciences*, 33(2) :219–228, 2007.
- [21] Guilherme Beneti Martins, Laurent Mahieu-Willame, Thomas Baudier, and Nicolas Ducros. Openspyrit : an ecosystem for open single-pixel hyperspectral imaging. *Optics Express*, 31(10) :15599–15614, 2023.
- [22] Henri Ma tre. *Du photon au pixel : L'appareil photographique num rique 2e  dition revue et augment e *. ISTE Group, 2016.
- [23] Zibang Zhang, Xueying Wang, Guoan Zheng, and Jingang Zhong. Hadamard single-pixel imaging versus fourier single-pixel imaging. *Optics Express*, 25(16) :19619–19639, 2017.
- [24] Graham M Gibson, Steven D Johnson, and Miles J Padgett. Single-pixel imaging 12 years on : a review. *Optics express*, 28(19) :28190–28208, 2020.
- [25] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D : nonlinear phenomena*, 60(1-4) :259–268, 1992.
- [26] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20 :89–97, 2004.
- [27] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM journal on imaging sciences*, 2(1) :183–202, 2009.
- [28] Andreas Langer. Automated parameter selection for total variation minimization in image restoration. *Journal of Mathematical Imaging and Vision*, 57 :239–268, 2017.
- [29] Triet Le, Rick Chartrand, and Thomas J Asaki. A variational approach to reconstructing images corrupted by poisson noise. *Journal of mathematical imaging and vision*, 27(3) :257–263, 2007.
- [30] Jacques Hadamard. Sur les probl mes aux d riv es partielles et leur signification physique. *Princeton university bulletin*, pages 49–52, 1902.
- [31] Per Christian Hansen. *Rank-deficient and discrete ill-posed problems : numerical aspects of linear inversion*. SIAM, 1998.
- [32] Bruce D Lucas and Takeo 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.
- [33] Jianbo Shi et al. Good features to track. In *1994 Proceedings of IEEE conference on computer vision and pattern recognition*, pages 593–600. IEEE, 1994.

- [34] Gunnar Farnebäck. Two-frame motion estimation based on polynomial expansion. In *Image Analysis : 13th Scandinavian Conference, SCIA 2003 Halmstad, Sweden, June 29–July 2, 2003 Proceedings 13*, pages 363–370. Springer, 2003.
- [35] Christopher Zach, Thomas Pock, and Horst Bischof. A duality based approach for realtime tv-l 1 optical flow. In *Pattern Recognition : 29th DAGM Symposium, Heidelberg, Germany, September 12-14, 2007. Proceedings 29*, pages 214–223. Springer, 2007.
- [36] Yousset I Abdel-Aziz, Hauck Michael Karara, and Michael Hauck. Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry. *Photogrammetric engineering & remote sensing*, 81(2) :103–107, 2015.
- [37] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60 :91–110, 2004.

9 Annexes

9.1 Algorithme de régularisation TV en Python

```
1 import numpy as np
2 from skimage.restoration import denoise_tv_chambolle
3
4 def regularization(A, b, eta, n_iter=10):
5     M, N = A.shape
6     n = int(math.sqrt(N))
7
8     f = np.zeros(N)
9     tau = 0.5 * 1. / N
10
11    ## ISTA
12    for k in range(n_iter):
13        f = f - tau * np.dot(A.transpose(), np.dot(A, f) - b)
14        f_square = f.reshape((n, n))
15        f_square = denoise_tv_chambolle(f_square, weight=tau / eta, eps=2e-5,
16        max_num_iter=2000)
17        f = f_square.reshape((N,))
18
19    return f
```

Listing 1 – Régularisation TV en Python

Un répertoire demo est disponible sur GitHub¹ pour le lecteur qui voudrait avoir un point de vue plus pratique de la méthode employée.

9.2 Effet du sous-échantillonage sur le temps de calcul

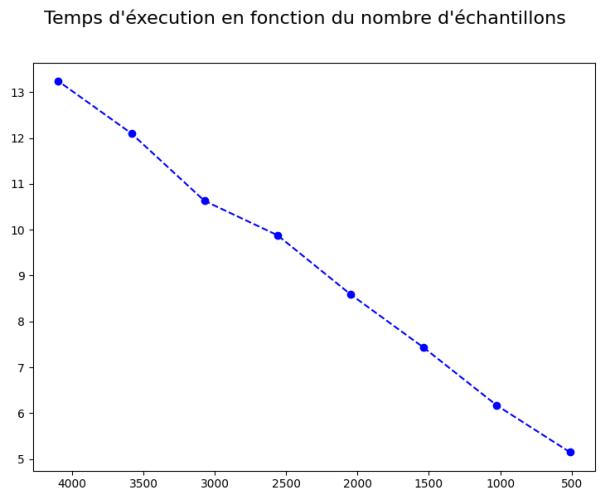


FIGURE 30 – Évolution du temps d'exécution (s) en fonction du nombre d'échantillons

9.3 Erreurs sur l'estimation du mouvement

Le meilleur des cas pour l'estimation du mouvement consiste à considérer une estimation parfaite du mouvement. Ainsi, $\forall x \in X, e_{x_i}^f(x) = 0$ et on voit facilement que $e_{L^2}^f = 0$. Le pire des cas pour l'estimation du mouvement serait de se tromper de la taille de l'image, c'est-à-dire

1. url : https://github.com/Tommte/sp_moco_demo/tree/main

qu'un pixel donné $x \in X$ qui devrait se trouver sur le bord droit de l'image après déplacement se trouve en réalité sur le bord gauche de l'image. Dans ce cas, on a pour $i \in \{1, 2\}$:

$$\begin{aligned} e_{x_i}^f(x) &= 1 \\ \text{donc } e^f(x) &= \sqrt{2} \\ \text{donc } e_{L^2}^f &= \frac{1}{\sqrt{N(N+1)}} \sqrt{\sum_{j=0}^{N-1} 2} \\ \text{donc } e_{L^2}^f &= 1 \end{aligned}$$

Les résultats sur l'erreur L^∞ dans le meilleurs comme dans le pire des cas sont triviaux.

9.4 Interpolation bilinéaire sur la tâche blanche

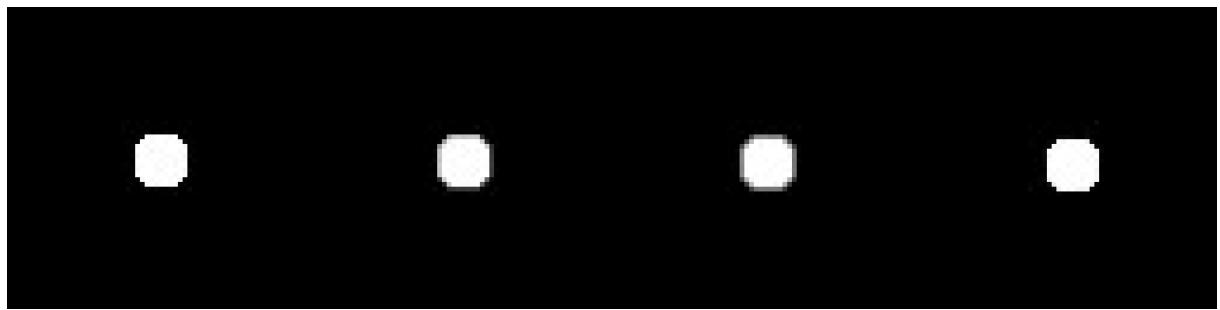


FIGURE 31 – Evolution de la tâche blanche sur différentes soumise à un mouvement de translation diagonal

Note : Les petits changements d'intensités à l'intérieur et à l'extérieur de la tâche blanche qu'on peut éventuellement apercevoir sur la Figure 31 sont des artefacts de compression.

10 Résumé en anglais

In previous team projects, a high-spectral-resolution imager that acquires a $64 \times 64 \times 2048$ hypercube in approximately 10 seconds was developed. This might be sufficient for imaging a motionless object, but blur artifacts are created when applying this method to in-vivo surgery where the patient's organs may be moving at a faster rate (e.g., 1 cardiac cycle per second) than the acquisition process. In order to perform the acquisition, a single-pixel camera coupled to a monochrome/RGB camera were used. While the former is convenient for fast hyperspectral reconstruction, the latter allows for an estimation of the scene motion from a video stream. The purpose of this internship is to propose strategies for a motion-compensated reconstruction algorithm, assuming that the motion of the imaged object is known. An Hadamard transform was used to address the inverse problem that arises from the single-pixel camera. Simulations incorporating different levels of Poisson noise on the measurements and noise on the parameters of the motion model were conducted to validate the proposed strategies. A total variation regularization method was employed to deal with these noise. Lastly, the proposed strategies were applied on real data acquired by fellow team members.