

CORRECTION DU FLOU DE MOUVEMENT SUR DES IMAGES PRISES DE NUIT DEPUIS UN VÉHICULE DE NUMÉRISATION TERRESTRE

Vincent DAVAL⁽¹⁾, Lâman LELÉGARD⁽²⁾, Mathieu BRÉDIF⁽²⁾

⁽¹⁾ Ministère de la Défense, 14 rue Saint-Dominique, 75007 Paris, France

⁽²⁾ Univ. Paris-Est, LASTIG MATIS, IGN, ENSG, F-94160 Saint-Mande, France
E-mail : {laman.lelegard, mathieu.bredif}@ign.fr

Résumé

Ce travail marque une première étape dans la définition d'une méthode de correction du flou de mouvement observé dans les clichés pris avec un long temps d'exposition par un véhicule de cartographie mobile. Dans l'approche proposée, nous prenons en considération à la fois les données inertielles provenant d'accéléromètres et de gyroscopes et les données de variation de la profondeur de la scène fournies par des mesures Lidar ou un modèle 3D. Notre algorithme utilise toutes les données utiles afin de déterminer au mieux la fonction d'étalement du point en chaque pixel. Nous proposons également un premier essai de correction du flou en utilisant les noyaux de flou non uniforme et spatialement variant que nous avons obtenus en suivant une approche de reconstruction spatiale. Notre méthode est actuellement validée sur des prises de vue floues non bruitées obtenues par images de synthèse qui reproduisent le mouvement réel du véhicule. Nous précisons enfin comment il est envisagé d'obtenir une correction de l'image complète et d'améliorer encore ces premiers travaux.

Mots-clés : Flou de bougé, Données inertielles, Flou non uniforme et spatialement variant, Décomposition en valeurs singulières, Fonction d'étalement du point, Transformée de Fourier

Abstract

This current work investigates the possibility of acquiring images with long exposure time, by defining a method of motion blur correction from photographs taken with the specific imaging equipment of a mobile mapping system. In the proposed approach, we take into consideration both inertial data provided by accelerometers and gyroscopes and spatial data coming from Lidar cloud or 3D models and bringing information about the variations of the scene depth. Our algorithm exploits all useful data provided by the mobile mapping system to compute in the most accurate way the point spread function for each pixel. We propose also a first attempt of blur removing using our non-uniform and spatially-varying blur kernel in each pixel and spatial considerations. Our method is currently tested on blurred and non-noisy images generated by synthetic images computing the real vehicle movement. We eventually discuss how to get a whole image corrected from motion blur and how to improve yet our first results.

Keywords: Motion blur, Inertial data, Non-uniform/spatially-varying blur, SVD, PSF, Fourier Transform

1. Introduction

La phase d'acquisition d'images reste une des premières étapes dans la production de données cartographiques. La cartographie traditionnelle repose essentiellement sur l'imagerie aérienne ou satellitaire. Néanmoins, de nouveaux produits géographiques tels que la cartographie à très grande échelle, la navigation immersive ou encore la mise à jour en continu font appel à de nouveaux systèmes d'acquisition tel que les véhicules terrestres de cartographie mobile ou les drones. Dans tous les cas, l'acquisition se fait depuis un objet en mouvement ce qui introduit un flou de bougé dans les images. L'extension de ce flou dépend du temps d'exposition et de la vitesse de la caméra. Dans le cas des prises de vue terrestres diurnes, les temps

d'exposition sont suffisamment courts pour produire des flous de bougé largement inférieur au pixel. Ce flou est invisible sur les images qui sont alors considérées comme nettes. Au contraire, les prises de vue par faible luminosité nécessitent des temps d'exposition plus longs qui sont à l'origine de flou de bougé s'étendant sur plusieurs pixels : ces images sont qualifiées de floues.

Nous nous intéressons ici plus particulièrement au cas des images acquises en faible luminosité, depuis un véhicule terrestre de cartographie mobile, bien que la problématique puisse s'étendre à d'autres configurations d'acquisition.

1.1. Contexte et données de l'étude

Dans le cadre de la montée des véhicules terrestres de cartographie mobile, l'IGN a développé son propre

système mobile de numérisation terrestre [PAPARODITIS et al., 2012] destiné à fournir des données pour la navigation immersive, nommé STEREOPOLIS (Figure 1), et qui produit à la fois de l'imagerie optique et laser obtenue par une tête panoramique formée de dix caméras Pike, deux jeux de caméras directionnelles stéréoscopiques, un scanner Lidar Rotatif Velodyne et de deux lasers à balayage RIEGL. L'ensemble est couplé à un système de navigation, composé d'un GPS et d'une centrale inertielle, assurant le géoréférencement des données.

En vue de proposer un caractère pluri-temporel à la navigation, des acquisitions à différentes heures de la journée ont été envisagées. Or les conditions de prise de vue actuelles des images de nuit par ce système ne permettent pas d'obtenir des clichés nets. Ceux-ci sont systématiquement flous et inutilisables en production. Plusieurs solutions, ainsi que leurs inconvénients respectifs, ont été envisagées pour pallier ce problème :

- Utiliser un flash puissant pour éclairer les façades aurait pu être une source de nuisance non négligeable pour les riverains. De plus, la portée réduite ainsi que les reflets du flash rendent cette solution inenvisageable.
- Réduire le temps d'exposition des caméras en augmentant le gain aurait apporté un niveau de bruit non acceptable.
- Ralentir la vitesse du véhicule imageur était susceptible de gêner la circulation et d'induire des coûts opérationnels dus à une durée d'acquisition plus longue qu'une mission diurne.

Ces différentes approches n'ayant pas été retenues, un traitement correctif du flou a posteriori au niveau des données image s'impose. Une première idée de l'extension du flou de bougé peut être obtenue à partir d'un cas de figure général simplifié. Considérons le véhicule en translation pure à une vitesse de 50km/h et intéressons-nous aux caméras les plus proches de la scène, à savoir les caméras latérales (correspondant aux numéros 31 et 34 sur la Figure 1) dont la focale est de l'ordre de 1400 pixels. Chaque image, prise avec un temps d'exposition de 10 millisecondes, est alors affectée d'un flou horizontal dépendant de la profondeur :

Profondeur du sujet	Extension du flou
5 m	39 pixels
10 m	19,5 pixels
20 m	9,73 pixels

Le flou de bougé est d'autant plus conséquent que les objets dans la scène sont proches de la caméra. On verra par la suite que ce phénomène a un fort impact lors de la correction du flou dans ces images (Figure 5).

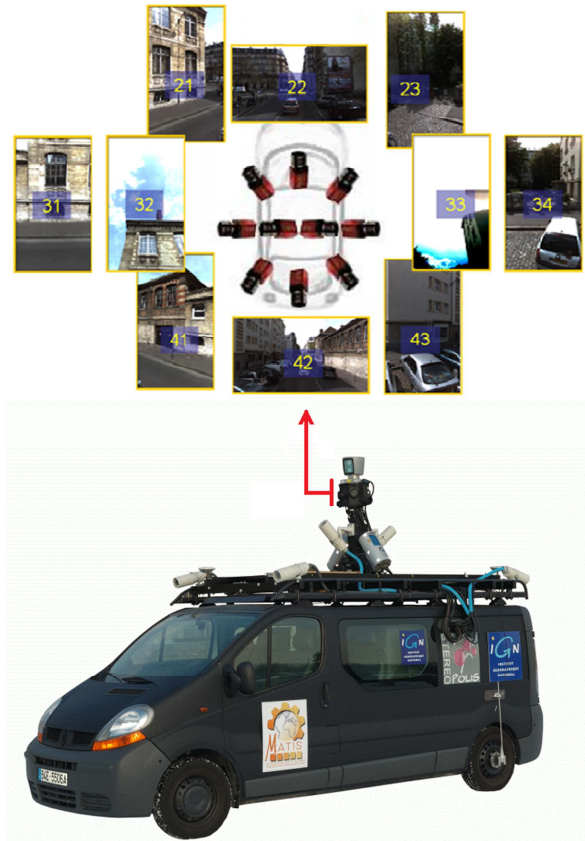


Figure 1. Véhicule STEREOPOLIS et configuration de sa tête panoramique

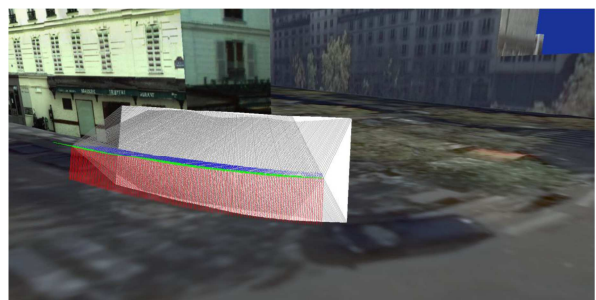


Figure 2. Représentation des 300 positions de caméra entrant dans la simulation d'une image floue

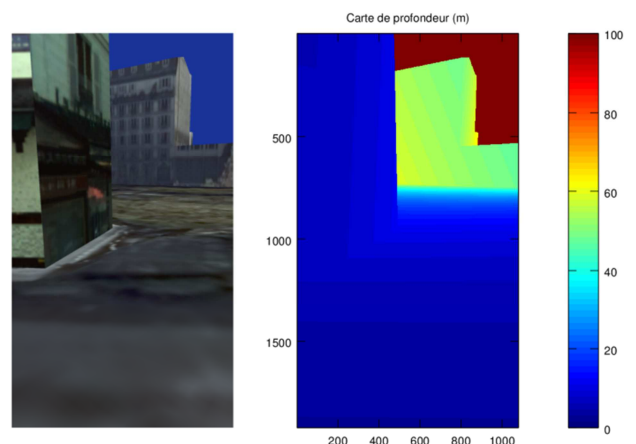


Figure 3. Carte de profondeur associée à une position de la caméra

1.1.1 Données réelles

Le jeu d'images de nuit floues sur Saint Mandé avec leurs poses estimées associées est un jeu de données réelles. Il n'a pas été possible de disposer de données 3D associées pour avoir une notion de profondeur de la scène même approximative. Elles seront utilisées pour l'étude du bruit et pour certains essais de reconstruction par des méthodes traditionnelles.

1.1.2. Données de synthèse

Ces images, texturées, ont été générées à partir de fichiers d'orientation de missions réelles. Ces images permettent de reproduire avec précision un flou dont on connaît et maîtrise précisément les paramètres. Elles sont appropriées à l'analyse approfondie du problème grâce aux données de profondeur de scène associées et à l'absence de bruit.

Carte de profondeur

À chaque image de synthèse est associée une carte de profondeur (Figure 3) calculée à partir du même modèle 3D (Figure 2) qui a servi pour les images de synthèse. Elle servira à calculer l'extension du flou local pour chaque pixel de l'image, en restant totalement cohérente avec l'image traitée. On se place ainsi dans un cas théorique idéal.

1.2. Analyse du problème

Une image issue d'un appareil photographique est obtenue par intégration de l'ensemble des photons lumineux qui sont venus rencontrer le capteur durant le temps d'exposition. Une image floue est une image dont la valeur radiométrique de chaque pixel a été altérée localement par les valeurs des pixels au voisinage. La taille du voisinage est d'autant plus importante que le flou est important. Le nouveau pixel est ainsi une moyenne pondérée des pixels alentours. Cela provoque un lissage des contours se traduisant dans le cliché par une perte de contraste. Une composante de bruit de capteur apparaît également pour chaque pixel sous la forme d'une composante additive. Il est généralement possible de modéliser ce phénomène sous la forme d'une équation simple :

$$pixel_i^{flou} = \sum_{j \in V_i} a_{i,j} \times pixel_j^{nette} + bruit_i, \quad (1)$$

où $a_{i,j}$ est la contribution du $pixel_j^{nette}$ d'un voisinage V_i de l'image nette théorique et $bruit_i$ le bruit du capteur au niveau du photosite i .

L'équation (1) peut se formaliser sous la forme d'un produit matriciel (2) ou bien, dans le cas où les caractéristiques du flou sont constantes sur toute l'image, par un produit de convolution (3). Elle se généralise ainsi à l'ensemble de l'image :

$$I^{floue} = A \cdot I^{nette} + B, \quad (2)$$

$$image^{floue} = a * image^{nette} + bruit. \quad (3)$$

L'élément a est appelé le noyau de flou ou fonction d'étalement du point, qui sera désignée par la suite par le terme PSF, de l'anglais *Point Spread Function*. D'un point de vue physique, cela correspond à l'image d'une source ponctuelle ayant subi ce flou (Figure 4).

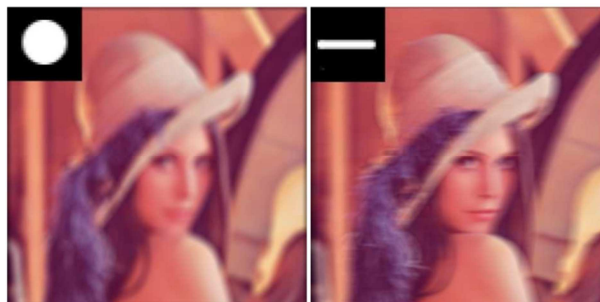


Figure 4. Exemples de PSF
À gauche : flou de mise au point (disque)
À droite : flou de filé horizontal (segment)



Figure 5. Déconvolutions de Wiener pour un modèle de filé de n pixels

Dans une représentation fréquentielle, obtenue par une transformation de Fourier, le produit de convolution devient une simple multiplication, l'équation (3) ainsi s'écrit sous la forme suivante :

$$TF(image^{floue}) = TF(a).TF(image^{nette}) + TF(bruit), \quad (4)$$

où TF désigne la transformée Fourier du signal image. Il est alors envisageable de retrouver le signal net $image^{nette}$ en divisant (4) par le terme $TF(a)$. On parle alors de déconvolution du signal.

Or cette méthode naïve ne donne généralement pas de résultats exploitables : $TF(a)$ pouvant généralement présenter des éléments nuls (correspondant à une extinction de contraste de l'image floue dans le domaine temporel), la division du terme $TF(bruit)$ par $TF(a)$ rend le résultat illisible (explosion du bruit due à une division par zéro).

L'approche proposée par [WIENER, 1949] permet de diminuer ce phénomène en atténuant les hautes fréquences lors de la déconvolution.

Le filtre de Wiener se caractérise comme étant le filtre qui minimise l'erreur quadratique moyenne entre la solution estimée et la solution vraie. Il prend la forme suivante :

$$W(a) = \frac{\overline{TF(a)}}{|TF(a)|^2 + 1/SNR(image^{floue})}, \quad (5)$$

où $SNR(image^{floue})$ est le rapport signal sur bruit de l'image. Lorsque le bruit est négligeable, on retrouve l'équation du filtre inverse. Cette méthode nécessite de disposer d'estimations sur le bruit et sur le flou. En pratique, et en l'absence d'information sur le niveau du bruit dans l'image, le terme de régularisation $1/SNR(image^{floue})$ est modélisé par une constante positive proche de zéro.

1.2.1 Une première application aux données réelles

Le filtre de Wiener donne des résultats intéressants dans le cas d'un flou de filé, dont la PSF est modélisée par un segment de droite. Une inscription, alors difficilement déchiffrable à cause du flou de filé, devient lisible après une déconvolution de Wiener pour une PSF correctement estimée (Figure 5). Dans cet exemple, on remarquera que l'extension du flou n'est pas le même pour l'inscription au premier plan (au niveau du store) et celle à l'arrière-plan (située au niveau de la façade). Cela s'interprète par le fait que les objets au premier plan parcourent une distance plus importante dans l'image que ceux à l'arrière-plan.

1.2.2 Configuration des données de synthèse

L'image floue simulée (Figure 2) illustre typiquement le cas particulier d'une photo acquise dans un virage. En appariant manuellement (Figure 6) les points de la scène en début et fin du temps d'exposition, on constate que les déplacements ne sont pas tous horizontaux et ne présentent pas tous la même direction.

Ces deux exemples (Figure 5 et Figure 6) amènent au constat suivant : une formulation du phénomène par un modèle de flou uniforme sur toute l'image (c'est-à-dire une PSF identique pour tous les pixels) n'est pas adaptée au problème présent.

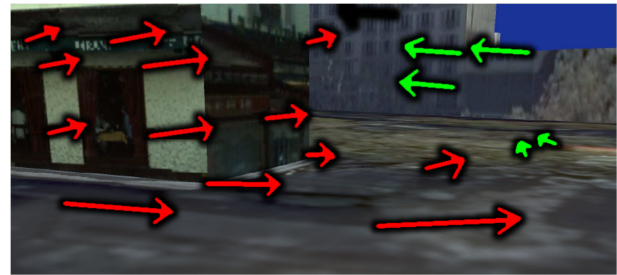


Figure 6. Approximation manuelle de PSF par des segments : déplacements entre le début et la fin de l'exposition

1.3. Travaux antérieurs

La déconvolution du flou dans les images numériques a été traitée depuis plusieurs décennies [WIENER, 1949] [RICHARDSON., 1972] [LUCY, 1974] mais les puissances de calcul de l'époque en limitaient fortement les applications. Les avancées techniques de la dernière décennie, se traduisant notamment par des puissances de calcul accrues, ont ouvert le domaine à de nouvelles perspectives.

Par exemple, en adoptant un système de prise de vue particulier, [RASKAR et al., 2006] obtient des résultats intéressants pour la reconstitution d'images affectées par un flou de bougé. Cette méthode a pour effet de mieux conserver les hautes fréquences au moment de la prise de vue et d'améliorer ainsi après déconvolution le rendu des détails lors de la reconstruction de l'image floue. Une image floue a par définition perdu un certain nombre de détails, hautes fréquences, qu'il est très difficile de faire réapparaître quelle que soit la méthode utilisée pour la reconstruction. La proposition de [RASKAR et al., 2006] est de modifier la manière dont l'obturateur de l'appareil de prise de vue fonctionne. Au lieu de laisser ouvert celui-ci en permanence pendant le temps d'exposition, l'auteur propose d'effectuer une série d'ouvertures et de fermetures non-périodiques dont la durée est parfaitement connue. La séquence binaire d'obturation optimale est déterminée de telle sorte que le conditionnement de la matrice de flou soit le plus faible possible ce qui améliore la qualité de l'inversion en diminuant les régions, dans le domaine fréquentiel, où la PSF s'annule. La méthode de reconstruction utilisée ensuite est soit une méthode de déconvolution classique [WIENER, 1949] ou encore [RICHARDSON., 1972] [LUCY, 1974] soit une résolution du système linéaire matriciel. Par contre, le flou est modélisé comme provenant d'une caméra en translation photographiant une scène à une profondeur donnée, ce qui ne correspond pas à notre problème où la profondeur de la scène est variable et où le véhicule n'est pas toujours en translation (dans les virages par exemple comme illustré en Figure 2 et Figure 6).

Il existe aussi des méthodes dites de « déconvolution aveugle » lorsque l'on ne possède pas une connaissance de la PSF. Ces approches font souvent appel à des processus itératifs qui permettent de retrouver progressivement la PSF dans les images floues pour les reconstruire ensuite par des méthodes traditionnelles ou plus performantes comme [FERGUS et al., 2006] et [SHAN et al., 2008] qui proposent respectivement deux algorithmes qui donnent tous deux de très bons résultats, notamment, pour [SHAN et al., 2008], en limitant les effets de rebond lors de la correction en utilisant une version modifiée de [RICHARDSON., 1972] [LUCY, 1974]. Or ces approches se limitent à retrouver une PSF par image, ce qui ne convient pas à une scène présentant des objets répartis à différentes profondeurs. De plus, nous disposons de données de navigation qui rendent envisageable une estimation directe de la PSF. Ces approches dites « aveugles » ne conviennent donc pas à notre étude.

Le fait de ne pas se limiter à une seule PSF (flou homogène) pour toute l'image est une avancée significative introduite notamment dans [WHYTE et al., 2012] qui présentent une nouvelle modélisation des PSF qui prend en compte les propriétés géométriques des caméras. Le principe utilisé est de considérer que chaque rotation de la caméra peut être ramenée à une rotation autour de son centre optique. Une translation est appliquée pour ramener l'ensemble des centres des rotations sur le centre optique. Dans le cas d'un flou de bougé, ces translations restent peu importantes. Ce modèle n'est donc pas adapté pour de grandes translations où les effets de parallaxe deviendraient significatifs. Or c'est le cas des images prises avec notre système de numérisation (Figure 2).

Une méthode proposée par [LEVIN et al., 2008] permet de corriger une image floue prise de sujets en mouvement horizontal à partir d'une caméra qui effectue un mouvement bien particulier sur le même axe. En imposant à sa caméra de faire un mouvement à l'accélération constante le long d'un axe, une nouvelle image floue est obtenue dont le mouvement est connu avec précision. Le mouvement retenu est le suivant : effectuer un aller-retour pendant le temps d'exposition en conservant une accélération constante. Les résultats obtenus montrent que le flou résultant d'un objet en mouvement unidirectionnel peut être très correctement corrigé par la déconvolution de l'image indépendamment de l'amplitude du mouvement horizontal. Cette approche est très intéressante car au-delà du fait de ramener un problème complexe à une déconvolution avec un noyau de flou uniforme dans une image, il s'agit d'une première approche d'un flou causé par des objets en mouvement. Elle n'est cependant pas applicable à notre étude car le mouvement du véhicule d'acquisition ou des objets en mouvement photographiés n'est ni constant ni unidirectionnel (Figure 6).

L'introduction de données inertielles par [JOSHI et al., 2010] qui propose une méthode automatique de

correction d'image peut s'appliquer à des flous qui vont jusqu'à 100 pixels avec une moyenne de 30 pixels. Les données inertielles sont collectées à partir de gyroscopes et d'accéléromètres positionnés sur chaque axe et permettent d'obtenir les mouvements de la caméra en translation et en rotation (6 degrés de liberté). Ces travaux prolongent ceux de [PARK et al., 2008] qui utilisaient des accéléromètres pour mesurer le mouvement de translation sur trois axes. La prise en compte des rotations permet d'améliorer la détermination du noyau de flou pour chaque pixel. La formulation de la formation d'une image floue suit l'équation (2). Cette méthode se rapproche de notre problème. L'impact de la variation de la profondeur de la scène pour le calcul des PSF apparaît de manière claire. Les véhicules terrestres de cartographie mobile disposent généralement de données inertielles qui pourront être utilisées en s'inspirant de cette approche.

Enfin les travaux plus récents de [HUE et al., 2014] et [ARUN et al., 2015] abordent le problème sous l'angle d'une « déconvolution aveugle » où les paramètres à estimer sont les mouvements de la caméra (rotations et translations) ainsi que la profondeur de la scène photographiée. Les estimations se font par itérations. L'approche « aveugle » à la fois du problème de déconvolution et d'estimation de la profondeur est très complète et peut s'appliquer en théorie à notre problème. Néanmoins, le modèle évoqué dans les articles (composition de rotations et de translations uniformes) ne convient pas à la complexité du mouvement de notre véhicule. Le fait d'avoir à notre disposition des données inertielles ainsi qu'une information de profondeur nous invite encore une fois à ne pas retenir ces approches dites « aveugles ».

1.4. Approche retenue

La restauration d'images de scènes complexes prises avec un long temps d'exposition depuis un véhicule en mouvement est un problème qui est encore assez peu traité. Le fait de pouvoir disposer, grâce au système de numérisation mobile utilisé dans cette étude, d'informations supplémentaires comme les données de positionnement et de navigation ainsi qu'une représentation 3D de l'environnement de la scène photographiée offre des opportunités intéressantes et des pistes inédites dans la résolution de ce problème.

On s'attachera donc à exploiter au maximum par la suite cette spécificité technique liée à l'acquisition des images prises par notre système de cartographie mobile. Pour cela, une approche locale, pixel par pixel, suggérée par la formulation (2) et une modélisation fine des PSF affectée à chaque pixel composant chaque ligne de la matrice A sera mise en œuvre afin d'obtenir une correction du flou en tous points de l'image.

2. Méthodologie

La configuration complexe des acquisitions terrestres et l'inhomogénéité du flou résultant nous invitent à adopter une approche locale formulée par l'équation (2) et à déterminer des méthodes de résolution de cette dernière en considérant deux méthodes d'inversion de la matrice A , méthodes dont les performances respectives seront discutées ultérieurement.

2.1. Détermination locale de la PSF

L'impact de la distance objet-appareil de prise de vue est ainsi à prendre en compte car les variations observées peuvent être significatives (Figure 5 et Figure 6). Ce qui d'emblée rend moins intéressantes un certain nombre des méthodes vues précédemment car, dans le cas présent, le flou de bougé ne peut être considéré comme uniforme dans l'image. Se restreindre à un flou uniforme c'est perdre de la précision dans les calculs. Il n'est pas envisageable non plus de se rapporter à la méthode de [LEVIN et al., 2008] compte tenu des spécificités matérielles actuelles des systèmes terrestres de cartographie mobile.

La solution proposée dans cet article semble s'approcher de la méthode développée par [PARK et al., 2008] et [JOSHI et al., 2010] dont l'intérêt est de traiter un flou non uniforme par détermination du mouvement de la caméra avec les données inertielles de son mouvement. Cependant, notre approche diffère sur deux aspects. Les variations de la profondeur de la scène sont prises en compte avec notre approche non pas par zones mais pixel par pixel pour la reconstruction. De plus, le flou de bougé considéré ici est sensiblement plus important que celui étudié par [SHAN et al., 2008], il n'est donc pas envisageable d'utiliser la méthode d'interpolation bilinéaire pour obtenir chaque PSF. La conservation de l'énergie et la continuité du mouvement imposent de rechercher une méthode de discrétisation des segments reliant les différents points de la PSF.

2.1.1. Gestion de la profondeur

Les variations de la profondeur de la scène sont prises en compte avec notre approche non pas par zones mais pixel par pixel pour la reconstruction. A ce stade, le choix de travailler sur les images de synthèse s'impose car ces dernières proposent une estimation parfaite de la trajectoire et disposent de données de profondeur pour chaque pixel. Pour parvenir à un tel résultat avec les images réelles, il serait nécessaire d'obtenir une modélisation précise de la profondeur.

Plusieurs pistes sont envisageables :

- Segmentation de l'image par zones de profondeur constante (profondeur à estimer).
- Modélisation 3D de la ville.
- Utilisation de nuages lasers 3D acquis par le véhicule ou un autre scanner laser.

Un travail ultérieur permettrait d'obtenir la modélisation nécessaire au traitement complet de reconstruction des images floues réelles.

2.1.2. Interpolation de la trajectoire

Afin de déterminer le mouvement du véhicule durant l'acquisition d'une image (10 millisecondes) il est nécessaire de disposer de données inertielles à une fréquence supérieure à 300Hz. Or certaines centrales inertielles fournissent des données à une fréquence moindre. Par exemple, dans le cadre de ce travail, nous disposons de données à 100Hz. Cette fréquence n'est pas assez élevée pour permettre une modélisation précise pendant le temps d'exposition. Il est donc nécessaire de densifier cette donnée : connaissant plusieurs points de la trajectoire, il est possible de faire une interpolation, avec une spline cubique, et d'extraire plusieurs valeurs de position pendant le temps de prise de vue.

2.1.3. Estimation de l'extension du flou

La détermination de la PSF repose sur les deux points précédents : les données de trajectographie associées à une carte de profondeur donnent plusieurs points correspondant à des points clefs de la PSF qui doit par la suite être intégrée sous la forme d'une image. Cette intégration passe par une seconde interpolation dans l'espace image, elle sera détaillée dans la partie suivante qui traite de la construction de la matrice A .

2.2. Construction de la matrice A

La matrice A introduite à l'équation (2) peut être interprétée de deux manières, à savoir suivant ses lignes ou bien suivant ses colonnes :

- Une ligne de A correspond à la contribution des pixels de l'image latente (nette) pour un pixel de l'image observée (floue). Cela correspond à la formulation d'un produit de convolution.
- Une colonne de A correspond à la contribution d'un pixel de l'image latente sur plusieurs pixels de l'image observée. Cela correspond au déplacement d'un point de l'image latente dans l'image observée.

Une colonne de la matrice A peut donc être directement interprétée comme la réponse impulsionnelle du système (PSF). La matrice A sera donc construite colonne par colonne, ce qui revient, pour chaque pixel de l'image nette, à calculer son déplacement pendant le temps d'exposition de l'image floue. Le schéma ci-après (Figure 7) détaille notre méthode pour remplir une colonne de la matrice.

2.2.1. Déplacement du pixel durant l'exposition

Dans un premier temps, les coordonnées du pixel (dans notre exemple, le pixel d'indice 12 correspondant au douzième pixel en lecture ligne par ligne de l'image) associé à la colonne considérée sont calculées dans un

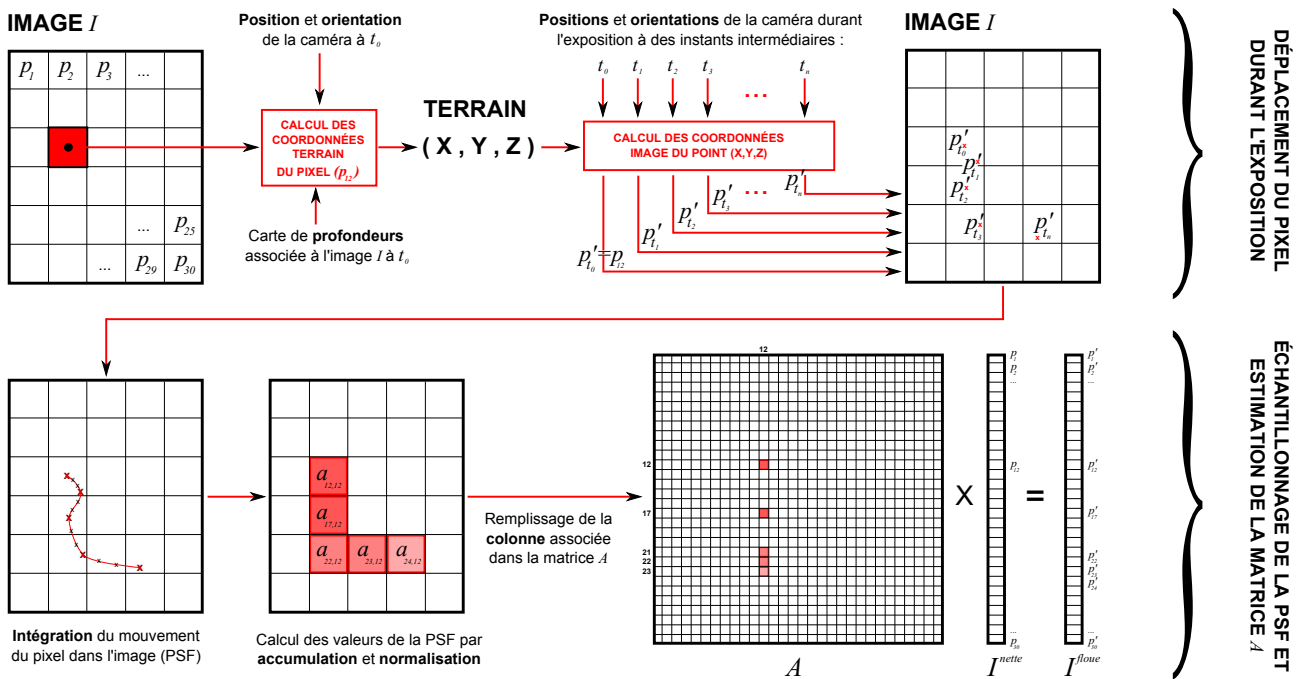


Figure 7. Détermination de la PSF et construction de la matrice A

référentiel terrain (X, Y, Z) . Il est pour cela nécessaire de disposer de données de positionnement de la caméra (position et orientation obtenues par traitement des données de navigation, GPS et centrale inertielle) ainsi que d'une carte de profondeur (obtenue ici par le modèle 3D de synthèse mais théoriquement calculable à partir d'un nuage de points laser) correspondant à la position et à l'orientation de la caméra à un instant précis : il a été ici choisi de le faire correspondre au début du temps d'exposition.

La deuxième étape consiste à projeter dans l'image le point (X, Y, Z) pour différents positionnements du sommet de prise de vue fournis par les fichiers d'orientation. Ces derniers sont calculés à partir des données de navigations : les orientations correspondent généralement à la position des caméras à l'instant du déclenchement, mais peuvent aussi être déterminées pour d'autres instants tels que les mesures de la centrale inertielle.

2.2.2. Échantillonnage de la PSF

On obtient ainsi une représentation vectorielle du mouvement du point dans l'image, c'est-à-dire la PSF du point. Afin de représenter la PSF sous forme de pixels, une étape d'intégration par interpolation est nécessaire. En effet, l'échantillonnage des points obtenus par calcul à partir des poses estimées peut être insuffisant pour une représentation sous forme d'une image. Dans notre exemple, les points projetés occupent les pixels p_{12} , p_{17} , p_{22} et p_{24} . Or la trajectoire du point passe par le pixel p_{23} . Le pas d'interpolation doit être inférieur à la fréquence de Nyquist, à savoir, dans notre cas, un tiers de pixel. Le remplissage des pixels de l'image de la PSF

se fait par accumulation et normalisation : pour chaque pixel, on compte le nombre de points de la PSF qui s'y trouvent puis on divise par le nombre de points total, calculés et interpolés, de la PSF. Cette étape de normalisation garantit une conservation de l'énergie entre l'image observée I^{floue} et l'image latente I^{nette} , corrigée du flou. Plus la valeur en un pixel de la PSF est élevée, plus le point a séjourné longtemps au niveau de ce pixel, prenant ainsi en compte l'accélération. Dans le cas illustré par notre exemple, le point associé au pixel p_{12} de l'image latente apparaît dans les pixels p_{12} , p_{17} , p_{22} , p_{23} et p_{24} avec une pondération respective de $a_{12,12}$, $a_{17,12}$, $a_{22,12}$, $a_{23,12}$ et $a_{24,12}$.

2.2.3. Estimation de la matrice A

L'image de la PSF est finalement représentée sous forme de colonne pour être placée dans la matrice A à la colonne correspondante à l'indice du pixel considéré dans l'image latente I^{nette} . Par définition du produit matriciel, la valeur de ce dernier sera répartie dans les pixels de l'image observée I^{floue} en suivant la PSF localement associée.

2.3. Inversion de la matrice A

Des difficultés apparaissent assez rapidement lorsqu'il s'agit d'inverser la matrice A puisque il n'y a pas de garantie que l'inversion puisse se faire. On parle de problème mal posé. On recherche alors une valeur approchée de l'inverse : le pseudo-inverse.

Une étude particulière du conditionnement de cette matrice est alors nécessaire. Le conditionnement mesure la difficulté de calcul numérique d'un problème.

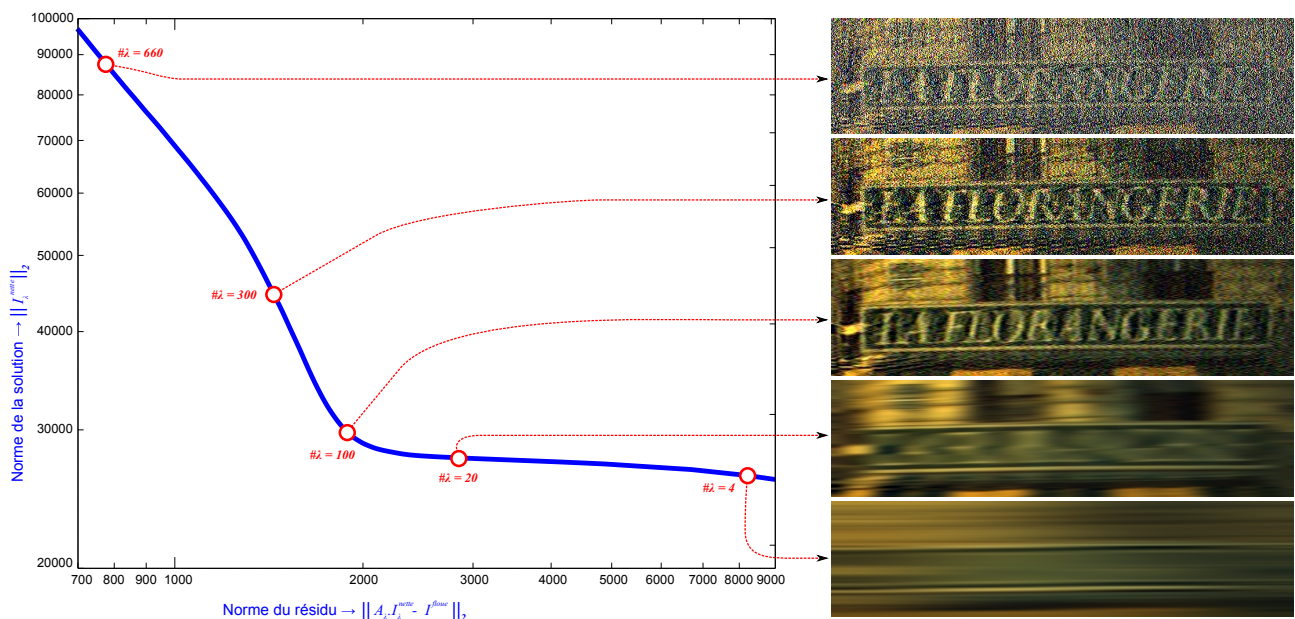


Figure 8. Exemple de reconstruction par SVD pour plusieurs seuils de troncature

Plus le conditionnement est bas, plus stable numériquement est la résolution du problème.

Dans le cas d'un système linéaire à résoudre, ce qui est notre cas, le conditionnement donne l'erreur commise sur la solution lorsque les données sont perturbées. Si cette erreur est trop importante, la solution peut être inexploitable. Des méthodes de régularisation plus ou moins complexes existent et permettent d'améliorer le conditionnement. Une fois pris en compte le problème de conditionnement, il s'agit de déterminer comment calculer l'inverse ou le pseudo-inverse. Un certain nombre de méthodes existent pour résoudre de tels systèmes. Parmi celles-ci, deux ont été retenues pour ce travail : la décomposition en valeurs singulières et la méthode du gradient conjugué.

2.3.1 La décomposition en valeurs singulières

Afin de généraliser aux matrices non carrées la décomposition en valeurs propres [GOLUB et al., 1996] propose une décomposition de matrice appelée décomposition en valeurs singulières, notée par la suite SVD pour *Singular Value Decomposition*. Cette méthode apporte une première réponse à l'inversion de notre système tout en pouvant s'appliquer au cas général d'images I^{floue} et I^{nette} de tailles différentes impliquant une matrice A non carrée.

Considérons la matrice A de l'équation (2), matrice de réels de dimension $m \times n$, alors il existe une décomposition de cette matrice telle que :

$$A = U.S.V^T, \quad (6)$$

où U et V sont deux matrices orthogonales de vecteurs singuliers respectivement à gauche et à droite. S est la matrice de taille $m \times n$ dont la diagonale à valeurs réelles positives porte les valeurs singulières ordonnées

par ordre décroissant. La pseudo-inverse, notée A^{-1} , s'obtient de la manière suivante :

$$A^{-1} = V.S^{-1}.U^T, \quad (7)$$

où S^{-1} est une matrice de taille $m \times n$ dont la diagonale se compose de l'inverse des valeurs singulières non nulles ordonnées par ordre croissant. L'utilisation de cette méthode offre de très bons résultats pour la reconstruction d'image.

Les valeurs singulières les plus faibles sont les plus sensibles au bruit et, par conséquent, leur prise en compte lors de la restauration dégrade le résultat obtenu. Il est préférable de tronquer la suite de ces valeurs à partir d'un certain seuil pour améliorer le rendu de la reconstruction.

Choix des paramètres et courbe « en L »

Il existe plusieurs méthodes pour déterminer le paramètre de seuil optimal pour la troncature. L'une d'entre elle, [HANSEN, 1992], permet d'obtenir celui-ci de manière graphique. Ce graphique connu sous le nom de courbe « en L » est une représentation de l'erreur de reconstruction en abscisse et de la valeur du terme de régularisation n en ordonnée. La présentation de ces résultats avec les échelles logarithmiques fait apparaître cette forme caractéristique (Figure 8) dite « en L ». Pour les faibles valeurs du seuil, le bruit est dominant, pour les fortes valeurs, le lissage est prépondérant. Le paramètre optimal à retenir est « le coin », c'est-à-dire le point où le rayon de courbure est minimal.

Une particularité de cette méthode est d'améliorer le conditionnement de la matrice en modifiant le seuil de troncature. En effet, le conditionnement est égal au rapport entre la valeur singulière la plus grand et la valeur la plus faible.

Elle est en théorie très intéressante dans notre cas, car elle permet non seulement de pouvoir déterminer un pseudo-inverse à la matrice A mais également en paramétrant la reconstruction en fonction du nombre de valeurs singulières conservées de contrôler l'influence du bruit et ses effets sur la reconstruction.

2.3.2. Méthode du gradient conjugué

Le premier article à proposer cette approche [HESTENNES, 1975] présente la méthode du gradient conjugué pour résoudre des systèmes d'équations linéaires, avec une matrice A définie positive.

Il s'agit d'un algorithme itératif qui converge après un nombre fini d'itérations, au plus égal à la dimension du système linéaire, vers une solution approchée. L'intérêt pratique de cette méthode est de réduire considérablement le temps de calcul nécessaire par rapport à la recherche classique de l'inverse (lorsqu'il existe). Elle est d'autant plus optimale que la solution approchée utilisée pour initialiser l'algorithme est proche de la solution recherchée.

Cette méthode est applicable dans notre cas en considérant la matrice carrée $A^T.A$ et ajoutant éventuellement un facteur d'amortissement positif $\varepsilon.Id$ afin d'apporter plus de robustesse au bruit. Ce qui revient finalement à minimiser la forme quadratique suivante :

$$f(x) = \frac{1}{2} x^T . (A^T . A + \varepsilon . Id) . x - x^T A^T . I^{floue} . \quad (8)$$

Cette approche purement itérative peut s'avérer très intéressante en pratique étant donné la taille conséquente des matrices à inverser. Cependant, elle est moins robuste à paramétrer que la précédente, faisant intervenir la SVD.

Choix des paramètres

Dans le cadre de ce travail, le choix des paramètres ε ainsi que le critère d'arrêt des itérations entrant en jeu dans l'algorithme du gradient conjugué a été purement empirique. Une analogie avec la courbe « en L » serait éventuellement à envisager.

3. Résultats

La méthode a été testée sur des données synthétiques obtenues à partir de données d'orientation réelles interpolées tous les centièmes de seconde grâce aux données brutes de trajectoire à 100Hz. On verra par la suite que la configuration très particulière de cette prise de vue synthétique (Figure 2) met en évidence certaines limites de notre approche.

3.1. Représentation du flou

Les résultats intermédiaires présentés ici visent d'une part à illustrer l'approche retenue pour le calcul des PSF reposant sur les trajectoires et les informations relatives à la scène 3D et d'autre part à vérifier la cohérence

entre notre modèle de flou introduite à l'équation (2) et la réalité de l'acquisition obtenue par simulation (Figure 2).

3.1.1. Carte de représentation des PSF

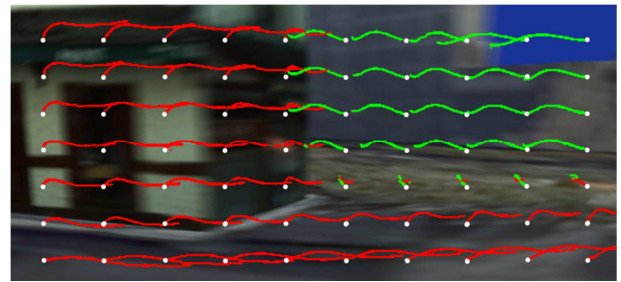


Figure 9. PSF locales calculées suivant une grille

Dans la méthode introduite précédemment (Figure 7) la construction de la matrice A passe premièrement par une estimation locale de la PSF en chaque pixel. Une représentation des PSF calculées sur une grille régulière de pixels (Figure 9) renvoie un résultat globalement cohérent avec le champ de déplacements obtenu manuellement (Figure 6). Néanmoins, en examinant attentivement cette carte de PSF (Figure 9) on constate que malgré une certaine cohérence globale, la PSF, pour de longs temps de pause, ne peut pas se réduire à un segment comme c'est le cas de l'image réelle (Figure 5) prise suivant une direction constante à une vitesse uniforme et parallèlement à la scène photographiée.

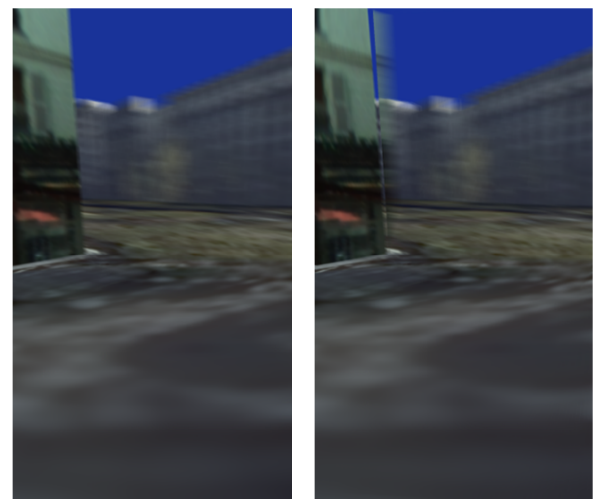


Figure 10. Deux manières de simuler une image floue.
À gauche : par intégration d'images de synthèse nettes.
À droite : après multiplication d'une image nette par A .

3.1.2. Validation de l'approche matricielle

Afin de valider la construction de la matrice A , un test a été effectué en rendant floue la première image simulée en multipliant cette dernière par A . En comparant (Figure 10) avec une image floue simulée par moyenne sur des images de synthèse prises suivant différentes positions, correspondants à la trajectoire du véhicule durant la prise de vue, et considérée comme proche de la réalité, on constate que le résultat est quasi-identique à

l'exception du bord du bâtiment. Ce phénomène peut être interprété par le fait que notre modélisation de la PSF ne tient pas compte des parties cachées et intègre ainsi des pixels des objets du premier plan alors qu'elle modélise un objet de l'arrière-plan.

3.2 Performances

Avant d'aborder la reconstruction à proprement parler, une remarque pratique concernant la méthode d'inversion de la matrice A nous invite plutôt à privilégier une approche par rapport à une autre.

Malgré des résultats très encourageants et une grande robustesse, la SVD n'est pas applicable à l'inversion de très grandes matrices en un temps raisonnable. En effet l'examen des courbes de temps de calcul pour ces deux méthodes (Figure 10 et Figure 11) montre que l'approche itérative du gradient conjugué est beaucoup plus performante en temps de calcul et par conséquent plus adaptée à notre problème. L'inconvénient est son manque de robustesse : le choix des paramètres étant laissé au soin de l'opérateur.

Après plusieurs essais, les paramètres empiriques retenus dans le cas du gradient conjugué sont $\varepsilon = 0,01$ et $nb_{itérations} = 80$.

3.3. Reconstruction

Le test effectué sur la donnée synthétique donne des résultats acceptables. L'image reconstruite par notre méthode (Figure 13, gauche) ressemble globalement à l'image nette théorique recherchée (Figure 13, droite). On relève néanmoins des artéfacts de reconstruction au niveau de l'arrête verticale de l'immeuble au premier plan qui semble se répéter au second plan.

Comme évoqué plus tôt, lors de la validation de la matrice A (Figure 10, droite). Le modèle de PSF ne tient pas compte des parties cachées, typiquement dans notre jeu de donnée, lorsque l'immeuble à l'arrière-plan passe derrière celui du premier plan. Cela met en avant une limite majeure de notre modèle.

3.4. Discussion

La méthode retenue dans cette étude est applicable en l'état pour les scènes où la profondeur varie sans discontinuité et, par conséquent, dépourvues de partie cachée. La configuration idéale serait une acquisition avec une trajectoire quasi-rectiligne et parallèle aux façades : ce qui correspond à la très grande majorité des cas de prises de vues.

3.4.1. Vers une gestion des parties cachées

Cette étude met en avant les limites du modèle de convolution local formulé à l'équation (2) dans la mesure des scènes comprenant des parties cachées. Une alternative envisagée serait, lors du calcul de la PSF, de ne retenir que les points visibles en s'aidant d'une carte de profondeur pour chaque position élémentaire

composant la prise de vue durant le temps d'exposition considéré.

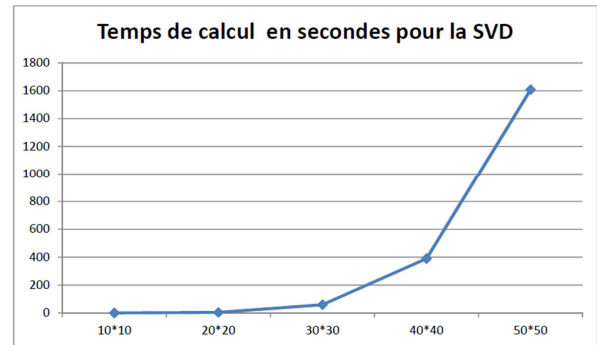


Figure 11. Temps de calcul (secondes) d'une SVD en fonction de la taille d'une image carrée (pixels)

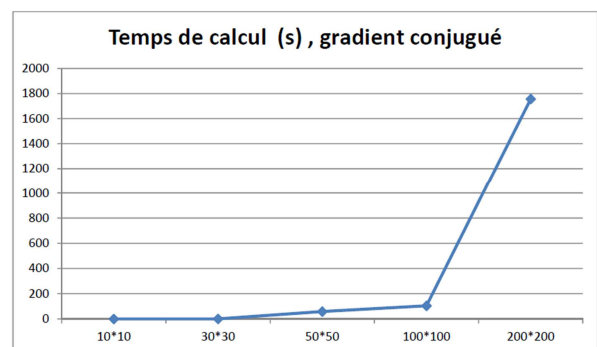


Figure 12. Temps de calcul (secondes) de l'approche par gradients conjugués en fonction de la taille d'une image carrée (pixels)

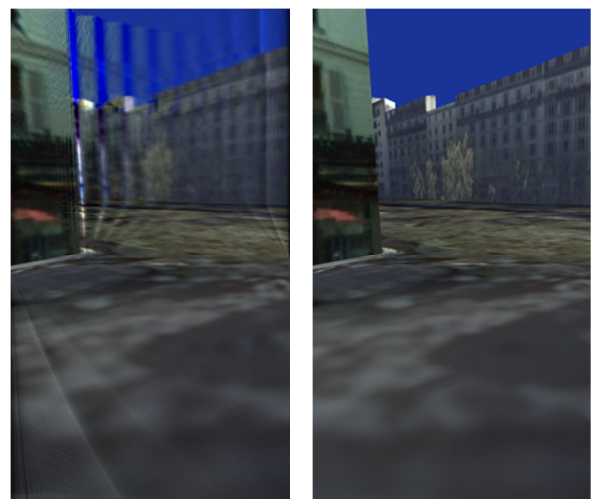


Figure 13. Exemple de reconstruction (à gauche) à partir d'une image floue de synthèse (Figure 10) comparée à l'image nette théorique (à droite).

L'image ne sera pas corrigée du flou au niveau des parties cachées, mais cela palliera en partie les défauts de reconstruction observés (Figure 13). Ce cas de figure se présentant assez rarement, à savoir essentiellement dans les virages, en présence de ruelles perpendiculaires ou de certains mobiliers urbains tel que

les lampadaires ou les troncs d'arbre. Notons que ces derniers n'apparaissent pas dans le modèle 3D utilisé pour le jeu de données de synthèse.

Sinon, une autre approche serait de ne pas chercher à reconstruire l'image à l'instant t_0 correspondant au début de l'exposition mais à un instant moyen entre le début et la fin du temps d'exposition. L'impact des parties cachées sur l'image reconstruite sera alors moindre.

3.4.2. Autres pistes d'améliorations

Le modèle optique (fonction de transfert de modulation) n'a pas été évoqué dans ce travail car les tests ont été menés sur des images de synthèse. Ce modèle peut néanmoins être intégré dans l'estimation de la PSF du flou de bougé en accumulant non pas des points (Figure 7) mais des PSF relatives à l'optique. La PSF du flou de bougé ainsi modélisée serait alors l'équivalent d'une source ponctuelle vue par la caméra tout en tenant compte de ses défauts optiques.

Le fait que la matrice A se présente bien intrinsèquement comme une matrice creuse n'a pas été pris en compte dans la méthode d'inversion. La détermination d'un mode de résolution optimal sera le sujet de travaux ultérieurs.

3.4.3. Limitations possibles sur les données réelles

L'application de notre approche sur des données réelles est donc à envisager avec circonspection. Les problèmes pouvant possiblement apparaître lors d'une application sur des données réelles sont nombreux :

- Défaut de synchronisation, de recalage et d'échantillonnage entre les données d'acquisition (images, laser, navigation).
- Présence d'objet en mouvement (voitures).
- Reconstruction des reflets des lampes dans les vitres (non modélisés par notre approche).

Bien qu'offrant certaines limitations, notre approche peut raisonnablement augmenter la qualité d'images présentant des extensions de flou moins important. Cela peut être le cas lors des acquisitions par temps très couvert ou au crépuscule.

3.4.4. Autres applications de la méthode

Il est enfin judicieux de considérer deux autres applications à nos travaux :

- Un outil de quantification local *a posteriori* de la netteté des images.
- Un outil d'aide à l'acquisition *a priori*.

Dans le premier cas, seule la partie calcul local des PSF sera retenue et pourra servir de métadonnées dans le cadre des applications nécessitant une connaissance la qualité des images.

Dans le second cas, une connaissance *a priori* de la scène 3D et un calcul des PSF pour une vitesse et un

temps d'exposition donnés à partir de trajectoires simulées indiquera à l'opérateur les zones où l'acquisition risquerait d'être floue, comme dans une ruelle étroite par exemple.

4. Conclusions et perspectives

En conclusion de ce travail, une approche du complexe problème de la reconstruction d'images floues provenant d'un système terrestre de cartographie mobile a pu être proposée. Cette méthode est innovante dans le sens où elle prend en compte à la fois les données inertielles du mouvement mais également les données de profondeur de la scène et ce pour chaque pixel. De plus, il serait envisageable d'étendre son application à d'autres systèmes d'acquisitions mobiles disposant des mêmes spécificités techniques, à savoir des données de trajectographie ainsi que d'un modèle 3D de la scène, tels que par exemple les drones.

Le choix de travailler sur des données de synthèse est justifié par le fait qu'il s'agit avant tout d'une étude de faisabilité. Ces travaux, bien qu'essentiellement prospectifs, ont abouti à des constats globalement encourageants.

Des résultats cohérents ont ainsi pu être obtenus dans le domaine de la mesure globale du flou, qui s'est traduit notamment par la production de cartes précises de champs de PSF. Enfin, un algorithme d'inversion par la méthode du gradient conjugué a été retenu donnant des résultats intéressants en termes de temps de calcul et de qualité d'inversion. Cependant, les résultats de reconstruction obtenus sont loin de tous donner pleine satisfaction et nécessitent, dans un travail ultérieur, de s'attacher à l'étude des parties cachées et à leur reconstruction ainsi qu'à celle des obstacles en mouvement.

Dans l'état actuel, ce travail a avant tout abouti à un nouvel outil de quantification de flou fournissant une information précieuse sur la qualité des images prises pendant une session d'acquisition d'un système de cartographie mobile. Enfin, il n'est pas pour autant envisagé de mettre de côté les tests réels. Bien au contraire, les évolutions récentes dans le domaine de la sensibilité des capteurs des appareils photographiques numériques, qui va en augmentant, ouvrent de nouvelles perspectives pour les prises de vue nocturnes, notamment en réduisant les temps d'exposition.

Remerciements

Nous remercions sincèrement MM. Bertrand Cannelle, Bruno Vallet et Jean-Pierre Papelard qui ont consacré une partie de leur temps à la génération des données nécessaires au travail présenté ici.

Références

- Arun M., Rajagopalan A. N., Seetharaman G., 2015. Multi-Shot Deblurring for 3D Scenes. In 2015 IEEE International Conference on Computer Vision Workshop (ICCVW): 19-27
- Fergus R., Singh B., Hertzmann A., Roweis S.T., Freeman W., 2006. *Removing camera shake from a single photograph*. ACM Transactions on Graphics 25 (2006): 787-794.
- Golub G., van Loan C.F., 1996. *The Singular Value Decomposition*. John Hopkins University Press, Baltimore.
- Hansen C., 1992. *Analysis of Discrete Ill-Posed Problems by Means of the L-Curve*. SIAM Review 34 n°4 (1992): 561-580.
- Hestennes M.R., 1975. *Pseudoinversus and conjugate gradient*. Communications of the ACM. 18, n°1 (1975): 40-43.
- Hu Z., Xu L., Yang M.-H., 2014. Joint depth estimation and camera shake removal from single blurry image. In 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR): 2893–2900.
- Joshi N., Kang S.B., Zitnick C.L., Szeliski R., 2010. *Image Deblurring using Inertial Measurement Sensors*. ACM SIGGRAPH.
- Levin A., Sand P., Cho T.S., Durand F., Freeman W.T., 2008. *Motion-invariant photography*. ACM Transactions on Graphics 27 (2008): 1-71.
- Lucy L., 1974. *An iterative technique for the rectification of observed distributions*. Journal of Astronomy 79 (1974): 745-754.
- Paparoditis N., Papelard J.P., Cannelle B., Devaux A., Soheilian B., David N., Houzay E., 2012. *Stereopolis II: A multi-purpose and multi-sensor 3D mobile mapping system for street visualisation and 3D metrology*. Revue Française de Photogrammétrie et de Télédétection 200: 69-79.
- Park S.Y., Park E.S., and Kim H.I., 2008. *Image deblurring using vibration information from 3-axis accelerometer*. Journal of the Institute of Electronics Engineers of Korea. SC, System and control: 1-11.
- Raskar A., Agrawal A., Tumblin J., 2006. *Coded Exposure Photography: Motion deblurring using a fluttered shutter*. SIGGRAPH.
- Richardson W.H., 1972. *Bayesian-based iterative method of image restoration*. Journal of the Optical Society of America (1917-1983) 62 (1972): 55-59.
- Shan Qi, Jiaya Jia, Aseem A., 2008. *High-quality Motion Deblurring from a single Image*. ACM Transactions on Graphics 27, n°3, article 73.
- Whyte O., Sivic J., Zisserman A., Ponce J., 2012. *Non-uniform Deblurring for Shaken Images*. IJCV, n°98 (2012): 168-186.
- Wiener, N., 1949, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, Cambridge, Technology Press of Massachusetts Institute of Technology, and New York, Wiley.