

Recalage et fusion de modèles numérisés tridimensionnels de grande taille

MEMO-F-403 - Préparation au mémoire

Tim Lenertz

17 août 2014

Table des matières

1	Introduction	3
2	Etat de l'art	4
2.1	Documentation 3D	4
2.1.1	Scanner tridimensionnels	5
2.2	Théorie préliminaire	5
2.2.1	RANSAC	5
2.2.2	Méthode de Newton	6
2.2.3	Analyse procustéenne généralisée	7
2.3	Recalage	7
2.4	Iterative Closest Point	8
2.4.1	Calcul de la transformation par quaternions	9
2.4.2	Variantes de l'algorithme	10
2.4.3	Version simultanée de l'ICP	11
2.4.4	ICP généralisé	11
2.5	4-Points Congruent Sets	13
2.6	Normal Distributions Transform	14
2.7	Extended Gaussian Images	15
2.7.1	Application sur le recalage de nuages de points	16
2.8	Identification et correspondance de droites	17
2.9	Identification et correspondance de plans	17
2.10	Détecteurs de zones d'intérêt	17
2.11	Recalage via des photos	17
2.12	Photogrammétrie	17
2.12.1	Optique	17
2.12.2	Modèle pin-hole	18
2.12.3	Calibrage	19
2.12.4	Problème de correspondance	21
2.12.5	Reconstruction 3D	21
2.12.6	Combination photo-nuage de points	22
3	Projet prévu pour le mémoire	24

1 Introduction

Pour des projets de documentation 3D, des objets ou environnements sont souvent numérisés sous forme de *nuages de points*. Un nuage de points consiste uniquement en un ensemble de points situés sur les surfaces des objets, définis par des coordonnées dans un repère orthonormal donné, et attribués par des valeurs scalaires ou vectorielles comme par exemple une couleur RGB.

Ces jeux de données sont usuellement capturés par des scanners 3D à télémètre laser, ou par photogrammétrie. Ces données brutes ne représentent qu'une partie de l'objet, et sont limitées par le champ de vision de la caméra, les surfaces cachées (sur le côté opposé au scanner de l'objet), les limites de résolution pour des parties éloignées ou qui ont une texture complexe. Afin d'en synthétiser un nuage de points représentant le modèle complet, plusieurs traitement doivent être effectués, notamment les recalage : Pour toutes les nuages de points partielles, une transformation affine est appliquée aux points afin de les mettre dans un repère commun.

Plusieurs techniques (semi)-automatisées ont été développées qui permettent de trouver une approximation des positions et orientations relatifs des scans. Ils peuvent faire intervenir des photos, ou des marqueurs visuels ajoutés sur la scène. La matrice de transformation précise est ensuite déterminée algorithmiquement, afin de bien aligner les surfaces dans les différents ensembles de points. Typiquement, une forme de l'algorithme ICP¹ est utilisé, un algorithme qui ajuste la transformation en itérativement minimisant la distance entre les deux ensembles de points.

Ce mémoire se concentre sur le recalage de numérisations à distance d'un objet à grande dimensions, avec des numérisations à courtes distances de détails de l'objet. Donc les ensembles de points à recaler pourront avoir des densités de points très différentes, et auront un nombre de points très élevé. Le but est d'établir un workflow et de développer des algorithmes qui permettent d'effectuer ce type de recalage. Le travail est basé sur un projet de documentation 3D actuel du LISA.

1. Iterative Closest Point

2 Etat de l'art

Ce chapitre est essentiellement une synthèse de plusieurs ouvrages scientifiques sur des projets de documentation 3D par nuage de points, différents algorithmes de recalage, et des techniques photogrammétriques. Certains algorithmes et leurs bases mathématiques seront développés plus en détail. Le but est d'exposer une partie de la recherche existante qui pourrait être rélevante pour le mémoire.

2.1 Documentation 3D

Des nuages de points créés via des processus de numérisation 3D sont utilisés dans plusieurs domaines pour représenter des objets réels. Les modèles 3D peuvent passer de représentations détaillées d'objets petits à des bâtiments ou même des sites entiers. Des exemples d'usage sont par exemple la documentation détaillée de sites archéologiques [Center for Advanced Spatial Technologies, 2014] [Keiner, 2011] [Grussenmeyer *et al.*, 2012] et d'environnements urbains [Kersten *et al.*, 2006], des acquisitions aériennes de terrains, la vision algorithmique en robotique [Biber, 2003], la capture de formes 3D pour création de prothèses en médecine dentaire et orthopédique, la rétroingénierie, la création de modèles en animation 3D.

Pour collectionner les données brutes de l'objet pour lequel on veut générer un modèle 3D, deux techniques sont généralement utilisées : La *photogrammétrie* consiste à prendre des photos (possiblement stéréoscopiques) de l'objet, à partir desquelles on peut extraire algorithmiquement de l'information sur la profondeur des pixels. D'autre part, on utilise des *scanners tridimensionnels*, des appareils qui effectuent un balayage par laser de l'objet et produisent directement un nuage de points. Souvent les deux techniques sont combinées. Cela permet par exemple de compléter un scan laser très détaillé avec de l'information sur les couleurs des points.

Pour passer des données brutes au modèle final envisagé, un nombre important d'étapes de planification et de post-traitement des données doivent être effectués. Par exemple, un projet de documentation archéologique pour lequel des scans 3D et des données photogrammétriques sont utilisés peut passer par les étapes suivantes : [Lerma *et al.*, 2009]

- Planification et acquisition des données brutes. Les positions à partir desquelles les scans 3D sont pris doivent être planifiées minutieusement en fonction du terrain, de la forme de l'objet et des spécifications du scanner, afin d'assurer une couverture complète de l'objet, avec la résolution et fidélité ciblée. De même pour les positions où des photos sont prises.
- Filtrage des nuages de points bruts. Le but est d'enlever des erreurs grosses du scanner, comme des points isolés.
- Recalage des nuages de points pour les mettre dans un repère commun. Beaucoup de méthodes et algorithmes ont été développés pour automatiser un recalage sous différentes conditions, ce qui est le sujet de ce mémoire. Dans certains cas des markers sont placés sur la scène avant l'acquisition des scans et photos. Les nuages de points peuvent ensuite être fusionnés par simple concaténation des points.
- Maillage des nuages de points, par exemple par triangulation de Delaunay. Le but ici est d'obtenir un modèle qui reprend l'information sur les surfaces du modèle. Pour des formes complexes, par exemple de la végétation sur un mur, des procédures additionnelles d'analyse et de filtrage sont requises.
- Par des méthodes photogrammétriques [Tournas & Tsakiri, 2009] et radiométriques [Girod, 2013], les photos sont modifiées et analysées : On corrige des vignettes, distorsions et autres erreurs optiques induites par la lentille de la caméra, égalise les valeurs colorimétriques des différentes photos. Par des détecteurs de zones d'intérêt en reconnaissance d'image, on retrouve des markers, ou d'autres points marquants, afin de mettre en correspondance plusieurs photos et d'en extraire la pose de la caméra et des informations de profondeur.

- Recalage des photos sur le modèle pour texturer le maillage. Au cas un modèle non-maillé qui reste sous forme de nuage de points, les photos peuvent être utilisées pour coloriser les points. Ensuite la densité des points peut être uniformisée si nécessaire.
- Dans certains contextes, par exemple en rétroingénierie ou en architecture, le but est de finir avec un modèle CAD. Des méthodes ont été développées pour segmenter le modèle, et pour reconnaître des formes ou objets.

2.1.1 Scanner tridimensionnels

Les scanners tridimensionnels utilisent des rayonnements, typiquement des lasers, pour physiquement mesurer des informations de profondeur de leur environnements. Dans les dernières années, cette technologie a connu un développement important. Les scanners 3D disposent d'un niveau avancé d'automatisation, et peuvent produire des données de bonne qualité [Grussenmeyer *et al.*, 2012] à des coûts relativement abordables.

Scanners par temps de vol (LIDAR)

Les scanners à temps de vol utilisent un rayon laser qui est projeté dans le champ de vision en différentes directions en succession rapide. Ils disposent aussi d'un capteur qui détecte la lumière reflétée du point laser sur l'objet. On mesure le temps de retour des pulses laser.

Ces scanner sont aussi appelés scanners *LIDAR*¹, par analogie au *RADAR*² qui fonctionne selon un principe similaire avec des rayonnements radio, et au *SONAR*³ pour des ondes sonores dans l'eau.

Comme la direction du laser est connue, il manque la distance que le rayonnement laser a traversé du scanner jusqu'à l'objet pour arriver à des coordonnées sphériques du point. Pour trouver cette distance, le scanner mesure le temps de retour Δt entre l'envoi du pulse laser jusqu'à la détection de la lumière reflétée. On peut en déduire la distance $d = c \times \frac{\Delta t}{2}$, où c est la vitesse de la lumière. Pour pouvoir acquérir une valeur exacte de d , le scanner doit être muni d'un chronomètre très exact : La lumière prend 3.3 picosecondes pour parcourir un millimètre. Typiquement, les scanners LIDAR collectionnent aussi la valeur d'intensité du signal de retour reçu. Celle-ci peut par exemple être utile pour déterminer le type de matériel de l'objet.

Le balayage du champ de vision peut être fait par des miroirs rotatifs.

Le système fonctionne bien pour des surfaces à réflexion diffuse, mais pas pour des surfaces qui manifestent une réflexion spéculaire (par exemple des surfaces métalliques), ou des surfaces transparentes (par exemple des vitrines). En fonction du type d'objet à scanner différentes fréquences de rayonnement sont utilisés, allant de l'ultraviolet par le visible au proche infrarouge. Dans quelques applications (par exemple des mesures aériennes d'un terrain), la fréquence est choisie pour ne pas endommager des yeux, dans d'autres (par exemples applications militaires) les balayages doivent être invisibles.

Scanners par triangulation

Scanners par décalage de phase

Scanners par lumière structurée

2.2 Théorie préliminaire

2.2.1 RANSAC

L'algorithme RANSAC ("Random Sample Consensus"⁴), initialement décrit en [Fischler & Bolles, 1980], est une méthode générale qui permet d'aligner un *modèle*, défini par un vecteur de paramètres θ , à des échantillons p_i de données mesurés. L'algorithme procède en sélectionnant aléatoirement des échantillons pour y déduire un modèle possible, qui est ensuite raffiné itérativement. La présence d'une fraction élevée

1. Light detection and ranging
 2. Radio detection and ranging
 3. Sound navigation and ranging
 4. Consensus par échantillons aléatoires

d'erreurs de mesure parmi les échantillons (“outliers”⁵) rend inutilisable des solutions classiques comme la méthode des moindres carrés. Avec l'approche RANSAC, la probabilité inférieure de sélectionner parmi tous les échantillons un outlier mène à un bon alignement avec les échantillons corrects. (“inliers”) Contrairement à des méthodes non-probablistiques, RANSAC essaye de trouver une estimation initiale du modèle par un nombre minimal d'échantillons, et non pas par un nombre le plus grand possible.

Soit m le nombre minimal d'échantillons requis pour trouver un modèle $\hat{\theta}$ qui comprend ces m échantillons. Soit P l'ensemble d'échantillons. Soit $E(\theta, p)$ une métrique d'erreur qui est zéro si l'échantillon p correspond parfaitement au modèle $\hat{\theta}$. RANSAC procède en répétant la procédure suivante :

1. Soit $S \subset P$ un sous-ensemble composé de m échantillons choisis aléatoirement.
2. Calculer à partir S un modèle estimatif $\hat{\theta}$. Soit $S^* \subset P$ tous les points p de P qui correspondent au modèle : $S^* = \{p \in P : E(\hat{\theta}, p) < \epsilon\}$, pour une tolérance d'erreur ϵ donnée. S^* est appelé le *consensus set*.
3. Si $|S^*| \geq t$ pour un seuil t déterminé en fonction du nombre d'outliers, calculer l'estimation finale du modèle θ à partir de tous les échantillons du consensus set S^* , et terminer.
4. Si $|S^*| < t$, on rejète le modèle $\hat{\theta}$, et recommence à l'étape 1. Si cela arrive plus que N (nombre maximal d'itérations) fois, l'algorithme a échoué.

Une autre procédure possible serait d'au lieu d'arrêter à 3, répéter pour un nombre fixe d'itérations, et garder à la fin le modèle pour lequel $|S^*|$ était maximal.

Un exemple simple d'application serait d'aligner une droite (modèle) $y = mx + b$, définie par les paramètres $\theta = (m, b)$, à des points (échantillons) $p_i = (x_i, y_i)$. Une droite peut être trouvée à partir de $m = 2$ points, en posant $m = \frac{x_2 - x_1}{y_2 - y_1}$ et $b = y_1 - mx_1$. Pour ajuster une droite à plus que deux points (pas nécessairement alignés), la méthode des moindres carrés peut être utilisée. Une métrique d'erreur serait la distance du point à la droite, donnée par $E(\theta, p) = \frac{|y - mx - b|}{\sqrt{m^2 + 1}}$.

2.2.2 Méthode de Newton

La méthode de Newton est un algorithme qui a le but de trouver itérativement des racines d'une fonction réelle continue et dérivable. En partant d'un point x_0 proche à la racine, on construit la série $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$. Le point x_{k+1} est la solution de l'équation $f(x_k) + f'(x_k)(x - x_k) = 0$, donc le point d'intersection de la tangente de f en x_k avec l'abscisse. Le principe d'algorithme est que les tangentes donnent une approximation du comportement de la fonction. Donc quand la racine a été trouvée, x_k sera un point stationnaire ($x_k = x_{k+1}$).

La méthode peut aussi être appliquée pour résoudre des problèmes d'optimisation, c'est à dire de minimisation d'une fonction, on considérant les racines de sa fonction dérivée f' . La formule devient

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} \quad (2.1)$$

Ceci peut aussi être étendu pour la résolution de problèmes d'optimisation sur une variable multidimensionnelle \vec{x} . On remplace f' par le gradient⁶ $\nabla f(\vec{x})$, et $(f'')^{-1}$ par l'inverse de la matrice Hessienne⁷ $(\mathbf{H}f(\vec{x}))^{-1}$:

$$\vec{x}_{k+1} = \vec{x}_k - (\mathbf{H}f(\vec{x}_k))^{-1} \nabla f(\vec{x}_k) \quad (2.2)$$

Pour éviter de passer par le calcul de la matrice inverse, le vecteur $\hat{\Delta}x_k = (\mathbf{H}f(\vec{x}_k))^{-1} \nabla f(\vec{x}_k)$, la quantité à retrancher du résultat après chaque itération, peut être trouvée par résolution du système d'équations

-
5. dans le contexte de nuages de points, des points qui ne sont pas sur une surface de l'objet
 6. La gradient d'une fonction vectorielle dérivable $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est le vecteur des ses n dérivées partielles :

$$\nabla f(\vec{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right)^T$$

7. La matrice Hessienne d'une fonction vectorielle $f : \mathbb{R}^n \rightarrow \mathbb{R}$ dérivable deux fois est composée de toutes les combinaisons de ses dérivées partielles secondes :

$$\mathbf{H}_{i,j} f(\vec{x}) = \frac{\partial^2 f}{\partial x_i \partial x_j}$$

linéaires $(\mathbf{H}f(\vec{x}_k)) \hat{\Delta}x = \nabla f(\vec{x}_k)$.⁸ Ecrit sous forme non-matricielle :

$$\left\{ \begin{array}{l} \frac{\partial^2 f}{\partial x_1 \partial x_1} \hat{\Delta}x_1 + \frac{\partial^2 f}{\partial x_1 \partial x_2} \hat{\Delta}x_2 + \cdots + \frac{\partial^2 f}{\partial x_1 \partial x_n} \hat{\Delta}x_n = \frac{df}{dx_1} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} \hat{\Delta}x_1 + \frac{\partial^2 f}{\partial x_2 \partial x_2} \hat{\Delta}x_2 + \cdots + \frac{\partial^2 f}{\partial x_2 \partial x_n} \hat{\Delta}x_n = \frac{df}{dx_2} \\ \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} \hat{\Delta}x_1 + \frac{\partial^2 f}{\partial x_n \partial x_2} \hat{\Delta}x_2 + \cdots + \frac{\partial^2 f}{\partial x_n \partial x_n} \hat{\Delta}x_n = \frac{df}{dx_n} \end{array} \right. \quad (2.3)$$

Dans le cas unidimensionnel, on prend la tangente de f' afin d'arriver au point minimal de f , qui correspond à la racine de f' . Pour trouver celle-ci on a besoin de prendre la dérivée seconde f'' . Or pour des fonctions multidimensionnelles $\mathbb{R}^n \rightarrow \mathbb{R}$, il y aura n fonctions dérivées partielles, toujours du type $\mathbb{R}^n \rightarrow \mathbb{R}$. Par conséquent il y aura n^2 fonctions dérivées secondes partielles. Ceci explique la nécessité de résoudre un système d'équations $n \times n$.

2.2.3 Analyse procustéenne généralisée

Le problème de trouver une transformation rigide qui minimise une distance d'un ensemble de couples de points donnés $(x_{1,i}, x_{2,i})$ peut être résolu par *analyse procustéenne*. L'*analyse procustéenne généralisée* ("GPA") permet d'avoir au lieu de couples des k -uplets $(x_{1,i}, x_{2,i}, \dots, x_{k,i})$, choisis parmi k ensembles de points. En général, l'*analyse procustéenne* sert à aligner des *formes*, ce qui peut être vu comme l'information géométrique qui reste d'un objet après avoir éliminé sa position, orientation et échelle. La méthode est utilisée dans beaucoup de domaines, notamment en statistique.

Soit n le nombre de couples. On encode la liste des k -uplets via k matrices modèle $n \times 3 \mathbf{X}_j$, dont les colonnes contiennent les 3 coordonnées des j -ième éléments des tuples. Ces matrices contiennent donc tous approximativement les mêmes points, hormis une transformation rigide.

On définit la *cible* \mathbf{K} comme

$$\mathbf{K} = \frac{1}{k} \sum_{i=1}^k \mathbf{X}'_j \quad (2.4)$$

La matrice \mathbf{X}'_j , encore inconnue, correspond à la version transformée \mathbf{X}_j .

Le principe de analyse procustéenne généralisée est le suivant : D'abord \mathbf{K} est initialisé à une valeur donnée. Pour toute matrice modèle \mathbf{X}_j , on calcule via une solution directe \mathbf{X}'_j qui est aligné le mieux avec le centroid. Par après, on recalcule \mathbf{K} par la formule 2.4. Le processus est itéré jusqu'à ce que \mathbf{K} se stabilise. [Toldo *et al.*, 2010]

Donc on doit trouver, pour tous les k matrices \mathbf{X}_j , une matrice orthogonale de rotation \mathbf{R}_j , un vecteur de translation \vec{t} et un facteur de homothétie c qui minimise

$$\|\mathbf{X}'_j\| = \|c \mathbf{X}_j \mathbf{R}_j + j \vec{t}^T - \mathbf{K}\| \quad (2.5)$$

où j est un vecteur unitaire 3×1 , et $\|\mathbf{X}\|$ représente la norme de Frobenius.⁹

Une solution de l'*analyse procustéenne* qui permet de trouver \mathbf{X}'_j et la procédure qui permet de calculer ce résultat sont données dans [Schönemann & Carroll, 1970] et [Schönemann, 1966].

2.3 Recalage

Afin de créer un modèle complet, on doit fusionner plusieurs scans qui sont prises de différents points de vues et qui sont tous limités par le champ de vision du scanner. Ces différents nuages de points ont

8. $\hat{\Delta}x_k$ est une approximation d'un vecteur Δx_k qui résoudrait le problème en une étape.

9. La norme de Frobenius d'une matrice $m \times n$ \mathbf{X} est défini comme racine carrée de la somme des carrés de tous les éléments de \mathbf{X} . On a :

$$\|\mathbf{X}\| = \sqrt{\sum_{i=1}^m \sum_{j=1}^n \|x_{i,j}\|^2} = \text{tr}(\mathbf{X}^* \mathbf{X}) \quad (2.6)$$

tous des repères différents, relatifs à la position et à l'orientation du scanner, et possiblement ayant des échelles différents.

Puisqu'il s'agit toujours de repères cartésiens orthonormaux, on peut mettre un nuage de points dans un autre repère en appliquant une matrice de transformation rigide à tous les points du nuage. Une telle transformation consiste en une translation, rotation, et dans certains cas un facteur de redimensionnement¹⁰. Dès que les nuages de points sont tous dans le même repère, on peut les fusionner par simple concaténation des points. En pratique, des post- et pré-traitements devront être effectués pour ajuster les densités et la distribution des points, corriger des erreurs dans les données brutes comme des points *outliers*, et autres.

Par recalage, on entend le procédé de trouver ces transformations rigides, de manière automatisé ou semi-automatisé. D'une part on a les algorithmes de recalage approximatif, donc le but est d'analyser le contenu des nuages de points et d'en identifier des régions qui représentent la même partie du modèle afin de trouver un recalage approximatif. Ensuite, ce recalage initial peut être raffiné par un algorithme de recalage précis pour bien aligner les surfaces des nuages de points. Généralement ICP est utilisé pour cela. Sachant que les nuages de points sont déjà alignés de manière à ce que les points qui représentent les mêmes parties du modèle sont proches l'un de l'autre, ICP procède en minimisant les distances entre ces points.

On peut distinguer entre les algorithmes locaux (séquentiels) qui opèrent sur deux nuages de points à la fois, et les algorithmes globaux (simultanés) qui effectuent un recalage de plusieurs nuages de points. Cette dernière méthode est plus complexe, mais a l'avantage qu'il n'y a pas d'erreur de recalage qui s'accumule lors des recalages deux-à-deux. Ces algorithmes tentent au lieu de distribuer l'erreur uniformément parmi les nuages.

La méthode préférée pour déterminer un recalage approximatif dépend surtout du type et de la forme de l'objet numérisé, et des données brutes dont on dispose. Certains algorithmes procèdent en identifiant des droites [Lichtenstein, 2011] ou des plans [Dold & Brenner, 2006] dans le modèle, et sont donc plus appropriés pour par exemple des façades simples de bâtiments. Si les scans ont des zones de chevauchement assez grandes, il peut être possible d'en déduire le recalage de façon complètement automatisée. D'autres méthodes peuvent être utilisées avec des données collectionnées additionnelles, comme des photos prises par le même point de vue que le scanner [Tournas & Tsakiri, 2009] ou des markers placés sur l'objet avant la numérisation [Matiukas & Miniotas, 2011]. Par des algorithmes de détection de zones d'intérêt [Tuytelaars & Mikolajczyk, 2007] sur des images ou directement sur les nuages de points, on peut alors identifier des points communs dans les différents scans. Aussi les informations attribuées aux points, comme une couleur, une mesure de température, ou autres, peuvent servir en tant que dimension en plus des coordonnées 3D pour déterminer leur distance. Une première estimation de la pose des scans peut être issue de sensors odométriques ou GPS du scanner.

Dans les pages suivantes certains algorithmes et méthodes seront décrits en détail, avec leurs bases mathématiques. Ils pourront possiblement être utilisés ou développées pour le projet de mémoire.

2.4 Iterative Closest Point

L'algorithme ICP¹¹ est considéré comme la méthode standard pour l'alignement précis de deux nuages de points, après qu'un recalage approximatif initial a déjà été fait. Parmi les deux nuages de points pris comme entrée, l'un (*référence*) reste fixé, tandis que l'autre (*source*) est transformé par translation et rotation. L'algorithme procède en itérativement raffinant la transformation à chaque étape, de manière à ce que la distance entre les deux nuages de points soit minimisée.

Plus précisément, pour tout point s_i de l'ensemble *source*, on choisit un point c_i de l'ensemble *cible* pour lequel on estime qu'il représente plus au moins la même position dans le nuage de points. Comme les deux nuages de points sont déjà recalés approximativement, une méthode courante est de choisir le point le plus proche, avec la transformation actuelle. L'algorithme nécessite donc que les densités de

10. Pour pouvoir exprimer une transformation rigide (translation, rotation, redimensionnement) d'un point tridimensionnel \vec{p} en utilisant uniquement une multiplication matricielle $\vec{x}' = \mathbf{M} \vec{x}$, des *coordonnées homogènes* peuvent être utilisées pour la matrice de transformation et les points.

11. itératif point le plus proche

points des deux ensembles sont similaires. Ensuite une transformation est appliquée aux points s_i qui minimise une métrique donnée de l'erreur de recalage.

La *cible* ne doit pas forcément être donnée sous forme de nuage de points, mais peut par exemple être une surface mathématique continue donnée sous forme paramétrique $(x(\vec{t}), y(\vec{t}), z(\vec{t}))$ ou implicite $g(x, y, z) = 0$. On peut alors calculer numériquement les coordonnées d'un point c_i correspondant à chaque point s_i . [Besl & McKay, 1992]

Quand les vecteurs normaux sur la cible sont connus, il est possible d'utiliser une variante *point-to-plane*¹² : Au lieu de minimiser une métrique des distances entre deux points, les écarts entre les points source et les plans tangents aux points cible sont minimisés. Cette méthode a généralement une vitesse de convergence supérieur et peut mener à des meilleurs résultats. Les algorithmes utilisés à chaque itération pour calculer la minimisation ont cependant un coût numérique plus élevé. L'ICP *point-to-plane* est une alternative intéressante puisque les surfaces des objets scannés sont souvent localement planaires. [Low, 2004]

Plusieurs variantes de l'algorithme ont été développées qui se distinguent par le choix des couples de points, des poids associés aux couples, le rejet de certains couples, et la façon comment la métrique d'erreur est définie et minimisée. [Rusinkiewicz & Levoy, 2001]

Pour la minimisation, on doit calculer une transformation rigide (généralement translation et rotation, donc 6 degrés de liberté) transforme les points s_i . Il existe plusieurs méthodes déterministes qui nécessitent pas de prendre plusieurs échantillons dans l'espace des transformations rigides. Pour le calcul de la rotation on peut par exemple utiliser la décomposition en valeurs singulières, par une représentation sous forme de matrice orthonormale ou sous forme de quaternion [Horn, 1987].

En général, un algorithme ICP procède selon les étapes suivantes :

1. Sélectionner des sous-ensemble de points s_i et c_i à utiliser (possiblement tous les points).
2. Former les couples (s_i, c_i) .
3. Associer des poids aux couples. (optionnel)
4. Rejeter certains couples. (optionnel)
5. Calculer une métrique d'erreur à partir des couples. Par exemple $\sum_i \|s_i - c_i\|^2$.
6. Appliquer une transformation rigide aux points s_i qui minimise l'erreur.
7. Si l'erreur est inférieur à une valeur de seuil donnée, arrêter. Sinon, continuer avec 1 ou 2.

2.4.1 Calcul de la transformation par quaternions

Une manière de calculer la translation et rotation qui est appliquée à chaque itération à l'ensemble de points *source* est décrite dans [Besl & McKay, 1992] et [Horn, 1987]. La procédure prend uniquement les n couples de points (\vec{s}_i, \vec{c}_i) comme entrée, où \vec{c}_i est un point de l'ensemble *cible* C , et \vec{s}_i un point de l'ensemble *source* S . On suppose que les transformations données par les itérations précédentes ont déjà été appliquées aux points s_i . La méthode reste valable pour tout choix de points, et C ne doit pas forcément être donné sous forme d'ensemble de points, mais ses points \vec{c}_i peuvent aussi par exemple être calculés à partir d'une surface paramétrique.

Le but de trouver \vec{q}_T et \dot{q}_R qui minimisent le carré des distances des couples $f(\vec{q})$.

$$f(\vec{q}) = \frac{1}{n} \sum_{i=1}^n \|\vec{c}_i - \mathbf{R}\vec{s}_i - \vec{q}_T\|^2 \quad (2.7)$$

$\vec{q}_T = (q_4 q_5 q_6)^T$ donne la translation, et \mathbf{R} la matrice de rotation 3x3.

On cherche une transformation de la forme $\vec{t}' = \mathbf{R}(\vec{s}') + \vec{q}_T$, donc une rotation, suivie par une translation à appliquer aux points s_i , qui minimise les distances des points transformés t_i aux points c_i correspondants dans le nuage de points *cible*. Car on ne peut pas assumer que les deux nuages S et C auront exactement la même distribution de points, les erreurs $e_i = \vec{c}_i - \vec{t}_i$ ne peuvent en général pas devenir 0. On minimise la somme des carrés des erreurs $\sum \|e_i\|^2$. Le développement décrit en [Horn, 1987] inclut aussi un facteur de redimensionnement, mais pour ICP on suppose généralement que les deux nuages de points ont déjà

12. au lieu de *point-to-point*

la même échelle. Un redimensionnement durant les itérations introduirait des problèmes avec le choix des couples de points.

Pour le développement suivant, on prend les coordonnées des points \vec{s}_i et \vec{c}_i , relativement au centres de masse des ensembles. On définit

$$\vec{\mu}_s = \frac{1}{n} \sum_i \vec{s}_i \quad \text{et} \quad \vec{\mu}_c = \frac{1}{n} \sum_i \vec{c}_i \quad (2.8)$$

Et pour tout i , on pose $\vec{s}'_i = \vec{s}_i - \vec{\mu}_s$ et $\vec{c}'_i = \vec{c}_i - \vec{\mu}_c$. Donc $\sum \vec{s}'_i = 0$ et $\sum \vec{c}'_i = 0$. L'erreur e_i peut être réécrit comme $e_i = \vec{c}'_i - \vec{t}'_i = \vec{c}'_i - \mathbf{R}(\vec{s}'_i) - \vec{q}'_T$. On peut montrer que la somme des carrés des erreurs sera minimale quand $\vec{q}'_T = 0$, et donc la translation à appliquer correspondra à $\vec{q}_T = \vec{q}'_T - \mathbf{R}(\vec{\mu}_s)$:

En fait, on a que $\vec{q}'_T = \vec{q}_T - \vec{\mu}_c + \mathbf{R}(\vec{\mu}_s)$. La somme des carrés peut être développée en

$$\sum_{i=1}^n \|\vec{c}'_i - \mathbf{R}(\vec{s}'_i)\|^2 - 2\vec{q}'_T \sum_{i=1}^n (\vec{c}'_i - \mathbf{R}(\vec{s}'_i)) + n\|\vec{q}'_T\|^2 \quad (2.9)$$

Le second terme sera toujours 0 puisque les coordonnées sont pris relativement aux centres de masse. Le premier ne dépend pas de \vec{q}'_T . Donc l'expression est minimisée par $\vec{q}'_T = 0 \iff \vec{q}_T = \vec{\mu}_c - \mathbf{R}(\vec{\mu}_s)$.

Pour trouver la rotation \mathbf{R} , on maximise

$$\sum_{i=1}^n \vec{c}'_i \cdot \mathbf{R}(\vec{s}'_i) \quad (2.10)$$

En effet, le produit scalaire de deux vecteurs est maximal quand ils ont la même direction. On cherche la rotation sous forme d'un quaternion \dot{q}_R .¹³

Une rotation peut être encodée dans un quaternion unitaire, c'est à dire pour lequel $q_0 \geq 0$ et $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$. \dot{q} est purement imaginaire si $q_0 = 0$. (Les trois autres composants peuvent être considérés comme termes imaginaires comme chez les nombres complexes.) On représente un vecteur sous forme de quaternion purement imaginaire, en mettant ses coordonnées dans q_1, q_2, q_3 . On peut montrer que si \dot{s} est purement imaginaire et \dot{q}_R est unitaire, alors $\dot{q}_R \dot{s} \dot{q}_R^*$ reste purement imaginaire, et peut donc représenter une opération sur un vecteur. \dot{q}_R^* est le conjugué de \dot{q}_R .

L'article [Horn, 1987] montre qu'en utilisant les propriétés des quaternions, on peut former à partir des ensembles de points \vec{s}_i et \vec{c}_i une matrice symétrique 4x4 \mathbf{N} telle que

$$\sum_{i=1}^n \vec{c}'_i \cdot \mathbf{R}(\vec{s}'_i) = \sum_{i=1}^n \vec{c}'_i \cdot (\dot{q}_R \vec{s}'_i \dot{q}_R^*) = \dot{q}_R \mathbf{N} \dot{q}_R^* \quad (2.12)$$

Il est aussi montré que le quaternion unitaire \dot{q}_R qui maximise l'expression est le vecteur propre correspondant à la valeur propre la plus grande de \mathbf{N} . Le calcul des valeurs propres nécessite la résolution d'une équation polynomiale du quatrième degré. Il existe des solutions de forme fermée pour ce problème.

Donc, cette méthode permet de déterminer la rotation et translation à appliquer à \mathbf{M} par une série de calculs, sans passer par des méthodes approximatives.

2.4.2 Variantes de l'algorithme

Comme mentionné, diverses variantes de l'ICP sont possibles. [Rusinkiewicz & Levoy, 2001] donne un comparatif des différentes possibilités.

13. Une matrice de rotation \mathbf{R} peut en être déduite à partir d'un quaternion $\dot{q} = (q_0 q_1 q_2 q_3)^T$ par la formule

$$\mathbf{R} = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) \\ 2(q_1 q_2 + q_0 q_3) & q_0^2 + q_2^2 - q_1^2 - q_3^2 & 2(q_2 q_3 - q_0 q_1) \\ 2(q_1 q_3 - q_0 q_2) & 2(q_2 q_3 + q_0 q_1) & q_0^2 + q_3^2 - q_1^2 - q_2^2 \end{bmatrix} \quad (2.11)$$

2.4.3 Version simultanée de l'ICP

Une approche qui permet de faire un recalage ICP *simultané* de plusieurs nuages de points en utilisant l'analyse procustéenne (c.f. section 2.2.3) généralisée est décrite dans [Toldo *et al.*, 2010].

Comme montré dans la section précédente, une solution du problème GPA est l'aligner les nuages de points à une *cible* \mathbf{K} générée artificiellement à partie de tous les nuages de points, au lieu de les aligner à un autre nuage de points donné comme entrée. Ceci est appliqué pour rendre l'ICP *simultané*, de manière à ce qu'il n'y a pas d'erreur qui s'accumule.

Pour le choix des tuples $(x_{1,i}, x_{2,i}, \dots, x_{k,i})$, l'article propose de choisir des tuples dans lesquelles tous les paires de points sont *mutuellement voisins les plus proches*. Deux points $\vec{a} \in A$ et $\vec{b} \in B$ sont considérés mutuellement voisins les plus proches si \vec{a} est le point le plus proche à \vec{b} dans A , et \vec{b} est le point le plus proche de \vec{a} dans B . On forme un graphe ayant comme sommets les points de tous les nuages, et les couples de sommets qui remplissent cette conditions sont connectés par des arrêtes. Les tuples à choisir correspondent aux sous-ensembles indépendants dans ce graphe. La solution de GPA est donnée qui fonctionne aussi si les tuples ne sont pas toujours complets (donc pour certains nuages de points il n'ont pas de point correspondant), et inclut la possibilité d'associer des poids aux tuples. Cela est nécessaire parce que pour en grand nombre de nuages de points à recaler, il n'y aura dans la plupart des cas seulement relativement peu de tuples complets.

La version initiale de la cible \mathbf{K} est formée par les centroids des tuples. L'algorithme GPA-ICP peut alors être formulé comme suit : [Toldo *et al.*, 2010]

Entrée : n nuages de points p_i

Sortie : n matrices de transformation \mathbf{M}_i

1. Initialiser \mathbf{M}_i par matrices d'identité
2. Trouver les tuples de points correspondants
3. Former la cible \mathbf{K} à partir des centroids des tuples
4. Estimer par GPA les transformations rigides qui alignent $\{p_i\}$ à la cible
5. Appliquer ces transformations aux points $\{p_i\}$, et ajouter les aux matrices \mathbf{M}_i
6. Répéter depuis 2, jusqu'à ce que \mathbf{K} est stabilisé (ou un nombre limite d'itérations a été dépassé)

Après l'exécution de cet algorithme, les n nuages de points sont recalés dans un repère commun qui n'est pas spécifié d'avance. Donc un recalage complémentaire est nécessaire pour mettre le nuage fusionné dans le repère désiré.

2.4.4 ICP généralisé

Pour ce mémoire, il sera nécessaire de trouver des méthodes qui permettent de recaler des nuages de points à densités de points différents. Une version *généralisée* du ICP est développée dans [Segal *et al.*, 2009] : On prend en compte que les nuages de points ne sont qu'une approximation des surfaces scannées, et on représente l'aspect de ces surfaces sous-jacentes via un modèle probabilistique.

Pour cet algorithme, l'étape de minimisation est altérée. Les autres étapes (choix des couples de points, etc.) peuvent toujours être implémentés via les différentes variantes proposées. On suppose qu'il existe des nuages de points $\hat{S} = \{\hat{s}_i\}$ et $\hat{C} = \{\hat{c}_i\}$ sous-jacents aux nuages de points source S et cible C donnés, qui sont tels qu'il existe une matrice de transformation rigide \mathbf{M}^* qui recalcera parfaitement \hat{S} et \hat{C} . Donc $\hat{c}_i = \mathbf{M}^* \hat{s}_i$ pour tout i .

\hat{S} et \hat{C} constituent une version idéalisée de la représentation de la surface par nuage de points. Avec S et C , un recalage parfait n'est généralement pas faisable.

On suppose aussi que les points de S et C sont des variables aléatoires tirés de \hat{S} et \hat{C} par une loi normale multidimensionnelle

$$s_i \sim \mathcal{N}(\hat{s}_i, \mathbf{C}_i^S) \quad \text{et} \quad c_i \sim \mathcal{N}(\hat{c}_i, \mathbf{C}_i^C) \quad (2.13)$$

où \mathbf{C}_i^S et \mathbf{C}_i^C sont des matrice de variance-covariance qu'on associe à chaque point de S et C . Ces matrices vont être paramétrisées pour décrire la surface autour de \hat{s}_i respectivement \hat{c}_i . Ils représentent les distribution de probabilité des positions de s_i et c_i .

La distance entre deux points pour une matrice de transformation \mathbf{M} est donnée par $d_i^{(\mathbf{M})} = c_i - \mathbf{M} s_i$. Le but sera de trouver \mathbf{M} qui minimise $\sum_i \|d_i^{(\mathbf{M})}\|^2$. Dans ce modèle, les $d_i^{(\mathbf{M})}$ sont eux-mêmes des variables aléatoires. Vu que a_i et b_i sont tirés d'une loi normale multidimensionnelle, la distribution de $d_i^{(\mathbf{M})}$ est

$$d_i^{(\mathbf{M})} \sim \mathcal{N}(\hat{c}_i - \mathbf{M} \hat{s}_i, \mathbf{C}_i^C + \mathbf{M} \mathbf{C}_i^S \mathbf{M}^T) \quad (2.14)$$

Par conséquent,

$$d_i^{(\mathbf{M}^*)} \sim \mathcal{N}(0, \mathbf{C}_i^C + (\mathbf{M}^*) \mathbf{C}_i^S (\mathbf{M}^*)^T) \quad (2.15)$$

La méthode d'estimation du maximum de vraisemblance est utilisée dont le principe est le suivant : On a n échantillons x_1, \dots, x_n issus d'une variable aléatoire X , et on cherche la fonction de densité $f^{(\theta^*)}$ de probabilité associée à X , en sachant qu'elle appartient à une famille de fonctions de densité de probabilité, paramétrisée par θ . Quand X est discrète, la fonction de masse $p^{(\theta^*)}$ est utilisée. Donc on calcule une estimation de θ^* à partir de x_1, \dots, x_n et f . Pour une valeur du paramètre θ donnée, et on supposant que les échantillons sont mutuellement indépendants, la *fonction de vraisemblance* peut être calculée par

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f^{(\theta)}(x_i) \quad (2.16)$$

Une estimation de θ^* peut alors être trouvée en maximisant L :

$$\hat{\theta} = \arg \max_{\theta} L(\theta; x_1, \dots, x_n) \quad (2.17)$$

On considère que $d_i^{(\mathbf{M})}$ sont des échantillons. En fixant \mathbf{C}_i^S et \mathbf{C}_i^C , a fonction de masse des échantillons $p^{\mathbf{M}}$ peut être calculée avec \mathbf{M} comme seule inconnue. Par une application récursive de l'estimation du maximum de vraisemblance, une valeur pour \mathbf{M} pourra être trouvée :

$$\mathbf{M} = \arg \max_{\mathbf{M}} \prod_{i=1}^n p^{\mathbf{M}}(d_i^{(\mathbf{M})}) = \arg \max_{\mathbf{M}} \sum_{i=1}^n \log(p^{\mathbf{M}}(d_i^{(\mathbf{M})})) \quad (2.18)$$

La somme des logarithmes peut être maximisée au lieu parque que la fonction \log est strictement croissante.

En utilisant 2.14, la formule peut être simplifiée en

$$\mathbf{M} = \arg \min_{\mathbf{M}} \sum_{i=1}^n (d_i^{(\mathbf{M})})^T (\mathbf{C}_i^C + \mathbf{M} \mathbf{C}_i^S \mathbf{M}^T)^{-1} d_i^{(\mathbf{M})} \quad (2.19)$$

C'est cela qui constitue l'amélioration du ICP généralisé par rapport au ICP standard : Au lieu de minimiser la métrique d'erreur $\sum_i \|d_i^{(\mathbf{M})}\|$, on minimise $\sum_i (d_i^{(\mathbf{M})})^T (\mathbf{C}_i^C + \mathbf{M} \mathbf{C}_i^S \mathbf{M}^T)^{-1} d_i^{(\mathbf{M})}$. On le recalcule plus directement les points S et P , mais les surfaces sous-jacentes. Les surfaces à recaler sont les mêmes (vu de différentes perspectives), mais les points S et P ne représentent en général pas exactement les mêmes positions sur la surface, contrairement aux "vrais" points \hat{S} et \hat{P} . Par cette méthode, on ne suppose plus cela, mais plutôt on introduit une incertitude sur la position des points donnés par rapport aux vrais points, qui est décrite par les matrices de variance-covariance \mathbf{C}_i^S et \mathbf{C}_i^C . Donc l'algorithme correspond à une approche *plane-to-plane*, augmentée par une distribution de probabilité sur les plane, de l'ICP.

Pour un point avec le premier vecteur directeur du repère orthonormal comme vecteur normal, la matrice

$$\begin{bmatrix} \epsilon & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.20)$$

avec ϵ un petit nombre, indique une grande certitude sur la position du point le long de la direction du vecteur normal, mais une petite certitude sur la position du point dans le plan tangent. Comme la surface est souvent localement planaire, ce plan est une approximation de la surface. Soit \mathbf{R}_{s_i} et \mathbf{R}_{c_i} les matrices de changement de base de ce dernier repère, vers un repère dont le premier vecteur directeur

est le vecteur normal du point s_i respectivement c_i , sans translation. \mathbf{C}_i^S et \mathbf{C}_i^C peuvent maintenant être calculées par

$$\mathbf{C}_i^S = \mathbf{R}_{s_i} \begin{bmatrix} \epsilon & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{R}_{s_i}^T \quad \text{et} \quad \mathbf{C}_i^C = \mathbf{R}_{c_i} \begin{bmatrix} \epsilon & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{R}_{c_i}^T \quad (2.21)$$

Il a été expérimentalement montré que la méthode génère des résultats supérieurs au ICP standard. [Segal *et al.*, 2009] Cependant, puisqu'elle suppose que les surfaces sont localement planaires, et que \mathbf{C}_i^S et \mathbf{C}_i^C donnent une bonne estimation de la forme de la surface.

L'article [Holz & Behnke, 2014] présente une manière de calculer les vecteurs normaux (et donc les matrices via 2.21), dans le contexte de nuages de points à densités non-uniformes pris par un scanner 3D composé d'un télémètre laser 2D tournant. Le télémètre capture seulement des *scan lines* 2D, c'est à dire des valeurs de profondeur de points dans un plan (vertical) sortant du scanner, et non pas dans un champ de vision tridimensionnel. Durant les rotations, plusieurs scan lines sont pris dans différentes. La distance de points voisins en direction horizontale (donc provenant de deux scan lines) sera plus grande à celle en direction verticale, sur une même scan line.

Il ne suffit donc plus de comparer les points voisins pour avoir une estimation du vecteur normal. La méthode proposée procède en créant un maillage à quadrilatères, formée par points voisins sur une scan line, connectés les deux points correspondants de la scan line voisine. Ces derniers points ne seraient pas inclus par une recherche de points voisins classique. Après filtrage du maillage, les normaux sont maintenant estimés en considérant les points adjacents sur le maillage, pour un rayon de recherche donné.

2.5 4-Points Congruent Sets

La méthode 4-Points Congruent Sets¹⁴ (4PCS) est une approche qui permet de recalier deux nuages de points avec du bruit (i.e. outliers), relativement peu d'information commune, et sans avoir besoin d'un recalage initial. Elle est basée sur une variante du RANSAC appliquée sur des échantillons bien choisis.

Le principe de RANSAC est de trouver par recherche aléatoire un nombre minimal d'échantillons à partir desquelles une approximation assez bonne d'un modèle peut être trouvée, et puis de raffiner cette approximation en utilisant uniquement les échantillons (inliers) assez proches. Pour le problème du recalage de deux nuages de points S et C , le modèle à chercher est une matrice de transformation \mathbf{M} qui recalcule assez bien les deux nuages. Une telle matrice peut être univoquement définie par deux triplets de points sans les repères de S et C respectivement. Une méthode pour en déduire \mathbf{M} est donnée en [Horn, 1987]. \mathbf{M} représente le recalage correct si les points des deux triplets représentent toujours les mêmes positions dans les deux nuages de points.

Une manière d'appliquer RANSAC est donc de choisir comme échantillons aléatoirement ces deux triplets de points, et de calculer et vérifier les matrices de transformation correspondantes. Le consensus set est l'ensemble de points de $s \in S$ pour lesquels $\min_{c \in C} \|s - c\| < \epsilon$. On peut raffiner le recalage par exemple avec ICP. La plupart de ces échantillons ne formeront pas un recalage correct. Une telle méthode aurait une complexité d'au moins $O(n^3 \log n)$. Pour la rendre utilisable, il faut donc altérer la façon de choisir des échantillons.

La méthode 4PCS utilise comme échantillon deux 4-points (i.e. quadruples de points) prises de S et de C . Ceci permettra d'arriver à une complexité quadratique au lieu de cubique. A partir de S , on choisit aléatoirement quatre points qui sont coplanaires. Ces quatre points sont appelés la *base*. En pratique, une coplanarité approximative est suffisante. Pour le deuxième quadruple, on choisit ceux qui sont congruents avec le premier, c'est à dire que le premier peut être transformé en le deuxième par une transformation rigide. L'algorithme fonctionne selon la procédure suivante :

1. Sélectionner aléatoirement base \dot{b} (quadruple de points coplanaires) de S .
2. Trouver l'ensemble U des quadruples congruents à la base.
3. Pour tout $\dot{u}_i \in U$, calculer la matrice de transformation \mathbf{M}_i qui aligne \dot{u}_i à \dot{b} . La solution à forme fermée développée en 2.4.1 peut être utilisée pour aligner deux ensembles de 4 points. Le consensus

14. ensemble congruents de 4-points

set $S_i \cup S$ est formé par les points de la source qui sont après la transformation \mathbf{M}_i assez bien recalés à la cible :

$$S_i = \{s \in S : \min_{c \in C} \|c - \mathbf{M}_i s\| < \delta\} \quad (2.22)$$

4. On retient le consensus set S_k et la matrice de transformation \mathbf{M}_k correspondante pour lequel $|S_i|$ était maximal. \mathbf{M}_k correspond à la meilleure transformation qui a pu être trouvée avec la base \vec{b} .
5. Un compteur est incrémenté. S'il n'a pas encore dépassé un nombre fixé L , on continue avec l'étape 1. Sinon, l'algorithme se termine. A la fin on garde la matrice \mathbf{M}_k pour laquelle $|S_k|$ était maximal.

L'opération fondamentale de l'algorithme est la recherche des quadruplets $\dot{u} = (\vec{u}_1, \vec{u}_2, \vec{u}_3, \vec{u}_4)$ congruents à une base $\vec{b} = (\vec{b}_1, \vec{b}_2, \vec{b}_3, \vec{b}_4)$ donnée. Les points de la base sont coplanaires, par définition. Il y aura toujours deux paires dans la base qui forment des droites qui s'intersectent. Sans perte de généralité, soit $\vec{u}_1 \vec{u}_2$ et $\vec{u}_3 \vec{u}_4$ ces droites, et \vec{e} leur point d'intersection. Il a été montré que les deux nombres

$$r_1 = \frac{\|\vec{u}_1 - \vec{e}\|}{\|\vec{u}_1 - \vec{u}_2\|} \quad \text{et} \quad r_2 = \frac{\|\vec{u}_3 - \vec{e}\|}{\|\vec{u}_3 - \vec{u}_4\|} \quad (2.23)$$

sont invariants par rapport à la transformation affine de \vec{b} . De plus, ils changent toujours quand une transformation non-affine est appliquée. Donc pour trouver des candidats pour \dot{u} , il suffit de chercher les quadruplets qui produisent les mêmes valeurs r_1 et r_2 .

On considère tous des paires de points $(\vec{s}_1, \vec{s}_2) \in S$, on faisant l'hypothèse qu'ils appartiennent à un des quadruplets \dot{u} cherchés, et qu'ils sont congruents à soit (\vec{b}_1, \vec{b}_2) ou (\vec{b}_3, \vec{b}_4) . Donc ils forment une des droites qui passent par leur équivalent de \vec{e} . Des paires sont cherchées dans les deux ordres, c'est à dire on considère (\vec{s}_1, \vec{s}_2) et (\vec{s}_2, \vec{s}_1) séparément. Puisqu'on ne sait pas quelles des deux droites ils représentent, il y a deux possibilités \vec{e}_1 et \vec{e}_2 pour la position du point d'intersection de ce quadruplet :

$$\vec{e}_1 = \vec{s}_1 = r_1(\vec{s}_2 - \vec{s}_1) \quad \text{ou} \quad \vec{e}_2 = \vec{s}_1 = r_2(\vec{s}_2 - \vec{s}_1) \quad (2.24)$$

Si deux paires de points (\vec{s}_1', \vec{s}_2') et $(\vec{s}_1'', \vec{s}_2'')$ ont pu être trouvées tels que $\vec{e}_1' = \vec{e}_2''$, alors il est probable qu'ils forment un des quadruplets congruents recherchés.

L'algorithme procède par filtrer les quadruplets qui ne sont pas congruents avec la base. Dans l'implémentation donnée par [Aiger et al., 2008], l'algorithme considère seulement les paires (\vec{s}_1, \vec{s}_2) dont la distance correspond approximativement à $\|\vec{b}_1 - \vec{b}_2\|$ ou $\|\vec{b}_3 - \vec{b}_4\|$. Les points \vec{e} sont mis sur une structure *range tree*, ce qui permet de trouver des partenaires coïncidents e' de façon efficace. La procédure a une complexité dans $O(n^2 + k)$, où k est le nombre de quadruplets \dot{u} qui seront trouvés. (L'algorithme est *output-sensitive*.)

Pour choisir une base \vec{b} , l'algorithme choisit d'abord aléatoirement trois points de C . On préfère des points à distance plutôt élevée, afin que la base ne soit pas perturbée par des imprécisions locales. On cherche ensuite le quatrième point de manière à ce qu'il soit coplanaire avec les trois autres.

2.6 Normal Distributions Transform

Dans le contexte de la robotique, la méthode NDT¹⁵ a été développée pour aligner deux nuages de points 2D qui représente les distances d'objets autour d'un robot mobile qui déplace sur un plan horizontal. [Biber, 2003] Ce recalage permet de construire une carte 2D des environs du robot, qui consiste en une translation 2D et une rotation sur une axe, dénotées par les paramètres $\theta = (t_x, t_y, \phi)$. La méthode est par exemple utilisé pour mesurer la position précise du robot en partant par une estimation calculée par des données odométriques.

L'idée est d'au lieu de recaler directement des points comme ICP, on recalcule leur *transformée en distribution normale*, une représentation continue et dérivable du nuage de points par morceaux du nuage de points qui consiste en la distribution de probabilité qu'un point soit à une position. La NDT est calculée à partir d'un nuage de points cible $C = \{\vec{c}_i\}$ de manière suivante. On commence par subdiviser l'espace 2D en une grille régulière avec une taille donnée des cellules. Soit C_j l'ensemble des points qui

15. transformée en distribution normale

se trouvent dans la cellule j . Pour tous les cellules qui contiennent au moins 3 points, on peut modéliser la probabilité $p(\vec{x})$ d'y trouver un point à la position \vec{x} :

Soit $\vec{\mu}_j = \frac{1}{n_j} \sum_i \vec{c}_{j,i}$ la moyenne des points dans la cellule. La matrice de variance-covariance qui représente la distributions des points C_j est calculée par $\mathbf{C}_j = \frac{1}{n_j-1} \sum_i (\vec{c}_{j,i} - \vec{\mu}_j)(\vec{c}_{j,i} - \vec{\mu}_j)^T$. La matrice exprime une distribution de Gauss en deux dimensions, ajusté sur la distributions des points. $p(\vec{x})$ peut être calculé par la fonction caractéristique :

$$p(\vec{x}) = \exp \left(-\frac{(\vec{x} - \vec{\mu}_j)^T \Sigma^{-1} (\vec{x} - \vec{\mu}_j)}{2} \right) \quad (2.25)$$

Similairement au ICP, l'algorithme procède en raffinant itérativement une estimation de la transformation (t_x, t_y, ϕ) cherchée. La fonction de transformation T est exprimée par

$$T_\theta : \vec{p} \mapsto \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix} \begin{bmatrix} p_x \\ p_y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (2.26)$$

Les étapes suivies par l'algorithme pour recalier une nuage de points source S à un nuage de points cible C sont les suivantes :

1. Créer la NDT pour C , et initialiser une estimation des paramètre de recalage $\theta = (t_x, t_y, \phi)$ (possiblement en utilisant des données odométriques enregistrées par le robot).
2. Mettre les points S dans le repère de C , en appliquant la transformation T_θ .
3. Pour tous les points transformés $\vec{s}'_i = T(\vec{s}_i)$, trouver sa cellule j et les $\mathbf{C}_j, \vec{\mu}_j$ correspondants.
4. On sait que $p(\vec{x})$ exprime la probabilité qu'un point de C se trouve à la position \vec{x} . La somme $\sum_i p(\vec{x}_i)$ est donc maximale si les \vec{x}_i représentent tous des positions de points, par exemple si $\{x_i\} = C$, donc si le recalage est parfait. L'algorithme va donc maximiser une fonction *score* des paramètres θ donnée par

$$\text{score}(\theta) = \sum_i \exp \left(-\frac{(\vec{s}'_i - \vec{\mu}_j)^T \mathbf{C}_j^{-1} (\vec{s}'_i - \vec{\mu}_j)}{2} \right) \quad (2.27)$$

5. Calculer des nouveaux paramètres de transformation θ par maximisation de la fonction *score*.
6. Tant que le recalage n'est pas encore assez bon, répéter avec l'étape 2.

Comme montré dans la section 2.2.2, on peut utiliser la méthode de Newton multidimensionnelle pour ce problème de maximisation : On minimise la fonction opposée $-\text{score}$. Puisqu'il y a trois paramètres (t_x, t_y, ϕ) et la fonction *score* calcule un réel, on est dans le cas $\mathbb{R}^3 \rightarrow \mathbb{R}$. Pour chaque itération de l'algorithme de Newton résoud le système

$$[\mathbf{H} \text{score}(\theta)] \Delta \theta = \nabla \text{score}(\theta) \quad (2.28)$$

en utilisant la définition de la fonction *score*.

Une façon simple d'étendre la méthode NDT pour utilisation avec des nuages de points 3D, mais toujours avec des transformations qui restent sur un plan horizontal, est décrite en [Dold *et al.*, 2007]. On extrait à partir des nuages de points 3D source et cible plusieurs couches horizontales 2D à différentes hauteurs. Le recalage NDT est fait sur toutes ces sous-nuages 2D en même temps, en utilisant toujours la même grille de cellules. Pour la fonction *score*, on prend la somme des différentes fonctions *score* appartenant aux sous-nuages.

2.7 Extended Gaussian Images

Si on cherche à recalier des nuages de points seulement au niveau de la rotation, sans avoir une estimation initiale du recalage, une possibilité intéressante pourrait être d'utiliser leur Extended Gaussian Images¹⁶ (EGI).

¹⁶ 16. Image gaussienne étendue

On commence par segmentiser les nuages de points en différentes surfaces planaires, pour lesquelles on calcule le vecteur normal. On disposera donc pour chaque nuage de point un ensemble de vecteurs normaux, qui peuvent être décrits par des coordonnées sphériques (θ, ϕ) . L'idée est de trouver une rotation 3D qui recalcule ces ensembles $\{(\theta_i, \phi_i)\}$, sachant que les angles sont modulo 2π .

Cette projection des vecteurs normaux d'un objet sur une sphère unitaire est appelée l'Extended Gaussian Image de l'objet. [Horn, 1984] Il a été montré que pour des objets convexes, l'EGI décrit univoquement l'objet, malgré la perte d'information. Des méthodes existent pour reconstruire un polyèdre convexe à partir de son EGI. Évidemment, l'EGI est invariant par rapport à une translation de l'objet, et par rapport à son échelle.

Pour pouvoir représenter un EGI numériquement, la surface de la sphère sera divisée en des petites cellules, et pour chacune d'elles on compte le nombre de normales $\vec{n} = (\theta, \phi)$ qui tombent dans cette cellule. Ce problème l'est pas évident puisqu'on essaye de passer d'une topologie sphérique à une sérialisation numérique. Un bon pavage de la sphère satisferait les conditions suivantes :

- Les cellules ont les mêmes formes, les mêmes aires, et des formes régulières compactes
- La division doit être assez fine pour produire une bonne résolution angulaire
- Il y a des rotations avec lesquelles des cellules avant et après la rotation coïncideront

Une manière simple est de diviser la sphère dans des intervalles de longitude et de latitude. Ce pavage peut être rendu arbitrairement fin, mais les cellules n'auront pas la même taille. Proche aux pôles, ils deviendront plus petits à cause de la convergence des méridiens. On pourrait améliorer le pavage en variant les intervalles. Aussi avec cette méthode, seulement les rotations autour de l'axe du globe peuvent faire coïncider les cellules. Pour arriver à des cellules de mêmes formes et aires, la sphère peut au lieu être approximée par un solide de Platon. Celui avec le plus de faces est l'icosaèdre régulier, ayant 20 faces triangulaires, ce qui est beaucoup trop peu pour un recalage précis.

Comme compromis, une possibilité est de commencer avec un dodécaèdre régulier (12 faces pentagonales), et de subdiviser ses faces en cinq triangles égaux pour obtenir un *pentakis dodecahedron* ayant 60 faces triangulaires avec la même aire. La subdivision peut ensuite être raffinée jusqu'au niveau désiré par la *division géodésique* : On divise uniformément les côtés de chaque triangle (faces) en f sections. Après avoir appliqué une triangulation, chaque ancienne face a maintenant été subdivisée en f^2 nouvelles faces triangulaires.

Lors de la création de la représentation numérique, le test suivant peut être utilisé pour vérifier si un vecteur normal \vec{n} appartient à la cellule triangulaire avec les sommets $\vec{v}_1, \vec{v}_2, \vec{v}_3$, tous donnés en coordonnées cartésiennes 3D : On calcule

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} v_{1,x} & v_{2,x} & v_{3,x} \\ v_{1,y} & v_{2,y} & v_{3,y} \\ v_{1,z} & v_{2,z} & v_{3,z} \end{bmatrix}^{-1} \times \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} \quad (2.29)$$

Si $\lambda_1, \lambda_2, \lambda_3$ sont tous positifs, alors \vec{n} appartient à cette cellule.

2.7.1 Application sur le recalage de nuages de points

L'article [Dold, 2005] décrit une implémentation d'un recalage de la rotation de deux nuages de points en utilisant les EGI. L'algorithme utilise plusieurs EGI à différents niveaux de pavage. On commence par un recalage brut en utilisant des EGIs avec un petit nombre de faces, puis on raffine itérativement la solution moyennant des EGI plus fins. Les EGI sont construits tel que le premier niveau de pavage produit 20 faces, et le deuxième en produit 80. Donc une solution approximative pourra être trouvée rapidement.

Une rotation 3D peut être représentée par 3 angles d'Euler (ω, ϕ, κ) , donc il y a degrés niveaux de liberté pour le recalage, tous modulo 2π .¹⁷ Pour la première itération, tous les triplets possibles sont considérées comme solution potentielles. Durant les itérations qui suivent, on n'essaie plus que des angles de rotations dans un certain rayon autour de l'ancienne solution trouvée, qui diminue progressivement, avec la taille d'incrément des triplets à tester.

17. La rotation qui correspond à un triplet d'angles d'Euler dépend de la convention utilisée. Elle définit l'ordre dans lequel on applique les trois rotations axiales, et si on applique la deuxième et troisième rotation dans le repère déjà transformé (*intrinsèque*) ou original (*extrinsèque*).

Durant chaque test, tous les vecteurs normaux $\{\vec{n}\}$ issue du nuage de points source sont transformés et l'EGI du niveau actuel en est créé. La représentation numérique de cet EGI est alors comparé avec celle du EGI de même niveau du le nuage de points cible. La problème principal est de définir la méthode selon laquelle la correspondance de deux représentations est évaluée. Soit $N_S(i)$ respectivement $N_C(i)$ les ensembles de vecteurs normaux associés à la i -ième face du EGI source ou cible. Trois mesures sont présentées dans l'article :

- Somme des différences absolues : $\sum_i |\|N_S(i)\| - \|N_C(i)\||$
- Somme des carrés des différences : $\sum_i (\|N_S(i)\| - \|N_C(i)\|)^2$
- En considérant N_S et N_C en tant que variables aléatoires, leur correlation : $\frac{\text{cov}(N_S, N_C)}{\|N_S\| \|N_C\|}$

Il est montré expérimentalement que ces deux dernières mesures peuvent donner de bons résultats. Pour la somme des différences absolues, les résultats se rapprochent trop pour des niveaux de pavage supérieurs. L'article propose aussi d'inclure l'aire des surfaces planaires des nuages de points, en tant que pondération pour les vecteur normaux.

2.8 Identification et correspondance de droites

2.9 Identification et correspondance de plans

2.10 DéTECTEURS de zones d'intérêt

2.11 Recalage via des photos

2.12 Photogrammétrie

La photogrammétrie peut être définie comme la science d'extraire algorithmiquement des informations sur un environnement tridimensionnel à partir de photos. En particulier, une opération photogrammétrique fondamentale est de retrouver les positions 3D correspondants à des points sur la photo.

Trois systèmes de coordonnées vont intervenir : Le *repère objet* 3D avec lequel on définit des positions dans l'espace, le *repère caméra* 3D, dont l'origine est la caméra, et le *repère image* 2D qui correspond aux pixels de la photo. Pour une caméra perspective idéalisée il est assez simple de déduire à partie d'un point dans le repère objet sa position correspondante dans de repère image. Pour cela on aura besoin des *paramètres extrinsèques* et des *paramètres intrinsèques* de la caméra¹⁸. L'opération inverse est beaucoup plus compliquée car les information de profondeur sont essentiellement perdues par la projection perspective 3D en 2D.

Les *paramètres extrinsèques* sont la position et l'orientation de la caméra par rapport à l'environnement. Dans un repère objet donné, ils peuvent être définis par les coordonnées (x_0, y_0, z_0) et les angles d'Euler (ω, ϕ, κ) . Ils permettent de passer du repère objet au repère caméra et vice versa, par une simple transformation affine. Les *paramètres intrinsèques*, d'autre part, définissent la projection perspective appliquée par la caméra. Ils dépendent de ses caractéristiques optiques et géométriques.

2.12.1 Optique

Chaque point sur un objet peut refléter des radiations de lumière provenant de différentes directions. Un objet devient visible pour un système optique quand ces radiations reflétées touchent la surface photo-sensitive du système, comme par exemple le capteur d'une caméra numérique, le film dans une caméra analogue, ou la rétine de l'œil animal.

Pour obtenir une image nette d'un objet, le système optique doit être tel que tous les rayons de lumière reflétés par un même point de l'objet convergent en un même point sur la surface photo-sensitive. Sinon, ils seront distribués dans un certain rayon sur la surface, et on obtient une image floue. Un système simple qui permet d'obtenir ceci est la *pin-hole camera* : L'aperture optique est rendu assez petit pour que idéalement qu'un seul rayon par point d'objet peut entrer. Ceci permet de captures des photos nettes,

18. Aussi appelés l'orientation externe et interne de la caméra.

peut en pratique, des caméras de ce type ont besoin d'un temps d'illumination très grand et produisent des photos de basse qualité. Ce modèle est cependant utilisé pour modéliser une caméra. Des systèmes optiques réels utilisent au lieu une lentille qui fait converger tous les rayons provenant d'un même point. La distance de ce point de convergence à la lentille varie en fonction de la distance du point sur l'objet. Donc pour pouvoir acquérir une image nette, il est nécessaire d'ajuster soit la distance lentille-capteur (caméra), ou la forme de la lentille (œil) en fonction de la distance de l'objet observé. En pratique, des lentilles introduisent des distorsions supplémentaires et d'autres effets optiques qui doivent être pris en compte pour effectuer des mesures précises sur base de photos. [Trucco & Verri, 1998]

La modélisation de la lumière en tant qu'ondes montre qu'un rayon lumineux correspond à une onde, qui peut être décrit comme une superposition d'un ensemble d'ondes sinusoïdes sur un continuum de fréquences. Quand un rayon est reflété, sa forme d'onde est altérée car certaines de ces fréquences sont absorbées. Trois types de cellules coniques existent dans la rétine humaine, qui sont sensibles à des histogrammes de fréquences différents dans le spectre visuel. Ceux-ci permettent la vision des couleurs rouge, vert et bleu. La combinaison de valeurs d'intensité de ces trois couleurs fondamentales donne lieu à la perception d'une couleur. Les capteurs d'une caméra numérique fonctionnent selon le même principe, et produisent des valeurs de couleur sous forme d'un triplet RGB. Comme ces valeurs de couleur ne sont pas directement une propriété de l'objet photographié mais dépendent des conditions lumineuses ambiantes et du système optique, des corrections radiométriques peuvent être nécessaires pour égaliser les couleurs si on veut fusionner des informations tirées de plusieurs photos d'un objet.

2.12.2 Modèle pin-hole

Comme décrit, on utilise le modèle pin-hole comme base pour modéliser la projection de 3D en 2D d'une caméra. Ce modèle ne prend pas en considération des distorsions et autres phénomènes produits par une lentille, et suppose que l'image sera nette partout. La figure 2.1 montre une représentation en deux dimensions : Un point P de l'objet est projeté vers un point P' sur le capteur. PP' représente d'un seul rayon lumineux qui passe à travers l'aperture en O . L'élément essentiel du modèle est que la position de P' sur le capteur est invariant à la position de P le long de PP' . C'est cette information qui est perdue lors de la projection 3D en 2D, et qu'on cherche à retrouver via des techniques photogrammétriques. Les scanners 3D laser mesurent directement la distance $\|OP\|$ pour construire une carte de profondeur et à partir de cela un nuage de points.

La distance $f = \|OO'\|$ est la *longueur focale*, et constitue le seul paramètre de cette projection. Avec la hauteur h du capteur, il détermine la champ de vision ω :

$$\omega = 2 \arctan \frac{h}{2f} \quad (2.30)$$

En tridimensionnel, le capteur est rectangulaire de champ de vision est donc décrit par deux angles, correspondant à la largeur w et hauteur h du capteur.

O' représente l'origine d'un *repère capteur* 2D intermédiaire. Les coordonnées (x_s, y_s) dans ce repère sont contraints par $-\frac{w}{2} \leq x_i \leq \frac{w}{2}$ et $-\frac{h}{2} \leq y_i \leq \frac{h}{2}$. Ils peuvent directement être traduits en indices de pixels (x_i, y_i) (repère image), en connaissant la taille du capteur (w, h) , la taille de l'image en pixels (W, H) , et la position (o_x, o_y) de O' dans le repère image. On définit $s_x = \frac{w}{W}$ et $s_y = \frac{h}{H}$, c'est à dire la taille d'un pixel sur le capteur en direction horizontale et verticale :

$$x_i = \frac{x_s}{s_x} - o_x \quad \text{et} \quad y_i = \frac{y_s}{s_y} - o_y \quad (2.31)$$

Des distorsions de l'image sur le capteur introduites par la lentille peuvent souvent être assez bien modélisées par les relations suivantes, où (x_{sd}, y_{sd}) sont les coordonnées d'un point sur le repère capteur avec ces distorsions, et (x_s, y_s) les coordonnées correspondantes après la correction : [Trucco & Verri, 1998]

$$x_s = x_{sd} (1 + k_1 r^2 + k_2 r^4) \quad \text{et} \quad y_s = y_{sd} (1 + k_1 r^2 + k_2 r^4) \quad (2.32)$$

avec $r = \sqrt{x_{sd}^2 + y_{sd}^2}$, et k_1, k_2 des paramètres de cette correction. Usuellement k_2 peut être mis à zéro, et seulement k_1 doit être configuré.

A partir de cela, on peut maintenant donner une définition des paramètres intrinsèques d'une caméra par le tuple

$$(f, o_x, o_y, s_x, s_y, k_1) \quad (2.33)$$

O est l'origine du repère caméra. On suppose que les coordonnées (x_c, y_c, z_c) du point P sont donnés dans ce repère. Par le théorème de Thalès, on a $\frac{-x_s}{x_c} = \frac{f}{z_c}$ et $\frac{-y_s}{y_c} = \frac{f}{z_c}$. On peut donc passer du repère caméra au repère capteur via

$$x_s = f \frac{x_c}{z_c} \quad \text{et} \quad y_s = f \frac{y_c}{z_c} \quad (2.34)$$

En coordonnées homogènes, et si on n'inclut pas la correction des distorsions mentionnée, la transformation du repère caméra dans le repère image en utilisant les paramètres intrinsèques peut être exprimé par une multiplication matricielle :

$$\begin{bmatrix} \ddot{x}_i \\ \ddot{y}_i \\ \ddot{w}_i \end{bmatrix} = \begin{bmatrix} -\frac{f}{s_x} & 0 & o_x \\ 0 & -\frac{f}{s_y} & o_y \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \quad (2.35)$$

La transformation du repère objet dans le repère caméra correspond à une rotation et une translation 3D, aussi exprimables par une matrice de multiplication en coordonnées homogènes. La projection complète du repère objet dans le repère image correspondra donc à une multiplication matricielle $(\ddot{x}_i, \ddot{y}_i, \ddot{w}_i)^T = \mathbf{M}(x_o, y_o, z_o, 1)^T$. \mathbf{M} est donnée par :

$$\mathbf{M} = \begin{bmatrix} -f_x r_{1,1} + o_x r_{3,1} & -f_x r_{1,2} + o_x r_{3,2} & -f_x r_{1,3} + o_x r_{3,3} & -f_x t_x + o_x t_z \\ -f_y r_{2,1} + o_y r_{3,1} & -f_y r_{2,2} + o_y r_{3,2} & -f_y r_{2,3} + o_y r_{3,3} & -f_y t_x + o_y t_z \\ r_{3,1} & r_{3,2} & r_{3,3} & t_z \end{bmatrix} \quad (2.36)$$

Ici, $r_{i,j}$ sont les 3×3 éléments de la matrice de rotation, $\vec{t} = (t_x, t_y, t_z)^T$ le vecteur de translation. De plus, on pose $f_x = \frac{f}{s_x}$ et $f_y = \frac{f}{s_y}$, pour que les paramètres dans l'expression soient indépendants.

2.12.3 Calibrage

La calibrage en photogrammétrie consiste à estimer les paramètres extrinsèques et/ou intrinsèques d'une caméra en utilisant une ou plusieurs photo d'un objet (pattern de calibrage) connu. [Trucco & Verri, 1998] Figure 2.2 montre un exemple d'un "polygone de calibrage" utilisé dans un laboratoire de recherche. [Elzbieta, 2009] Plusieurs cibles se trouvent sur l'objet qui peuvent être repérés sur les photos manuellement ou possiblement par des algorithmes de détection de zones d'intérêt. Les coordonnées 3D de chaque cible sont connus. En comparant les positions 2D des cibles sur plusieurs photos, un calibrage des paramètres extrinsèques et intrinsèques a pu été effectué dans ce projet.

La plupart des techniques photogrammétriques supposent que les paramètres extrinsèques et intrinsèques correspondant aux photos sont connus. Sauf si on travaille dans des conditions très particulières (par exemple avec un banc optique où les paramètres sont fixés), il est donc nécessaire d'effectuer un calibrage avant de procéder. Cependant, il existe aussi des techniques qui permettent d'extraire de l'information à partir de photos non-calibrés. [Trucco & Verri, 1998]

Comme vu en section 2.12.2, la projection de coordonnées du repère *objet* 3D dans le repère image 2D peut être exprimé en tant que matrice de transformation \mathbf{M} . Soit $m_{i,k}$ ses 3×4 éléments. Exprimé en forme non-matricielle, avec conversion de coordonnées homogènes en coordonnées on obtient

$$\begin{aligned} x_i &= \frac{\ddot{x}_i}{\ddot{w}_i} = \frac{m_{1,1}x_o + m_{1,2}y_o + m_{1,3}z_o + m_{1,4}}{m_{3,1}x_o + m_{3,2}y_o + m_{3,3}z_o + m_{3,4}} \\ y_i &= \frac{\ddot{y}_i}{\ddot{w}_i} = \frac{m_{2,1}x_o + m_{2,2}y_o + m_{2,3}z_o + m_{2,4}}{m_{3,1}x_o + m_{3,2}y_o + m_{3,3}z_o + m_{3,4}} \end{aligned} \quad (2.37)$$

En négligeant un facteur de redimensionnement, \mathbf{M} a 11 composants indépendants à chercher. 6 cibles pour lesquelles on connaît leur coordonnées image (x_i, y_i) , et leurs coordonnées objet (x_o, y_o, z_o) suffisent donc pour trouver \mathbf{M} par résolution d'un système linéaire formé par ces équations. En pratique, beaucoup

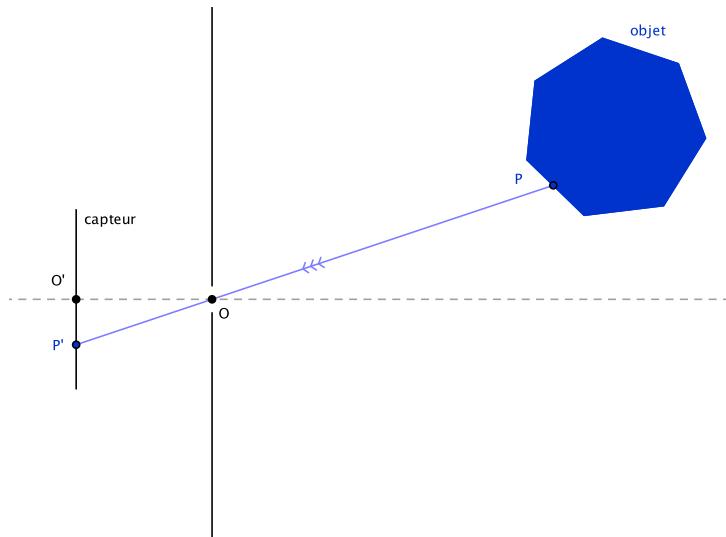


FIGURE 2.1 – Modèle caméra pin-hole



FIGURE 2.2 – Exemple de pattern de calibrage utilisé en photogrammétrie. Copié de [Elzbieta, 2009]

plus de cibles seront connus, et donc une meilleur estimation de \mathbf{M} pourra être obtenu par la méthode des moindres carrés.

La prochaine étape est de retrouver à partir des 12 valeurs $m_{i,j}$ les paramètres extrinsèques (\mathbf{R}, \vec{t}) et intrinsèques (o_x, o_y, f_x, f_y). Le développement correspondant est donné en [?].

2.12.4 Problème de correspondance

Afin de pouvoir reconstruire la position d'un point dans le repère objet à partir de deux ou plusieurs photos, il est d'abord nécessaire d'identifier la projection du même point sur les photos. Quand le but est de reconstruire un petit nombre de points spécifique, on peut manuellement identifier les points correspondants, ou utiliser des algorithmes de détection de zones d'intérêt.

Dans le but de générer un nuage de points par photogrammétrie, il sera par contre nécessaire d'associer une valeur de profondeur à chaque pixel des photos. Ici des méthodes de mise en correspondance peuvent être appliquées. En utilisant deux photos, l'*appariement dense* consiste donc à trouver pour chaque pixel (x_i, y_i) d'une image son homologue (x'_i, y'_i) sur l'autre image. Quand les deux poses de caméra se distinguent uniquement par un déplacement horizontal, alors des pixels homologues de trouveront toujours sur la même ligne dans les deux images, donc $y_i = y'_i$. La profondeur z_c peut alors être estimée par la *disparité* $d = \|x_i - x'_i\|$ des deux pixels. [Elzbieta, 2009] Ceci correspond au principe de fonctionnement de la vision humaine stéréoscopique.

Pour estimer le pixel (x'_i, y'_i) homologue à (x_i, y_i) , une méthode est de comparer visuellement les valeurs des pixels dans un voisinage V autour des pixels sur les deux images. Soit $I(x, y)$ respectivement $I'(x', y')$ la valeur d'un pixel dans le premier et le deuxième image. Pour des images en couleurs RGB, cela peut par exemple être une valeur de luminosité calculée à partie des canaux de couleur. En calcule donc $\hat{d}(x, y)$, la disparité à associer à une position dans le repère image, en maximisant une fonction de ressemblance du voisinage

$$\hat{d}(x, y) = \arg \max_{d_{\min} < d < d_{\max}} \sum_{\Delta x=-V}^V \sum_{\Delta y=-V}^V \psi(I(x + \Delta x, y + \Delta y), I'(x + \Delta x, y + \Delta y + d)) \quad (2.38)$$

Pour la fonction ψ , on utilise souvent $\psi(I_1, I_2) = I_1 I_2$ ou $\psi(I_1, I_2) = -(I_1 - I_2)^2$. Ils sont toutes les deux maximales quand $I_1 = I_2$. De tels méthodes d'appariement dense souffrent surtout d'problèmes d'occlusion (une partie d'une image n'est pas visible dans l'autre), et de fausses correspondances causés par du bruit dans les photos. Une contraint qui peut être appliquée pour éviter ces problèmes est que la même correspondance est trouvée en cherchant dans les deux directions. Pour la méthode décrite ici $\hat{d}(x_i, y_i) \simeq -\hat{d}'(x'_i, y'_i)$.

2.12.5 Reconstruction 3D

Quand les paramètres intrinsèques et extrinsèques d'une photo sont connus, on peut déterminer pour chaque pixel (x_i, y_i) la droite du rayonnement de lumière PP' (voir figure 2.1), dans le repère objet. Si on a deux photos pris de différentes poses et sur lequel le même point (x_i, y_i) à été identifié, on a une deuxième droite PP'' connue. On connaît donc en repère objet les points P' , P'' , les droites PP' , PP'' , et on sait que P, P', P'' sont coplanaires. Les coordonnées de P peuvent ainsi être calculés par triangulation.

Cela correspond à une résolution d'un système linéaire formés de deux équations du type $\vec{P}'_i \mathbf{M} = \vec{P}$, où \vec{P}'_i sont les coordonnées image de pixels correspondants sur les deux images, et l'inconnue \vec{P} les coordonnées 3D cherchés. Comme pour le calibrage, une amélioration du résultat est possible en utilisant plusieurs images.

Il n'est d'ailleurs pas nécessaire d'avoir les paramètres extrinsèques absolues des caméras, donc la rotation et translation dans le repère objet. Au lieu, des paramètres extrinsèques relatifs, c'est à dire par rapport à l'autre caméra sont suffisants.

Quand seulement les paramètres intrinsèques sont connus, une reconstruction est toujours possible à un facteur de redimensionnement près. Quand on ne dispose pas d'information sur les paramètres, une reconstruction à une transformation de perspective près est possible. Pour ces deux cas au moins 8 correspondances entre les deux images sont requis. [Trucco & Verri, 1998]

2.12.6 Combination photo-nuage de points

Comme mentionné, dans des projets de documentation 3D on combine souvent des scans laser avec des informations extraits de photos prises en combinaison avec les scans laser. La colorisation d'un nuage de points $P = \{p_i\}$ via une photo est assez simple en utilisant les techniques exposées. Le but est de trouver pour chaque point p_i du nuage de points, le pixel correspondant (x_i, y_i) sur la photo. On attribut alors la couleur de ce pixel au point du nuage.

Dans beaucoup de cas, la position de la caméra par rapport au scanner est connue. Il existe par exemple des dispositifs où une caméra est directement monté sur le scanner. Les points du nuage de points seront dans un repère propre au scanner. En utilisant une matrice de transformation qui exprime la position relative de la caméra, ils peuvent être reportés dans le repère caméra. Maintenant, les coordonnées du pixel peuvent être trouvées en utilisant les paramètres intrinsèques de la caméra, par la formule 2.35. Il peut cependant être nécessaire d'ajouter des corrections de distorsions comme ceux mentionnés. Pour des points qui sont occlus sur la photo, on aura des faux résultats, l'écart caméra-scanner doit donc être assez petit.

Quand le but sera de fusionner plusieurs nuages de points colorisés ainsi pour former un modèle complet, et pourra aussi être utile d'appliquer des corrections de homogénéisation des couleurs. [Girod, 2013] décrit une approche pour résoudre ce problème : D'abord le vignettage sur les photos est effacé, ensuite des corrections radiométriques sont appliquées sur l'ensemble des images.

Le vignettage est une réduction de la luminosité d'une photo aux périphéries. Un exemple est montré sur figure 2.3. Le phénomène peut avoir des différentes origines géométriques, optiques et électroniques, toutes liées à l'angle d'incidence du rayonnement sur le capteur. Il peut donc être modélisé par un modèle radial. L'article montre que $1 + \alpha r^2 + \beta r^4 + \gamma r^6$ décrit assez bien un facteur de l'intensité lumineuse perçue par le capteur de la caméra, en fonction de la distance au centre r . Les paramètres du modèle α, β, γ devront être estimés. Pour cela, on prend plusieurs photos pris d'un même objet avec la caméra, sur lesquelles on identifie des points de correspondance. On suppose que la luminosité ambiante est constante sur ces deux photos. Soit $G_{i,j}$ la valeur en niveaux gris du i -ième point de correspondance sur la j -ième photo. Dans ce modèle, on suppose donc que des valeurs distinctes d'un même point de correspondance sur différentes images sont dues au vignettage. Donc pour tout point de correspondance i ,

$$G_{i,1}(1 + \alpha r_{i,1}^2 + \beta r_{i,1}^4 + \gamma r_{i,1}^6) = G_{i,2}(1 + \alpha r_{i,2}^2 + \beta r_{i,2}^4 + \gamma r_{i,2}^6) = \dots \quad (2.39)$$

avec $r_{i,j}$ la distance au centre du point sur cette image. Il a été montré que la méthode RANSAC est appropriée pour trouver une bonne estimation de α, β, γ , avec les données d'entrée souvent erronées.

Après avoir corrigé les vignettes sur les photos individuelles, on veut préparer des photos pris de différents points de vue d'un objet afin de créer un nuage de points fusionné colorié de manière homogène. La méthode proposée en [Girod, 2013] cherche des points de correspondance sur des zones communes de différentes photos par le détecteur SIFT, puis calcule des matrices de correction de couleur à appliquer à l'ensemble de pixels de chaque image. La figure 2.4 montre un résultat obtenu avec cette méthode.

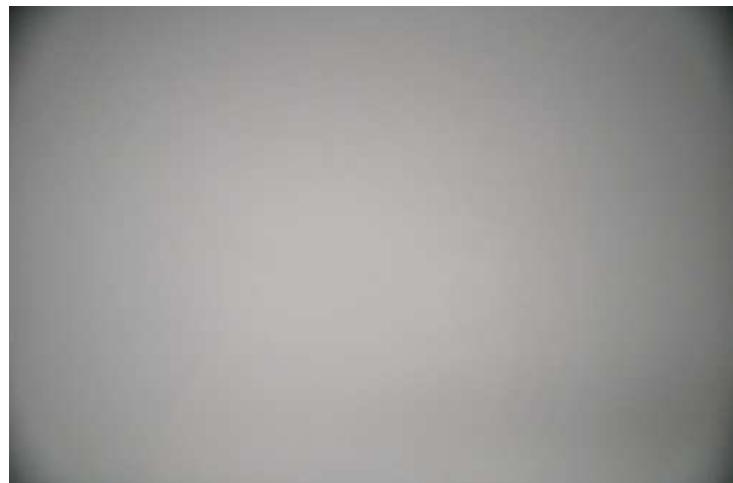


FIGURE 2.3 – Exemple de vignettage d'une photo. Copié de [Girod, 2013]

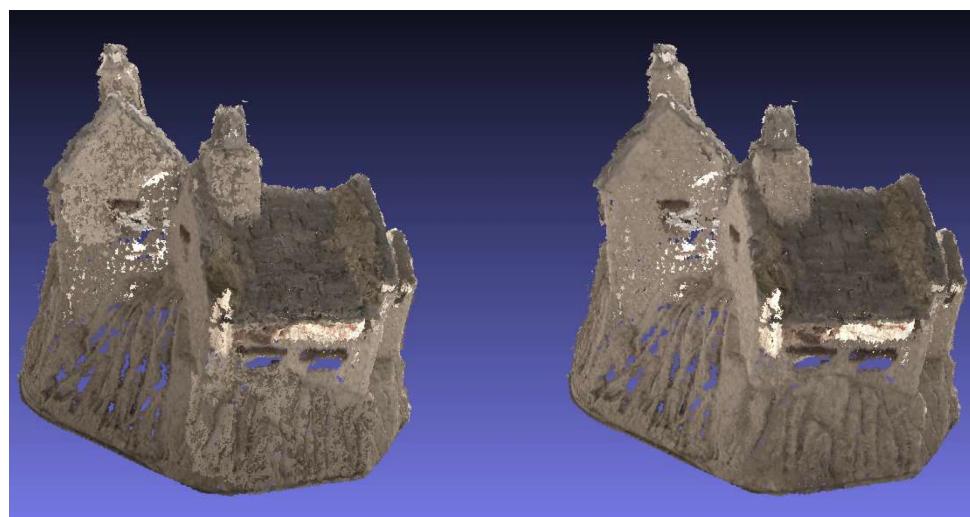


FIGURE 2.4 – Nuage de points sans et avec homogénéisation des couleurs. Copié de [Girod, 2013]

3 Projet prévu pour le mémoire

Bibliographie

- [Aiger *et al.*, 2008] Aiger, Dror, Mitra, Niloy J, & Cohen-Or, Daniel. 2008. 4-Points Congruent Sets for Robust Pairwise Surface Registration. *ACM Transactions on Graphics*, **27**(3).
- [Besl & McKay, 1992] Besl, Paul J, & McKay, Neil D. 1992 (February). A Method for Registration of 3-D Shapes. *Pages 239–256 of : IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14. IEEE.
- [Biber, 2003] Biber, Peter. 2003. The Normal Distributions Transform : A New Approach to Laser Scan Matching.
- [Center for Advanced Spatial Technologies, 2014] Center for Advanced Spatial Technologies, University of Arkansas. 2014. *Archaeological Laser Scanning*. <http://www.cast.uark.edu/home/research/archaeology-and-historic-preservation/archaeological-geomatics/archaeological-laser-scanning.html>.
- [Dold, 2005] Dold, Christoph. 2005. Extended Gaussian Images for the Registration of Terrestrial Scan Data. Institute of Cartography and Geoinformatics, University of Hannover, Germany.
- [Dold & Brenner, 2006] Dold, Christoph, & Brenner, Claus. 2006. Registration of Terrestrial Laser Scanning Data using Planar Patches and Image Data. *Pages 25–27 of : International Archives of Photogrammetry and Remote Sensing*. Leibniz University of Hannover.
- [Dold *et al.*, 2007] Dold, Christoph, Ripperda, Nora, & Brenner, Claus. 2007. Vergleich verschiedener Methoden zur automatischen Registrierung von terrestrischen Laserscandaten. *Beiträge der Oldenburger 3D-Tage*, 196.
- [Elzbieta, 2009] Elzbieta, Gardon. 2009 (Septembre). *Fusion de cartes de profondeurs en photogrammétrie architecturale*. Rapport de Projet Pluridisciplinaire, Ecole Nationale des Sciences Géographiques.
- [Fischler & Bolles, 1980] Fischler, Martin A, & Bolles, Robert C. 1980 (March). *Random Sample Consensus : A paradigm for model fitting with applications to image analysis and automated cartography*. Tech. rept. Artificial Intelligence Center, SRI International, Menlo Park, California.
- [Girod, 2013] Girod, Luc. 2013 (Septembre). *Égalisation radiométrique de nuages de points 3D : Principes et Algorithmique*. Stage de fin d'études, Ecole Nationale des Sciences Géographiques (France).
- [Grussenmeyer *et al.*, 2012] Grussenmeyer, P., Alby, E., Landes, T., Koehl, M., Guillemin, S., Hullo, J.-F., Assali, P., & Smigiel, E. 2012. Recording Approach Of Heritage Sites Based On Merging Point Clouds From High Resolution Photogrammetry And Terrestrial Laser Scanning. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **XXXIX-B5**, 553–558.
- [Holz & Behnke, 2014] Holz, Dirk, & Behnke, Sven. 2014 (June). Registration of Non-Uniform Density 3D Point Clouds using Approximate Surface Reconstruction. Joint 45th International Symposium on Robotics (ISR) and 8th German Conference on Robotics (ROBOTIK). Autonomous Intelligent Systems Group, University of Bonn, Germany, Munich.
- [Horn, 1984] Horn, Berthold K.P. 1984. Extended Gaussian Images. *Proceesings of the IEEE*, **72**(12), 1671–1686.
- [Horn, 1987] Horn, Berthold K.P. 1987 (April). Closed-form solution of absolute orientation using unit quaternions. *Page 629 of : Journal of the Optical Society of America A*, vol. 4. Department of Electrical Engineering, University of Hawaii at Manoa.
- [Keiner, 2011] Keiner, Christopher. 2011. *Bildregistrierung von 3D-Laserscans und Bildern für Wadi Sura*. M.Phil. thesis, Fachbereich Mathematik und Informatik der Freien Universität Berlin, April.
- [Kersten *et al.*, 2006] Kersten, Thomas, Przybilla, Heinz-Jürgen, & Lindstaedt, Maren. 2006. Integration, Fusion und Kombination von terrestrischen Laserscannerdaten und digitalen Bildern. November.

- [Lerma *et al.*, 2009] Lerma, José Luis, Navarro, Santiago, Cabrelles, Miriam, & Villaverde, Valentín. 2009. Terrestrial laser scanning and close range photogrammetry for 3D archaeological documentation : the Upper Palaeolithic Cave of Parpalló as a case study. *Journal of Archaeological Science*, **37**(October), 499–507.
- [Lichtenstein, 2011] Lichtenstein, Maria. 2011. *Strukturbasierte Registrierung von Punktwolken unter Verwendung von Bild- und Laserscannerdaten*. Ph.D. thesis, Rheinisch-Westfälische Technische Hochschule Aachen.
- [Low, 2004] Low, Kok-Lim. 2004 (February). *Linear Least-Squares Optimization for Point-to-Plane ICP Surface Registration*. Tech. rept. Department of Computer Science; University of North Carolina at Chapel Hill.
- [Matiukas & Miniotas, 2011] Matiukas, V., & Miniotas, D. 2011. Point Cloud Merging for Complete 3D Surface Reconstruction. ISSN 1392–1215 ELECTRONICS AND ELECTRICAL ENGINEERING, no. 7. Department of Electronic Systems, Vilnius Gediminas Technical University.
- [Rusinkiewicz & Levoy, 2001] Rusinkiewicz, Szymon, & Levoy, Marc. 2001. Efficient Variants of the ICP Algorithm. vol. 3-D Digital Imaging and Modeling. Stanford University.
- [Schönemann, 1966] Schönemann, Peter H. 1966. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, **31**(1), 1–10.
- [Schönemann & Carroll, 1970] Schönemann, Peter H., & Carroll, Robert M. 1970. Fitting one matrix to another using choice of a central dilation and a rigid motion. *Psychometrika*, **35**(2), 245–255.
- [Segal *et al.*, 2009] Segal, Aleksandr V., Haehnel, Dirk, & Thrun, Sebastian. 2009. Generalized-ICP. *Proceedings of Robotics : Science and Systems*.
- [Toldo *et al.*, 2010] Toldo, Roberto, Beinat, Alberto, & Crosilla, Fabio. 2010. Global registration of multiple point clouds embedding the Generalized Procrustes Analysis into an ICP framework. 3DPVT 2010 Conference.
- [Tournas & Tsakiri, 2009] Tournas, E., & Tsakiri, M. 2009. Automatic 3D Point Cloud Registration for Cultural Heritage Documentation. *Laser scanning 2009, IAPRS*, **XXXVIII**(September), 189–194.
- [Trucco & Verri, 1998] Trucco, Emanuele, & Verri, Alessandro. 1998. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall.
- [Tuytelaars & Mikolajczyk, 2007] Tuytelaars, Tinne, & Mikolajczyk, Krystian. 2007. Local Invariant Feature Detectors : A Survey. *Foundations and Trends in Computer Graphics and Vision*, **3**(3).