

École polytechnique de Louvain

Sphaeroptica

**Photogrammétrie appliquée à des images
d'arthropodes**

Auteur : **Yann POLLET**

Promoteur : **Sébastien JODOGNE**

Lecteurs : **Sébastien JODOGNE, Antoine LEGRAND, Patrick SEMAL**

Année académique 2022–2023

Master [120] en sciences informatiques

Remerciements

Je voudrais remercier mon promoteur de mémoire, Sébastien Jodogne, pour son suivi régulier et ses conseils utiles pendant tout ce travail.

Je voudrais ensuite remercier mon superviseur à l'Institut Royal des sciences naturelles de Belgique, Patrick Semal, de m'avoir proposé ce sujet de mémoire, ainsi que pour son aide régulière lors du développement de Sphaeroptica et de l'accès aux infrastructures de l'institut.

Je remercie également Adrien Gressin et Xavier Muth de la Haute école d'ingénierie et de gestion du canton de Vaud (HEIG-VD) / Haute école spécialisée de Suisse Occidentale (HES-SO), rattaché au groupe de recherche « Géomatique » pour leurs conseils avisés qui m'ont permis d'aller dans la bonne direction.

Enfin, je remercie mes parents pour leur aide précieuse pendant l'écriture de ce mémoire ainsi que mes amis, Noé, Guillaume et Maëlle, pour leurs encouragements.

Abstract

Depuis plusieurs décennies, les institutions de sciences naturelles dans le monde numérisent leurs collections pour pouvoir les archiver, diminuer les manipulations futures et les “prêter” virtuellement sans devoir envoyer le spécimen, ce qui représente toujours un risque de perte ou de dommage.

Malheureusement, pour les petits spécimens complexes comme les arthropodes, les solutions optant pour un calcul de modèle 3D, simplifient les structures et effacent des détails cruciaux pour l'étude de ces spécimens.

Pour répondre à cette problématique, nous avons d'abord mis en place un dispositif scAnt [1], un scanner open-hardware, pour l'acquisition d'images à haute résolution autour d'un spécimen.

Ensuite, nous avons développé Sphaeroptica, un logiciel de visualisation et de mesures sur base d'images centrées autour d'un point permettant d'effectuer des mesures 3D directement sur les images 2D calibrées. Malgré le fait que nous revenions aux bases de la photogrammétrie, aucun logiciel de photogrammétrie connu ne permet d'effectuer des mesures directement sur les images géoréférencées à haute résolution sans calcul d'un modèle 3D (mesh). Sphaeroptica contient également une caméra virtuelle synthétisant de nouvelles vues à partir des images acquises, avec pour objectif d'avoir un aperçu plus fluide de la visualisation du spécimen.

For many decades, natural sciences institutions in the world have been digitizing their collections to be able to archive them, to reduce the handling of it and to “loan” it virtually without having to send the specimen away, which always represents a risk of loss or damage. Unfortunately, for small complex specimens as arthropods, solutions computing a 3D model simplify structures and erase crucial details for the specimens' study.

To answer this problem, we first set up a scAnt device [1], an open-hardware scanner, for the image acquisition around a specimen. We then developed Sphaeroptica, a software that allows the visualization and 3D measurements of an object directly on 2D calibrated images. Even though we get back to the basics of photogrammetry, none of known photogrammetry softwares allows us to perform measurements directly on geo-referenced images with high resolution without the computation of a 3D mesh. Sphaeroptica also contains a virtual camera that synthesizes new views from existing ones, aiming at having a smoother overview of the specimen.

Lexique des acronymes

IRSNB : Institut Royal des Sciences Naturelles de Belgique. 1, 4, 8, 9, 11, 23, 43, 44

DiSSCo : Distributed System of Scientific Collections. 1

EDOF : Extended-Depth Out of Field. 6, 7, 9, 23, 45

DiNArDa : Digital Natural History Archive Darmstadt. 7

SfM : Structure-from-Motion. 7, 9, 10, 20-22

MVS : Multi-View Stereo. 10, 20

SIFT : Scale-Invariant Feature Transform. 21

ORB : Oriented FAST and rotated BRIEF. 21

RANSAC : RANdom SAmple Consensus. 21, 22

DLT : Direct Linear Transform. 28, 41, 42

Table des matières

1	Introduction.....	1
2	Etat de l'art de la numérisation de spécimens.....	3
2.1	Scanner à lumière structurée.....	3
2.1.1	Principe de la lumière structurée.....	3
2.1.2	Les scanners.....	4
2.1.3	Limitations.....	4
2.2	Scanner à images.....	6
2.2.1	Images Extended-depth out of field.....	6
2.2.2	DISC3D.....	7
2.2.3	scAnt.....	8
2.3	Les logiciels de photogrammétrie.....	9
3	Calibration de caméra.....	10
3.1	Modèle de caméra.....	11
3.1.1	Bases.....	12
3.1.2	Matrices intrinsèque et extrinsèque.....	13
3.1.3	Distorsion de la lentille.....	15
3.2	Méthodes de calibration.....	17
3.2.1	Calibration avec une mire 2D/3D.....	17
3.2.2	Structure-from-motion.....	18
4	Sphaeroptica.....	22
4.1	Acquisition des images.....	22
4.2	Calibration.....	23
4.3	Triangulation.....	25
4.3.1	Direct Linear Transform.....	27
4.3.2	Erreur de reprojection :.....	28
4.4	Comparaisons.....	28
5	Caméra Virtuelle.....	34
5.1	Calcul de la position géodésique.....	35
5.2	Sélection de l'image de référence.....	35
5.3	Calibration de la caméra virtuelle.....	35
5.4	Calcul de la matrice d'homographie H.....	38
6	Conclusion.....	41
6.1	Contributions.....	41
6.2	Discussions.....	41
6.3	Évolution future.....	43

Table des figures

Figure 1: image d'un scan raté d'un heliocopris à cause de sa couleur noire	5
Figure 2: photo stackée comparée avec une des 48 images utilisées pour son stacking [8] ..	7
Figure 3: calibration des images prises par scAnt autour d'un spécimen.....	8
Figure 4: génération des Images EDOF et de leur masquage par scAnt	9
Figure 5: la matrice de caméra P [11].....	10
Figure 6: les matrices intrinsèque et extrinsèque [12]	11
Figure 7: les transformations de systèmes de coordonnées [13]	11
Figure 8: schéma de caméra à modèle à sténopé [16]	12
Figure 9: : le modèle de caméra de sténopé [15, p. 154]	13
Figure 10: les systèmes de coordonnées [17]	13
Figure 11: les systèmes de coordonnées droitier et gaucher [19]	14
Figure 12: : la distorsion radiale[20].....	16
Figure 13: la distorsion tangentielle [21]	16
Figure 14: pattern 12x9 de carrés de 2mm de côté, prise par notre implémentation scAnt ..	18
Figure 15: implémentation du Sfm incrémental par COLMAP [10].....	19
Figure 16: image utilisée pour la calibration intrinsèque.....	23
Figure 17: image de la mire 3D construite en Lego	24
Figure 18: calibration des images reconnues avec le motif de Zhang.....	24
Figure 19: back projection du pixel en rayon [31]	25
Figure 20: triangulation avec minimisation de l'erreur géométrique [15, p. 314]	26
Figure 21: différences de mesures en mm	30
Figure 22: différences de mesures en %	30
Figure 23: image et modèle 3D de la deuxième mouche	32
Figure 24: comparaison des différences observées entre la deuxième mouche et le restes des arthropodes.....	33
Figure 25: position de base de la caméra.....	36
Figure 26: rotations effectuées pour calibrer la caméra virtuelle	37
Figure 27: comparaison rapprochée entre une image prise par scAnt et le modèle de référence calculée par Artec.....	42

Table des tables

Table 1: moyenne, médiane et écart type	31
Table 2: différences de ratio	31
Table 3: erreur de reprojexion sur les arthropodes mesurés	32

1 Introduction

Les Institutions de sciences naturelles poursuivent différents objectifs : préserver, étudier et valoriser les collections recueillies.

Grâce aux progrès de la technologie, la numérisation a permis de les aider dans cette quête [2].

Qu'elle soit en 2D ou 3D, la numérisation des collections offre de nouvelles possibilités pour les institutions :

- Un archivage pour garder une trace d'une pièce physique même en cas de destruction ;
- Un partage avec la communauté scientifique tout en préservant la pièce d'un risque de destruction, de perte ou de vol ;
- Une étude plus poussée grâce aux possibilités de nouveaux calculs, tels qu'une mesure exacte de volume, surface, épaisseur, tout en réduisant la manipulation de spécimens fragiles et rares.

C'est dans cette même optique que beaucoup d'institutions lancent des projets comme de la numérisation de masse ou la création de portails de partage de collections digitales.

Dans ce dernier cas, nous pouvons citer un projet européen du nom de [DiSSCo](#) (Distributed System of Scientific Collections).

C'est une infrastructure de recherche distribuée qui est dans sa phase de préparation et sera opérationnelle en 2026. DiSSCo consiste à construire une infrastructure commune entre les institutions de collections de sciences naturelles (NCS) de 23 pays en Europe pour rassembler l'ensemble des collections, archives scientifiques de recherches constituées par les scientifiques des différentes disciplines...

DiSSCo contribue aussi à coordonner les projets de numérisations de masse. On estime qu'il y a 1,5 milliard de spécimens dans les institutions dans le monde.

Si on prend seulement les chiffres de l'Institut Royal de Sciences Naturelles de Belgique (IRSNB), ce chiffre s'élève à 38 millions, et seulement 600 000 de ces spécimens ont été numérisés au niveau de leurs données/métadonnées et moins de 100 000 ont été photographiés ou scannés.

Pour la numérisation, de nombreux outils, comme différents types de scanners et de logiciels sont utilisés dans le but d'offrir la meilleure copie digitale des spécimens pour leur étude par les scientifiques. Le site [DIGIT-KEY](#) [3] synthétise les différentes techniques utilisées par les musées d'histoire naturelle selon le résultat désiré et la nature des spécimens.

Malheureusement, le calcul de modèles surfaciques et les algorithmes utilisés simplifient les modèles entraînant une perte de qualité. Des détails morphologiques importants à l'étude des spécimens petits et/ou fins, comme les arthropodes, avec des structures transparentes ou noires, sont toujours très difficiles à numériser avec une qualité suffisante.

C'est pourquoi l'objectif de ce travail est d'offrir un outil permettant la visualisation de spécimens de petite taille avec la meilleure résolution possible, afin de garder la texture du spécimen, tout en offrant la possibilité de prendre des mesures sur ce "modèle".

Le Museum für Naturkunde Berlin a déjà développé un logiciel permettant la visualisation 360° d'arthropodes [4] avec un tuilage des images pour avoir une résolution optimale. Cependant,

ce logiciel ne permet pas la prise de mesure, fonctionnalité absolument nécessaire à l'étude des arthropodes et au développement de l'utilisation de la morphométrie géométrique au sein de ces groupes de spécimens.

Dans le cadre de ce mémoire, nous allons nous concentrer sur la numérisation 3D des arthropodes. Pour cela, nous avons construit un équipement de numérisation 360°, ScAnt [1], sur base de leur publication disponible sur PeerJ. L'équipement de numérisation effectue automatiquement la prise d'images 360° du spécimen. Avec les scanners à lumière structurée type Artec Micro, il s'agit de l'approche la plus avancée pour la numérisation des arthropodes. Ce mémoire et ses développements se situent en aval de cette acquisition.

Nous avons tout d'abord effectué une calibration des images prises par le scanner scAnt.

La méthodologie utilisée pour la calibration sera développée dans le troisième chapitre. Nous présenterons ensuite le logiciel de photogrammétrie Sphaeroptica que nous avons développé afin de répondre aux besoins des entomologistes. Cette application comporte une visualisation 3D des spécimens étudiés et permet de procéder à des mesures de ces derniers.

2 Etat de l'art de la numérisation de spécimens

Dans le but de faciliter cette numérisation, divers outils sont utilisés, en fonction des besoins des numérisations à effectuer, qu'elles soient en 2D ou 3D, sur un spécimen grand ou très petit par exemple [2].

En se focalisant sur la numérisation 3D des collections, on peut tout d'abord y trouver des scanners facilitant la prise en charge de celle-ci et donc aidant à son automatisation.

Les premiers scanners présentés utilisent la technologie de la lumière structurée. Son avantage premier est sa précision. Ainsi, les modèles créés par ce genre de scanner seront utilisés comme modèles de références pour les tests.

D'autres scanners, eux, automatisent la prise d'images 2D tout autour d'un arthropode, limitant l'action humaine à la simple mise en place de l'objet. Plusieurs stations peuvent alors tourner simultanément avec une seule personne en charge. Cela permet d'augmenter la vitesse de numérisation des collections ou simplement d'offrir au scientifique la possibilité de se consacrer en même temps à d'autres tâches.

Enfin, nous terminerons cette section par la présentation des logiciels de photogrammétrie déjà présents sur le marché et utilisés pour la numérisation des collections.

2.1 Scanner à lumière structurée

2.1.1 Principe de la lumière structurée

Un scanner à lumière structurée peut être divisé en deux parties :

- Un projecteur qui permet de projeter des motifs de lumière parfaitement calibrés sur une surface à modéliser
- Une caméra capturant le motif déformé par la surface

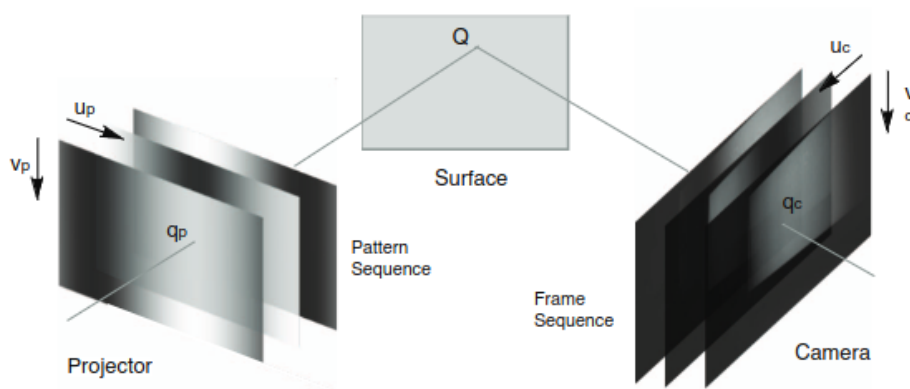


Fig. 1. Principle of structured light scanning with one camera and one projector, as implemented in SLStudio. Object surface points are determined by triangulating corresponding points in the camera and projector coordinate spaces.

Figure 2.1 : principe du scan à lumière structurée avec un projecteur et une caméra [5, p. 2]

Le résultat donne un modèle 3D très précis mais sans texture.

Certains scanners contiennent une caméra supplémentaire pour pouvoir ajouter la texture au modèle. Il est également possible d'ajouter la texture après coup grâce à des photos prises de l'objet.

2.1.2 Les scanners

La technologie présente l'énorme avantage d'offrir un large éventail de possibilités.

En effet, plusieurs types de scanners utilisant la lumière structurée ont été développés.

Il y a des scanners portables très facilement maniables. Ces scanners peuvent être utilisés pour la numérisation d'objets à taille variable (allant de petits objets à des êtres humains, ou même des scènes entières).

Dans le cadre de ce mémoire, le scanner utilisé pour créer les modèles de référence est un scanner Artec Micro. Il a été choisi car déjà utilisé par l'IRSNB et le musée royal de l'Afrique centrale.

En effet, ce scanner a été créé dans le but de scanner les très petits objets (9cm de diamètre maximum) et sa calibration permet d'être précis à 10 microns près.

C'est un scanner stationnaire, demandant peu d'intervention humaine pour la prise du modèle et la rotation de l'objet. Cela permet donc au scientifique de se consacrer à d'autres tâches pendant le scan de l'arthropode.

2.1.3 Limitations

Malheureusement, cette technologie a un coût important.

Un seul scanner peut coûter plusieurs dizaines de milliers d'euros. Ce n'est donc pas un outil facilement acquérable.

Ensuite, certaines surfaces d'objet sont difficilement capturables par un scanner à lumière structurée [6]:

- Les surfaces trop **sombres** : elles ne réfléchissent pas assez la lumière pour que le motif projeté puisse être reconnu.
- Les surfaces **réfléchissantes** : le motif risque de rebondir dans trop de directions imprévisibles pour une bonne capture par le senseur.
- Les surfaces **transparentes** : la lumière risque de traverser la surface et donc ne rebondira pas vers la caméra.
- Les surfaces **velues** : la finesse des poils risque de faire rebondir la lumière dans tous les sens, rendant la reconnaissance du motif projeté plus difficile.
- Les surfaces très **finies** : elles peuvent être entièrement illuminées par le motif et donc mal reconnues et retirées du modèle.
- Les surfaces **encastrées** : plus difficile d'accès, si la lumière ne peut pas être projetée dedans, il est impossible pour le scanner de savoir que cette surface existe.



Figure 1: image d'un scan raté d'un heliocopris à cause de sa couleur noire

Pour pouvoir effectuer le scan d'un objet contenant un de ces types de surface, il est possible de l'asperger d'un spray mat temporaire. Cependant, cette solution pourrait compromettre la préservation des spécimens très fragiles.

Enfin, comme avec tout modèle 3D, le modèle généré perd énormément en résolution. Ce qui peut être néfaste lorsque les détails du spécimen sont importants pour les scientifiques. Par exemple, le nombre de poils visibles et leur taille peuvent être indispensables pour la classification d'un insecte.

2.2 Scanner à images

Une autre manière de reconstruire un modèle 3D est de le faire simplement à partir d'images 2D prises par une caméra à haute résolution.

Tout commence par l'acquisition de photos tout autour de l'objet.

Celle-ci peut être faite manuellement grâce à un trépied et une plaque tournante.

Dans ce chapitre, nous vous présentons deux scanners spécialisés pour la prise d'images des petits objets:

- DISC3D
- scAnt

Mais avant de vous détailler les scanners, nous devons vous expliquer ce que sont les images extended-depth out of field (EDOF) qui sont calculées durant les scans.

2.2.1 Images Extended-depth out of field

Les images EDOF sont des images stackées en fonction de leur focus. C'est aussi plus communément appelé **focus stacking** [7].

Les images EDOF consistent à faire un stacking de plusieurs images sur une même prise de vue afin d'augmenter la résolution de l'objet pris en photo. Cela permet d'avoir une netteté complète sur cet objet.

Pour les petits spécimens qui ont des détails très précis, comme des poils, il est difficile de déterminer une bonne mise au point et un bon champ de profondeur pour avoir une netteté correcte sur l'entièreté du spécimen.

C'est pour contrer ce problème que la technique du focus stacking est utilisée.

En fonction de la netteté requise et de la taille de l'objet, plus ou moins d'images sont requises pour le focus stacking.

En général, pour un petit arthropode, trois ou quatre images sont suffisantes.



Figure 2: photo stackée comparée avec une des 48 images utilisées pour son stacking [8]

Sur la figure 2, vous pouvez voir à gauche une image EDOF qui est bien plus nette, avec des détails plus identifiables que sur l'image de droite, une des images ayant permis le focus stacking.

Cependant cette solution peut donner des erreurs géométriques sur l'image, puisqu'elle consiste à combiner plusieurs images à calibration différentes (soit intrinsèque soit extrinsèque) pour n'en faire qu'une seule. Pour cette raison, nous n'avons pas choisi d'utiliser ce processus dans la démonstration de Sphaeroptica dont le but est d'effectuer les mesures 3D sur les données les plus brutes possibles, sans transformation.

2.2.2 DISC3D

Le Darmstadt Insect Scanner (DISC3D) est un scanner développé par la Digital Archive of Natural History (DiNArDa) [9]. C'est un scanner spécifiquement conçu pour la numérisation des arthropodes épinglés.

Ce scanner utilise une méthode plus conventionnelle pour la création de modèles 3D.

Le processus consiste à faire une acquisition de plusieurs prises de vue différentes afin de constituer un modèle 3D grâce à la technique Structure-from-Motion (SfM) avec Agisoft Metashape.

En supplément, pour avoir une meilleure résolution des images, le scanner va calculer des images EDOF.

Limitations

Tout d'abord, le prix d'un scanner DISC3D peut varier entre 4.000 et 8.000€ dépendant du prix des composants.

Ensuite, le temps d'acquisition des images peut durer plus d'une heure, temps auquel il faut ajouter celui du calcul d'un modèle 3D.

Ainsi, il faut faire un choix entre la résolution des images, et donc la résolution du modèle, et le temps de calcul du modèle 3D.

2.2.3 scAnt

scAnt est un scanner open hardware créé par l'Imperial College UK [1].

C'est une version open source de DISC3D ; il fonctionne d'une manière similaire. L'ensemble des pièces qui le compose est réalisé soit grâce à de l'impression 3D soit grâce à du découpage, ce qui rend sa création accessible.

Limitations

En plus des limitations citées pour DISC3D, il faut ajouter le fait de devoir assembler ce dispositif soi-même.

Cependant, nous pouvons noter une réduction du prix par dix par rapport à celui de DISC3D.

Fonctionnement

Le scanner scAnt étant utilisé par l'IRSNB pour la numérisation de leurs spécimens, nous allons nous attarder sur le fonctionnement de ce scanner.

Pour information, c'est un outil que nous avons dû assembler préalablement à cette thèse de fin d'études.

ScAnt fonctionne grâce au mouvement de 3 moteurs.

Tout d'abord, les deux premiers ont pour but d'effectuer les rotations autour l'insecte pour pouvoir obtenir un nuage de photo tout autour de celui-ci, comme une "sphère" d'images (voir figure 3)

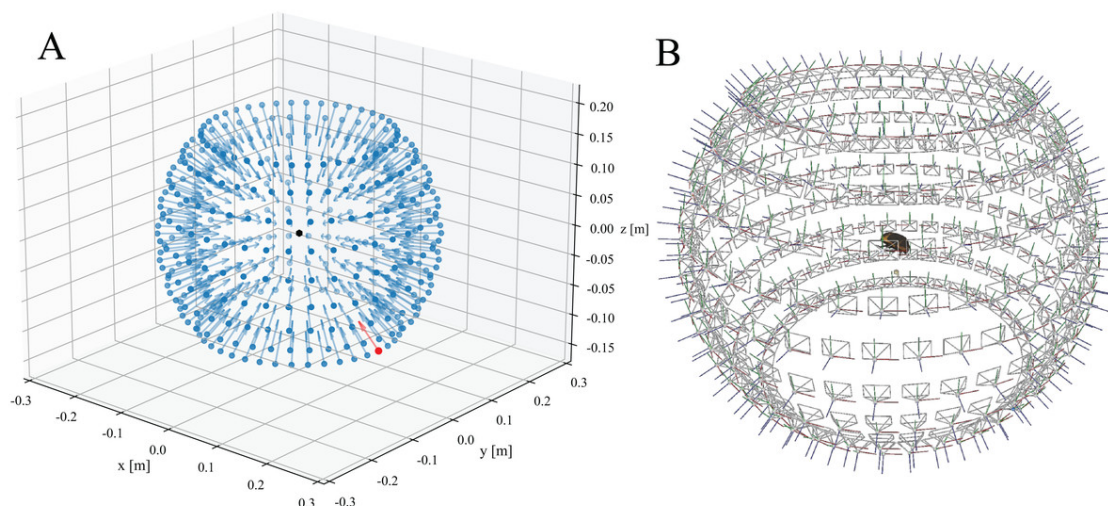


Figure 3: calibration des images prises par scAnt autour d'un spécimen

Le troisième a pour tâche de déplacer la caméra sur son axe de prise de vue pour pouvoir changer la profondeur de champ et s'assurer d'avoir l'entièreté de l'insecte en net.

Ces différentes photos pour une même prise de vue seront stackées afin d'obtenir une image EDOF.

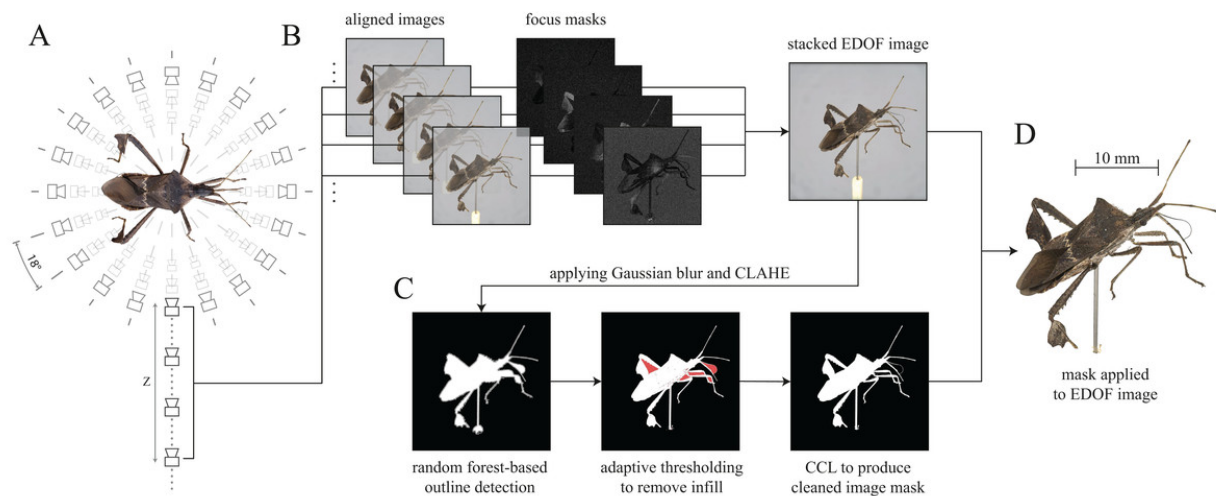


Figure 4: génération des Images EDOF et de leur masquage par scAnt

Enfin, lorsque le processing de stacking, complété par un processus de masquage pour retirer le fond, est terminé, les images peuvent être envoyées à un logiciel de photogrammétrie afin d'en extraire un modèle 3D.

2.3 Les logiciels de photogrammétrie

Des dizaines de logiciels de photogrammétrie sont disponibles pour le grand public. Nous nous limiterons à n'en citer que trois qui sont pertinents pour ce mémoire :

1. Agisoft Metashape, logiciel commercial utilisé à l'IRSNB
2. Meshroom, logiciel open source, véritable alternative à Agisoft Metashape
3. COLMAP, un logiciel open source très performant pour la partie SfM

Agisoft Metashape est un logiciel commercial développé par Agisoft, anciennement nommé Photoscan. Il permet d'effectuer aussi bien de la modélisation 3D que de la photogrammétrie aérienne. Cette solution est particulièrement polyvalente car beaucoup de formats sont acceptés et plus de 50 000 images peuvent être traitées à la fois sans problème.

Meshroom est une solution open source développée par AliceVision. Elle présente l'avantage que le processus de modélisation est paramétrable à volonté, permettant aux utilisateurs de maximiser les performances en fonction de leurs besoins.

COLMAP [10] est un logiciel open source créé par Johannes L. Schönberger, lorsqu'il était doctorant à l'ETH Zurich. Ce logiciel est un pipeline pour les projets en SfM traditionnels et en Multi-View Stereo (MVS).

Nous terminons ici la présentation de l'état de l'art, avec les avantages et les inconvénients des solutions actuelles.

3 Calibration de caméra

Nous arrivons au cœur du sujet de ce mémoire.

Après l'acquisition d'images autour d'un insecte réalisée avec scAnt, il faut calculer la calibration de ces images.

Une image étant un enregistrement du monde en 3D sur un plan en 2D, chaque image dépend du type de caméra utilisée, de sa position et de son orientation lorsqu'elle est prise. Ce sont, respectivement, les paramètres intrinsèques et extrinsèques de la caméra.

Pour éviter toute confusion pour la suite, il faut noter que le mot "caméra" représente deux concepts distincts : la caméra "réelle", celle qui prend la photo et la caméra "enregistrée", celle définie pour chaque image lors de la prise par la caméra réelle. Vu que cette dernière utilisation est beaucoup plus présente dans ce travail, nous utiliserons le mot "caméra" pour la caméra enregistrée. Et nous spécifierons si nous parlons de la caméra réelle.

Le processus de calibration de caméra a pour objectif d'estimer les paramètres intrinsèques et extrinsèques de chacune des caméras enregistrées.

Pour ce mémoire, nous avons dû comprendre ce processus afin d'obtenir la calibration de caméra la plus précise possible pour le traitement des données par Sphaeroptica, et ce avec les outils à disposition. Les outils utilisés ont été la librairie OpenCV, et le logiciel commercial Agisoft Metashape v1.8.3. Ce dernier a été choisi parce qu'il était déjà maîtrisé à l'IRSNB.

La calibration de caméra est la détermination de la matrice 3x4 de caméra P, tel que pour tout point 3D dans le monde réel, on puisse calculer sa correspondance sur le plan d'image 2D (voir "modèle de caméra à sténopé"), selon la formule suivante : $x = PX$.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} p_1 & p_2 & p_3 & p_4 \\ p_5 & p_6 & p_7 & p_8 \\ p_9 & p_{10} & p_{11} & p_{12} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

homogeneous
image
3 x 1

Camera
matrix
3 x 4

homogeneous
world point
4 x 1

Figure 5: la matrice de caméra P [11]

Cette matrice de projection peut être décomposée en deux matrices distinctes, les matrices intrinsèque et extrinsèque. La matrice intrinsèque correspond à la manière dont le monde est perçu par la caméra, et contient les valeurs de distance focale et du centre optique. La matrice extrinsèque, quant à elle, correspond à celle dont la caméra est perçue dans le monde. Elle contient l'orientation et la position de celle-ci.

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

2D Image Coordinates
Intrinsic properties (Optical Centre, scaling)
Extrinsic properties (Camera Rotation and translation)
3D World Coordinates

Figure 6: les matrices intrinsèque et extrinsèque [12]

Cette formule-ci comporte plusieurs transformations à travers quatre systèmes de coordonnées différents. Ces systèmes sont :

- Le système de coordonnées du monde centré en son origine (3D) ;
- Le système de coordonnées de la caméra centré en son centre (3D) ;
- Le système de coordonnées de l'image (2D) ;
- Le système de coordonnées des pixels, permettant de commencer au pixel (0,0) (souvent situé en haut à gauche).

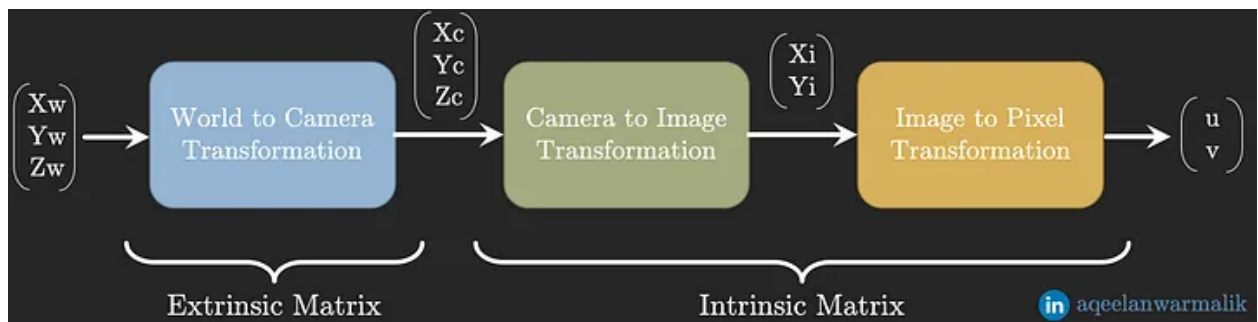


Figure 7: les transformations de systèmes de coordonnées [13]

Dans ce chapitre, nous commencerons par présenter le modèle de caméra utilisé pour la reconstruction 3D. Ce modèle est nécessaire pour comprendre les différentes méthodes de calibration, qui seront présentées dans la section d'après.

3.1 Modèle de caméra

Pour ce mémoire, nous avons opté pour le modèle de caméra à sténopé avec distorsion radiale et tangentielle.

C'est donc ce modèle que nous allons expliciter ci-après.

3.1.1 Bases

Pour pouvoir définir la matrice de projection P résultante du processus de calibration, nous devons donner des éléments théoriques sur les coordonnées homogènes et le modèle de caméra à sténopé.

Dans la photogrammétrie, les **coordonnées homogènes** forment la base de la projection du monde réel en 3D dans un espace 2D [14].

Dans un système euclidien dans l'espace 2D, un point est représenté par deux coordonnées (x,y) [15, p. 153]. Dans un espace projectif, un point, aussi appelé coordonnée homogène, est représenté par trois coordonnées telles que (xk,yk,k) correspond au même point pour tout $k > 0$. Ainsi, pour toute combinaison (x,y,k) , on peut retrouver le point $(x/k, y/k)$ dans le système euclidien.

Les coordonnées homogènes permettent aussi la représentation de points à l'infini, lorsque k égale 0 parce que $x/0$ et $y/0$ valent l'infini.

Ce système d'homogénéité fonctionne dans toutes les dimensions possibles. Nous verrons également des coordonnées homogènes $(X,Y,Z,1)$ dans un plan euclidien en 3D.

Le **modèle de caméra à sténopé** est le modèle de caméra le plus basique. C'est également le modèle de la XD utilisée pour l'acquisition des images dans le cadre de ce mémoire.

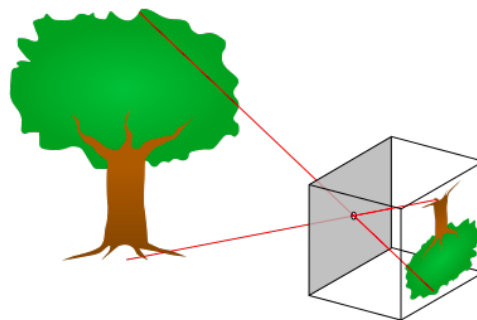


Figure 8: schéma de caméra à modèle à sténopé [16]

Dans la figure 8, nous pouvons constater que la lumière rentre par un point unique (le centre de la caméra) et se projette au fond de la boîte. L'image se retrouve donc retournée.

Des rayons de lumières passent à travers un point unique, le centre de la caméra ou le sténopé, et se projettent contre le plan d'image, situé à distance focale f . L'image est d'ailleurs retournée.

Mathématiquement, Hartley et Zisserman le présentent de la manière suivante :

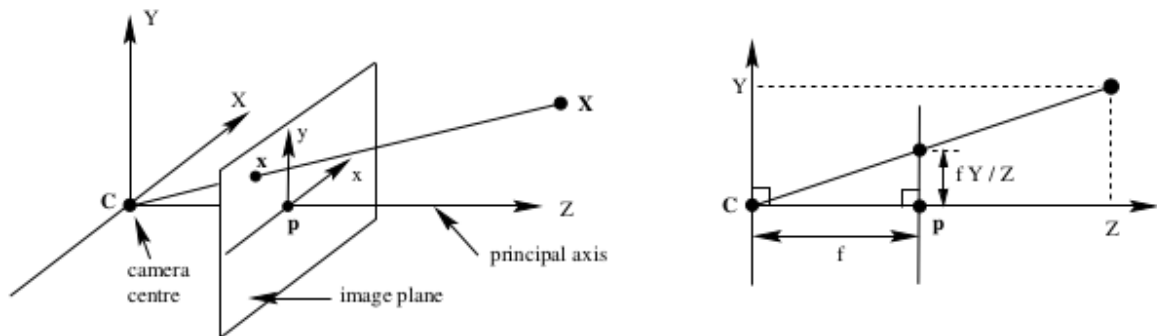


Figure 9: : le modèle de caméra de sténopé [15, p. 154]

Dans le schéma ci-dessus, le plan d'image est ramené devant le centre de caméra afin de faciliter la compréhension du modèle.

L'axe principal intersecte avec le plan d'image en un point p appelé point principal.
Enfin tout point dans le monde réel X sera projeté sur le plan d'image en fonction de l'intersection du vecteur dans la direction du centre de la caméra et son plan d'image.

3.1.2 Matrices intrinsèque et extrinsèque

Passons maintenant à la présentation des différents systèmes de coordonnées et des transformations effectuées pour passer d'un point 3D (dans le monde) à sa correspondance sur l'image 2D.

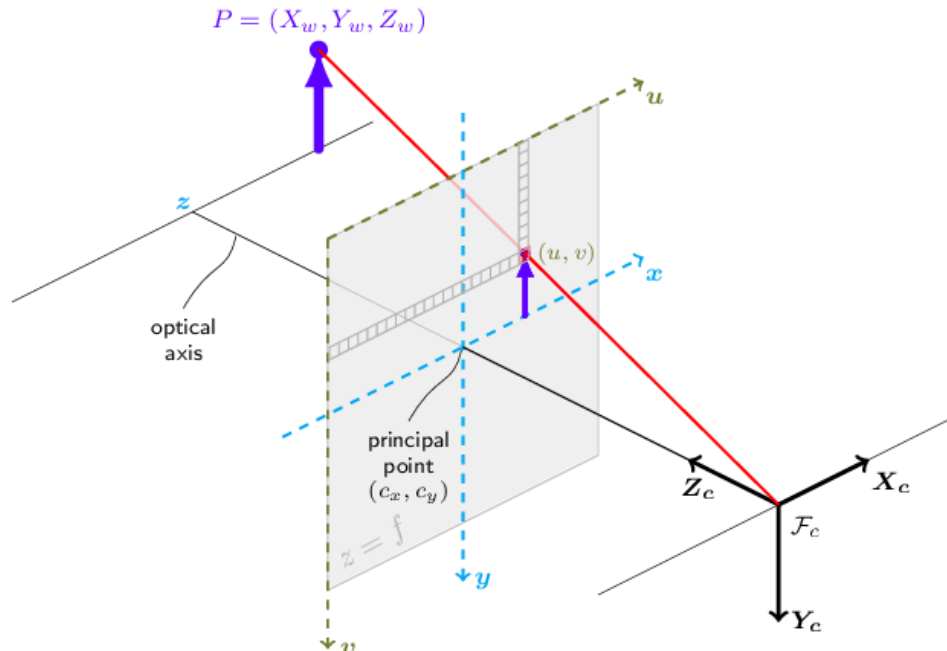


Figure 10: les systèmes de coordonnées [17]

La calibration de caméra permet de représenter un point (x,y,z) dans le monde réel en un pixel (x,y) sur une image [18].

La première transformation est la **transformation monde-caméra**.

Considérons un point (x_w, y_w, z_w) dans un système de référence fixe que l'on va appeler le système de coordonnées mondial et ce même point (x_c, y_c, z_c) dans le système de coordonnées de la caméra.

Cette transformation se fait grâce à une matrice 3x3 de rotation R et un vecteur 3x1 de translation t , comme dans la formule suivante :

$$\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = R_{3 \times 3} \begin{pmatrix} x_w \\ y_w \\ z_w \end{pmatrix} + t_{3 \times 1}$$

La matrice de rotation correspond à l'orientation de la caméra.

Il faut noter qu'il n'y a pas de conventions précises en photogrammétrie. Il y a deux types de systèmes de coordonnées, le gaucher et le droitier. Nous pouvons voir dans la figure 3.7, quand l'axe des X est orienté vers la droite et l'axe des Y vers le haut, un système de coordonnées gaucher va avoir l'axe des Z positif qui, visuellement, s'éloigne, tandis que pour le droitier, il va se rapprocher.

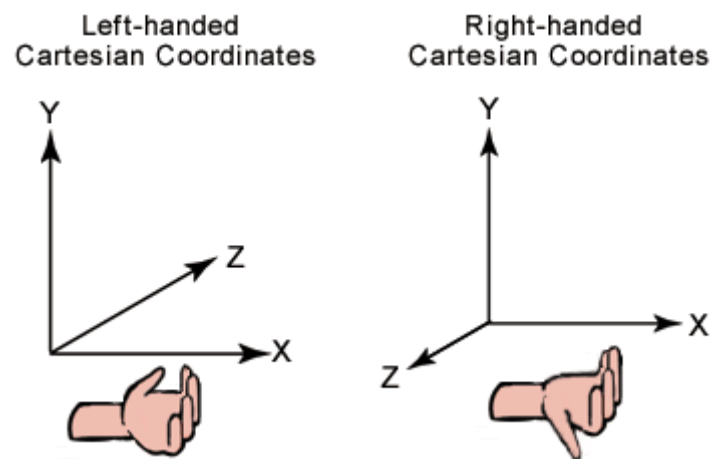


Figure 11: les systèmes de coordonnées droitier et gaucher [19]

Chaque logiciel de vision 3D choisit le système de coordonnées qui lui convient. Dès lors, lorsqu'on transfère des données entre logiciels, il faut parfois effectuer une conversion.

Pour Sphaeroptica, nous avons choisi de tourner l'axe des Z dans la direction de la caméra, l'axe des X vers la droite et l'axe des Y vers le bas. Cette orientation des axes de caméra est identique à celle utilisée par le logiciel Agisoft Metashape et OpenCV.

Le vecteur de translation, quant à lui, correspond à la position de la caméra. C'est le déplacement entre le centre du système de référence et le centre de la caméra, après rotation des axes.

Cette transformation du système de coordonnées de référence à celui de la caméra se fait grâce à la matrice 4x3 extrinsèque.

Ensuite, il faut effectuer la **transformation caméra-image**. Celle-ci est irréversible car c'est celle d'un plan en 3D vers un plan en 2D. En effet, la notion de profondeur est perdue avec cette transformation.

Il faut d'abord présenter le plan d'image, perpendiculaire aux axes X et Y et à distance focale f du centre de caméra.

Le système de coordonnées de l'image est un système euclidien en 2D centré au point principal, l'intersection entre le plan d'image et l'axe principal de la caméra [18].

Cette transformation du point (x_c, y_c, z_c) dans le système de coordonnées de l'image est donnée par les équations suivantes :

$$\begin{aligned}x_i &= f_x * x_c / z_c \\ y_i &= f_y * x_c / z_c\end{aligned}$$

Enfin, pour la **transformation image-pixel**, on peut trouver le pixel (u,v) grâce aux équations suivantes :

$$\begin{aligned}u - c_x &= s_x x_i \\ v - c_y &= s_y y_i\end{aligned}$$

avec (c_x, c_y) le point principal et s_x, s_y les signes des axes entre les conventions du système de coordonnées de l'image et celui du pixel.

Dans la figure présentant les systèmes de coordonnées, les deux valeurs seront positives. Ces deux dernières transformations sont effectuées grâce à la matrice 3x3 intrinsèque.

3.1.3 Distorsion de la lentille

Le calcul présenté ci-dessus aurait été suffisant si la lentille de la caméra était "parfaite". Malheureusement, avec l'industrialisation de la production de caméras, chaque caméra apporte une distorsion plus ou moins grande sur l'image, surtout en fonction de sa qualité. Celle-ci se présente de deux manières :

- La distorsion radiale
- La distorsion tangentielle

Premièrement, la **distorsion radiale** est une distorsion où les bords sont plus déformés que les pixels autour du point principal.

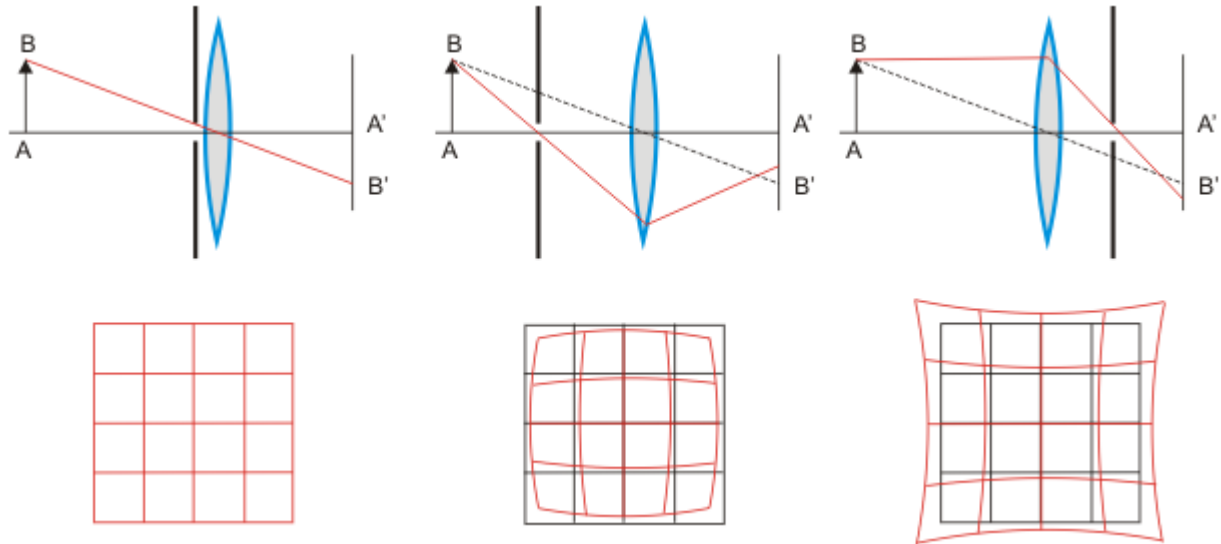


Figure 12: : la distorsion radiale[20]

Dans l'ordre, on peut voir, une image sans distorsion, une image avec une distorsion **avec barillet**, et une image avec une distorsion **avec coussinet**

La déformation radiale se calcule de la manière suivante :

$$x_d = x_u * \frac{1+k_1r^2+k_2r^4+k_3r^6}{1+k_4r^2+k_5r^4+k_6r^6}$$

$$y_d = y_u * \frac{1+k_1r^2+k_2r^4+k_3r^6}{1+k_4r^2+k_5r^4+k_6r^6}$$

avec (x_u, y_u) le pixel sans distorsion de caméra, (x_d, y_d) le pixel avec distorsion, les valeurs $k_1, \dots, k_6$ qui sont les valeurs de distorsions radiales et r qui est la distance entre le pixel et le point principal :

$$r^2 = x_u^2 + y_u^2$$

C'est la raison pour laquelle la distorsion de la caméra est non linéaire.

Deuxièmement, la **distorsion tangentielle** est un défaut dans la fabrication de la caméra. C'est quand le plan d'image n'est pas parallèle à la lentille.

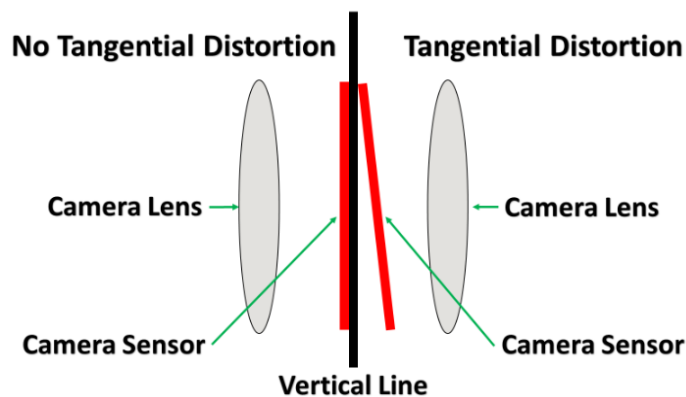


Figure 13: la distorsion tangentielle [21]

Elle se calcule de cette façon :

$$\begin{aligned}x_d &= x_u + 2p_1x_u y_u + p_2(r^2 + 2x_u^2) \\y_d &= y_u + 2p_2x_u y_u + p_1(r^2 + 2y_u^2)\end{aligned}$$

avec p_1 et p_2 , les valeurs de la distorsion tangentielle.

Combinés cela donne ces deux équations :

$$\begin{aligned}x_d &= x_u * \frac{1+k_1r^2+k_2r^4+k_3r^6}{1+k_4r^2+k_5r^4+k_6r^6} + 2p_1x_u y_u + p_2(r^2 + 2x_u^2) \\y_d &= y_u * \frac{1+k_1r^2+k_2r^4+k_3r^6}{1+k_4r^2+k_5r^4+k_6r^6} + 2p_2x_u y_u + p_1(r^2 + 2y_u^2)\end{aligned}$$

On peut voir dans ces équations que huit paramètres sont utilisés pour calculer la distorsion, six pour la distorsion radiale et deux pour celle tangentielle. Tout paramètre non donné par la calibration peut être simplement ignoré et retiré.

Par exemple, nous utilisons les valeurs fournies par Agisoft Metashape dans le format OpenCV qui n'en comprend que cinq : k_1 , k_2 , p_1 , p_2 et k_3 .

Les équations seront alors les suivantes :

$$\begin{aligned}x_d &= x_u * (1 + k_1r^2 + k_2r^4 + k_3r^6) + 2p_1x_u y_u + p_2(r^2 + 2x_u^2) \\y_d &= y_u * (1 + k_1r^2 + k_2r^4 + k_3r^6) + 2p_2x_u y_u + p_1(r^2 + 2y_u^2)\end{aligned}$$

3.2 Méthodes de calibration

Dans la section précédente, nous avons vu le modèle de caméra à sténopé avec distorsion radiale et tangentielle.

Maintenant, nous allons présenter deux méthodes différentes pour obtenir une calibration de caméra suffisamment précise pour une bonne photogrammétrie des arthropodes.

Les méthodes suivantes ont été utilisées lors de ce mémoire pour la calibration des différents sets d'images d'arthropodes :

- Calibration avec une mire 2D/3D
- Structure from motion

Dans un premier temps, nous nous focaliserons sur la théorie et les formules de ces méthodes, leur utilisation dans Sphaeroptica étant expliquée lors de la présentation du logiciel.

3.2.1 Calibration avec une mire 2D/3D

Cette première méthode de calibration de caméra est une méthode très précise qui consiste à prendre plusieurs images avec une mire en 2D/3D qui a un motif connu.

Une mire connue permet de trouver des correspondances entre les images et des coordonnées précises du monde.

Malheureusement, ce genre de calibration est souvent très couteux car elle demande une préparation préalable, le développement d'un logiciel précis pour chaque mire différente utilisée.

Toutefois, une méthode très flexible a été développée selon cette base. Celle-ci consiste à prendre les images avec un pattern très connu, un plateau d'échecs [22].

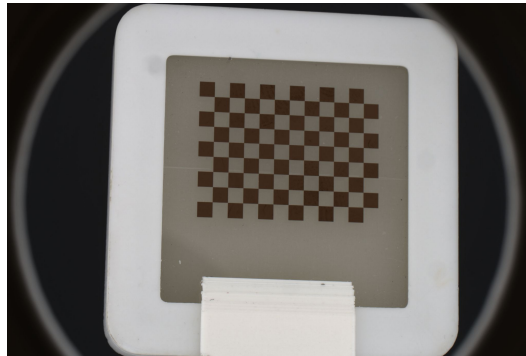


Figure 14: pattern 12x9 de carrés de 2mm de côté, prise par notre implémentation scAnt

Le processus tel que présenté par Zhang est très simple :

1. Imprimer le motif d'un plateau d'échecs sur une surface plane.
2. Calculer les dimensions de ce motif, la taille des cases et le nombre de lignes et colonnes
3. Détecter les intersections des rectangles sur chacune des images
4. Estimer les paramètres intrinsèques et extrinsèques des caméras ainsi que les coefficients de distorsions
5. Affiner les résultats en minimisant les erreurs de reprojection.

Cette méthode est très populaire car elle ne nécessite pas beaucoup d'efforts et de coûts pour fonctionner correctement.

En effet, simplement imprimer un motif avec une imprimante et coller la page sur une surface solide comme la couverture d'un livre est suffisant.

3.2.2 Structure-from-motion

Le SfM permet une reconstruction de modèle avec des images non ordonnées et dont la géométrie de l'objet n'est pas connue [10]. Cette méthode consiste à déterminer la structure d'un objet grâce au mouvement relatif de la caméra observatrice à l'objet, que le mouvement vienne de la caméra ou de l'objet.

Le SfM permet de déterminer la position des caméras et de donner un nuage de point éparpillé.

Pour pouvoir reconstruire un modèle 3D, on ajoute dans le processus un algorithme de MVS. Comme seule la calibration des caméras est nécessaire pour Sphaeroptica, nous n'approfondirons pas cette partie-là.

Différents types de SfM ont été développés :

- Le SfM incrémental [10] qui consiste à ajouter les images une par une dans la scène ;
- Le SfM hiérarchique [23] qui crée une scène par paire d'images et qui les combine une à une jusqu'à former la scène complète ;
- Et le SfM global [24] qui calcule la pose de toutes les caméras en même temps.

Les caractéristiques importantes pour un SfM sont les suivantes :

- Sa robustesse ;
- Sa complétude ;
- Sa précision ;
- Son évolutivité.

Le SfM incrémental est la méthode la plus populaire car elle coche toutes ces cases, sauf celle de l'adaptabilité. C'est cette méthode qui va être présentée dans ce chapitre pour démontrer les différentes étapes d'un pipeline SfM.

Pour le bon fonctionnement d'une technique SfM, il faut que les hypothèses émises au début de l'algorithme ou pendant son processus soient les plus correctes possibles.

En effet, une mauvaise initialisation peut fausser l'entièreté du modèle.

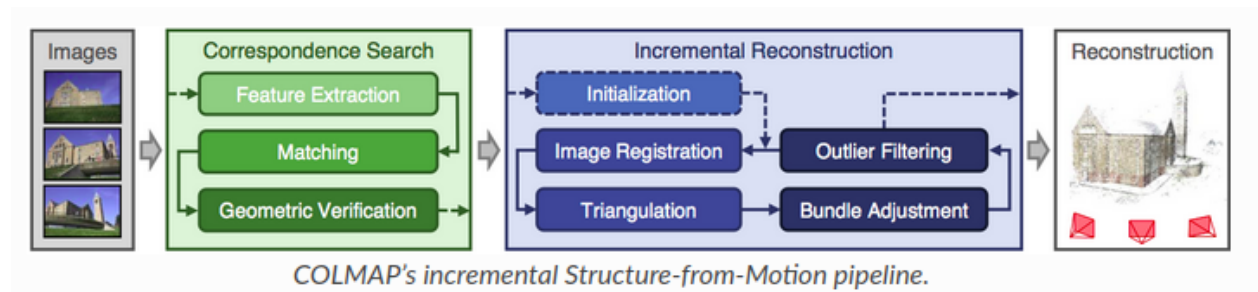


Figure 15: implémentation du SfM incrémental par COLMAP [10]

Le pipeline de SfM est séparé en 2 étapes distinctes :

- La recherche de correspondances
- La reconstruction incrémentale

Recherche de correspondances

Cette première étape cherche à trouver des points sur chacune des images et à trouver plusieurs correspondances sur des images qui se chevauchent.

Il faut effectuer, en premier, une extraction des pixels caractéristiques de l'image, aussi appelée en anglais **feature extraction**.

Cela consiste à trouver tous les points qui sont reconnaissables sur les images et pour chaque pixel caractéristique, retenir son descripteur. Par exemple, si on prend un simple dé avec une couleur par face, on peut voir qu'il est difficile de différencier deux points sur la même surface. Toutefois, chacun des coins de ce dé est reconnaissable facilement.

Les algorithmes [25] utilisés pour trouver ces pixels caractéristiques sont :

- le Scale-Invariant Feature Transform (SIFT) [26], un algorithme très robuste qui est l'état de l'art dans la feature extraction ;
- l'Oriented FAST and rotated BRIEF (ORB) [27], un algorithme plus rapide ;
- ...

Ensuite, il faut trouver le plus de correspondances entre les caractéristiques d'une paire d'images. Cela s'appelle le **feature matching**.

Cela consiste, pour chaque caractéristique, de trouver les correspondances les plus probables sur l'autre image grâce à une comparaison des deux descripteurs.

Enfin, parce qu'il est possible de trouver plusieurs features similaires sur plusieurs images, cela peut fausser les correspondances trouvées. Une **vérification géométrique** sera effectuée pour détecter les valeurs aberrantes.

Pour les éliminer, il faut utiliser des algorithmes comme le RANdom SAmple Consensus (RANSAC) [28].

Avoir un certain nombre de correspondances correctes permet le calcul de matrices de transformations entre la paire d'images, comme la matrice d'homographie H , la matrice fondamentale F et la matrice essentielle E .

La reconstruction incrémentale

La reconstruction incrémentale a pour but de déterminer les positions des caméras pour les images, et le nuage de points de la scène.

Pour cela, on commence par **initialiser** la scène avec la meilleure combinaison de deux images possible pour commencer la reconstruction. Ensuite, les images vont être ajoutées à la scène, une par une, pour pouvoir la compléter.

Pour ce faire, il faut effectuer un **recalage** de la prochaine image. Ce processus s'appelle en anglais **image registration**.

En se basant sur cette image, le processus va ajouter une caméra sur la scène en définissant sa position. Pour cela, il faut résoudre le problème Perspective-n-Point (PnP) [29] en comparant les caractéristiques de l'image avec les points 3D déjà calculés dans les scènes. Parce que toute valeur aberrante faussera la position de la caméra, il faut réutiliser un algorithme comme RANSAC pour les éviter. Une erreur de calibration pour cette caméra provoquera un effet boule de neige par la suite vu le côté incrémental du processus.

Dès que la position de caméra est déterminée, on peut ajouter de nouveaux points dans la scène, grâce à la **triangulation** [30].

Une triangulation de points est possible quand, pour un même point dans la scène, on peut retrouver au minimum deux images différentes qui l'enregistrent. Dans cette étape, le programme comparera les points sur l'image sélectionnée sur toutes les images déjà placées. Un algorithme de triangulation sera expliqué plus en détail dans le chapitre suivant.

Enfin, à chaque incrémentation, il faut réajuster toutes les positions des caméras.

En effet, avec la présence de bruit dans les images, la triangulation ne sera jamais parfaite. Cette partie s'appelle le **bundle adjustment**. Elle consiste à calculer les paramètres de chacune des caméras qui réduisent au maximum l'erreur de reprojection.

Dans cette section, on a pu voir que donner des approximations correctes à un programme SfM est primordial pour récupérer une scène précise.

Toutefois, ce n'est pas le seul critère nécessaire à un bon fonctionnement de cette technique. Une bonne résolution des images est également nécessaire pour récupérer un modèle précis. De plus, les objets avec un manque de texture peuvent aussi compliquer le bon fonctionnement du SfM.

Ce chapitre “Calibration de caméra” s’est concentré sur les éléments théoriques nécessaires à la calibration. Son application pratique pour les besoins de Sphaeroptica sera développée dans le chapitre suivant.

4 Sphaeroptica

La principale limitation des modèles surfaciques générés par les logiciels cités auparavant est la perte de résolution lors de la création du modèle. Cela peut être problématique pour les objets qui sont petits, qui ont des angles fins ou qui ont énormément de détails. Ce qui est le cas pour les arthropodes.

C'est pour cette raison que Sphaeroptica a été développé : donner la possibilité aux scientifiques d'effectuer des mesures 3D sur un support ayant la meilleure résolution possible, c'est-à-dire les images 2D elles-mêmes.

Dans ce chapitre, nous vous présenterons les différentes étapes d'une numérisation avec Sphaeroptica, du lancement du scanner à la prise des mesures.

Sphaeroptica est un logiciel photogrammétrique de visualisation 360° d'un spécimen avec des images déjà calibrées. Ce logiciel permet également la prise de mesures grâce à une triangulation des repères posés par l'utilisateur sur les images.

Les spécimens présentés dans ce chapitre sont des arthropodes.

En effet, l'IRSNB m'a proposé ce mémoire afin de mettre en place un système de numérisation approprié pour leur collection d'arthropodes.

4.1 Acquisition des images

Comme déjà expliqué plus haut, l'acquisition des images s'est faite au moyen d'un dispositif scAnt assemblé au sein de l'institut. La prise de photos s'est faite avec une caméra 20 MPX. Mais ce n'est pas un requis de l'application, tout objet, petit ou grand, dont des images ont été prises autour de celui-ci peut être sélectionné.

Le stacking d'images EDOF, fonctionnalité de scAnt, n'a pas été utilisé pour la prise des images. La raison de ce choix est que le stacking modifie les images prises et toute modification ajoute des erreurs géométriques sur les images. Pour m'assurer de la précision des mesures, nous avons préféré opter pour l'utilisation des images brutes.

Nous avons choisi de prendre 180 images, à partir de 9 rotations verticales et 20 rotations horizontales. Vu le choix de ne pas utiliser le focus stacking (EDOF), la caméra ne doit pas se déplacer pour chaque prise de vue. Cela réduit considérablement le temps du scan. Une prise complète dure environ de 5 à 10 minutes avec ce paramétrage de scAnt.

Il n'est pas utile de réduire la qualité de la caméra pour gagner en rapidité.

En effet, le temps de calcul pour aligner les images n'est que d'une dizaine de minutes. Cette durée est minime comparée au temps nécessaire pour créer un seul modèle 3D qui dure plus de 5 heures.

4.2 Calibration

Après chaque acquisition des images, il faut calibrer les caméras de la scène.

Une mauvaise calibration entraînerait des erreurs dans les mesures prises sur Sphaeroptica, soit des erreurs de reprojection, soit des erreurs d'échelle.

- Les erreurs de reprojection sont des erreurs d'incertitude du calcul du point 3D. C'est la distance entre les valeurs de repères choisis pour la triangulation et celles de la projection du point 3D.
- Les erreurs d'échelle surviennent quand les proportions de l'objet sont respectées mais que la distance des caméras par rapport à l'objet n'a pas été exactement respectée. Cela fait qu'un millimètre dans le monde réel n'équivaut pas à un millimètre dans la scène.

Il faut savoir qu'un énorme avantage des scanners est leur haute reproductibilité, c'est-à-dire le fait qu'il soit possible de reproduire identiquement ou presque les mêmes positions de caméra à chaque scan.

Une seule calibration est donc nécessaire pour l'entièreté du système, et si aucune valeur ne change, il est possible de réutiliser les mêmes données de calibration pour chaque acquisition d'images.

La méthode optimale pour cette calibration initiale aurait été une calibration de la matrice intrinsèque grâce au motif de Zhang suivi d'une calibration extrinsèque avec une résolution du Perspective-n-point (PnP) avec une mire 3D. Malheureusement, l'utilisation de la mire 3D n'a pas été possible. La raison est expliquée ci-dessous et une alternative a été adoptée.

Pour la calibration intrinsèque, nous avons utilisé une mire contenant ce pattern en 12x9 avec des carrés de 2mm de côté. Pour ce faire, nous avons développé une fonctionnalité de Sphaeroptica pour la calibration avec le motif de Zhang, en ayant recours à la librairie OpenCV.

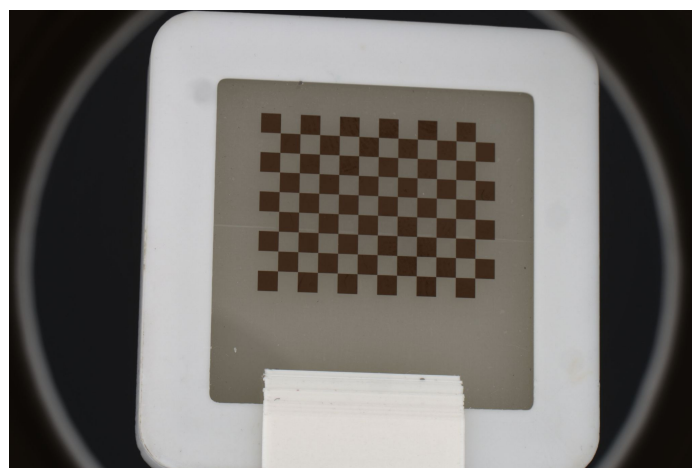


Figure 16: image utilisée pour la calibration intrinsèque

Concernant la mire 3D, un prototype a été construit avec des Lego. Malheureusement, par manque de temps, il n'a pas été possible de développer une solution rapide et efficace pour

faire automatiquement les correspondances 2D-3D sur les images. Ce sera un travail à effectuer plus tard pour avoir des mesures plus précises.

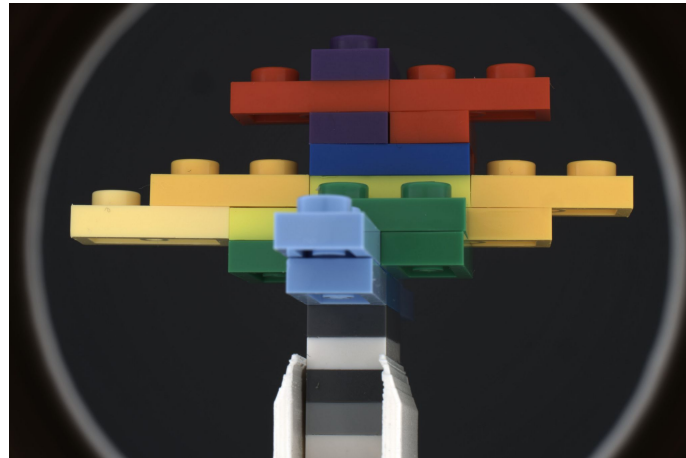


Figure 17: image de la mire 3D construite en Lego

En attendant, nous avons quand même dû trouver une solution pour définir les valeurs extrinsèques des 180 caméras.

La méthode présentée par Zhengyou Zhang permet également de calculer les vecteurs de rotations et de translation, donc les paramètres extrinsèques, pour toutes les images où le motif a été reconnu. Cela n'est pas complet car cela ne peut pas concerner l'entièreté des images d'une sphère 360°, mais cela peut donner une estimation du centre de rotation du monde par rapport aux images, ce qui est déjà un bon début. Ensuite, il a été possible d'estimer la distance moyenne entre les centres des caméras et le centre de rotation pour pouvoir extrapoler la position de chacune des caméras du scan grâce à cela.

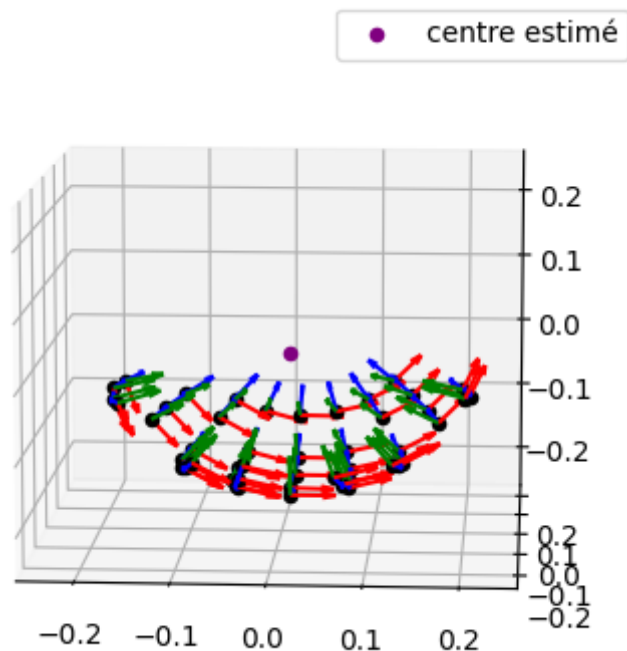


Figure 18: calibration des images reconnues avec le motif de Zhang

Un autre élément important à prendre en compte pour la calibration des caméras est le fait que la prise d'images avec scAnt peut amener quelques incertitudes. Ces dernières sont dues à une position pas toujours précise de l'arthropode dans le scanner, ou encore par des vibrations pouvant faire bouger l'insecte de manière imprévisible.

La calibration définie ci-dessus ne pouvant tenir compte de ces conditions, nous avons opté pour un affinage des résultats de la calibration initiale à chaque scan. Cet affinage a été effectué par le logiciel Agisoft Metashape.

4.3 Triangulation

La triangulation permet d'estimer des points 3D avec pour objectif d'effectuer des mesures entre ces différents points [30], [15, p. 312].

Pour effectuer un calcul de distance, l'entomologiste qui a besoin de ses mesures va devoir choisir 2 points 3D de l'arthropode. Ensuite, pour chaque point 3D, il doit sélectionner un pixel de repère sur deux images différentes (au minimum). La triangulation en a besoin pour estimer le point 3D. C'est ce processus qui est expliqué ci-dessous.

Lors de la projection du monde en 3D sur le plan d'image en 2D, un élément important a été retiré : la profondeur.

Par conséquent, il est impossible de trouver la position exacte dans un monde en 3D d'un point désigné sur l'image en fonction de sa distance et du centre de projection.

En effet, désigner un pixel sur une image calibrée permet seulement de donner un vecteur de direction ayant pour point initial le centre de la caméra et traversant ce pixel dans la direction de l'image. Il existe donc une infinité de solutions sur ce rayon.

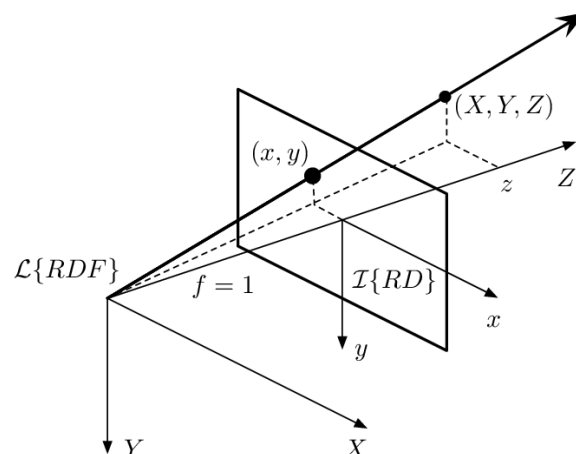


Figure 19: back projection du pixel en rayon [31]

Une triangulation à partir de plusieurs repères venant d'images différentes est donc nécessaire.

La triangulation est le calcul pour trouver le point dans l'espace 3D venant de l'intersection des rayons connus pour ce point.

Malheureusement, à cause du bruit présent sur les images, il y a rarement une intersection exacte.

Par conséquent, il faut trouver le point 3D le plus proche, grâce à une minimisation des moindres carrés. C'est-à-dire trouver le point 3D dont les projections sur les images sont les plus proches par rapport aux pixels posés.

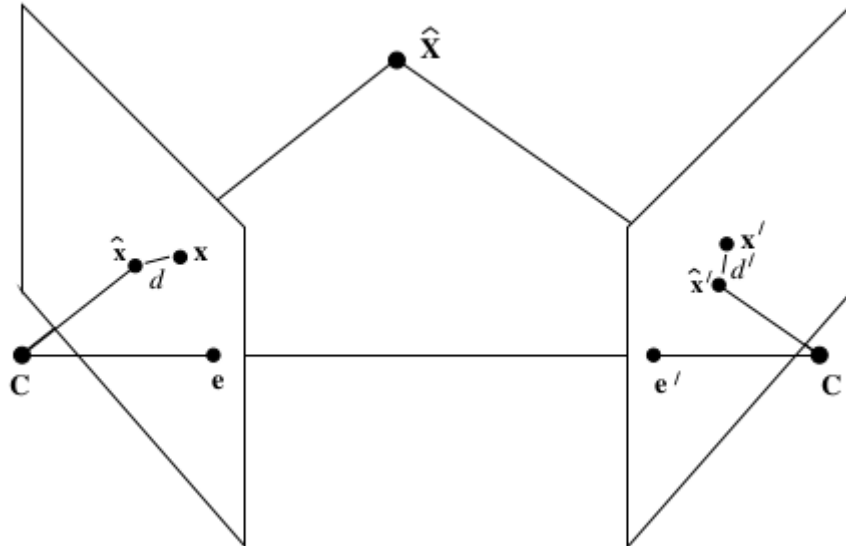


Figure 20: triangulation avec minimisation de l'erreur géométrique [15, p. 314]

Pour résoudre cette problématique de minimisation des moindres carrés, nous allons utiliser l'algorithme du Direct Linear Transform (DLT) qui permet de résoudre les équations de type $AX = 0$ A est une matrice $m \times n$ et X une matrice $n \times 1$ qui contient les valeurs à trouver [15, p. 312], [32], [33].

Pour déterminer la matrice A, nous savons qu'en théorie, $x = PX$ pour un point dans le monde X et la matrice d'une caméra calibrée P. Cela veut aussi dire que $x \times PX = 0$.

Si on considère que P^i est la i-ème ligne de la matrice P, pour toute image i, nous avons la formule suivante :

$$x_i \times P_i X = \begin{pmatrix} y_i P^3 X - P^2 X \\ P^1 X - x_i P^3 X \\ x_i P^2 X - y_i P^1 X \end{pmatrix}$$

Parce que la 3e ligne de cette matrice est une combinaison linéaire des 2 premières, on peut l'enlever et nous récupérons l'équation suivante :

$$A = \begin{pmatrix} y_i P_i^3 X - P_i^2 X \\ P_i^1 X - x_i P_i^3 X \\ \dots \end{pmatrix} = 0$$

Il faudra simplement ajouter deux équations pour chaque nouvelle vue ajoutée (c'est-à-dire pour chaque vue ayant un repère posé pour ce point).

4.3.1 Direct Linear Transform

Maintenant que la matrice A a été calculée, on peut résoudre le DLT grâce aux valeurs singulières de la matrice A. Pour que la solution optimale puisse être calculée, il faut que la Matrice A m x n ait plus d'équations que de valeurs à calculer : m >= n.

Nous présentons les algorithmes suivants tels qu'ils le sont par Hartley et Zisserman dans l'annexe 5 de "Multiple View Geometry" à la page 588.

Singular Value Decomposition

Toute matrice rectangulaire m x n A peut être décomposée en une matrice orthogonale m x m U, une matrice diagonale m x n D et une matrice orthogonale n x n V.

$$A = UDV^T$$

Toute valeur singulière présente sur la diagonale de D tel que $\sigma_i = D_{ii}$, sera non négative et généralement présenté de manière décroissante :

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \dots \geq \sigma_n$$

Algorithme quand Ax = 0

Dans le cas où il faut minimiser $\|Ax\|$ où $\|\cdot\|$ est la norme de la matrice, la valeur optimale de x sera égale à la dernière colonne de V dans $A = UDV^T$.

Algorithme quand Ax = b

Nous cherchons à trouver la matrice x, tel que $\|Ax - b\|$ est minime.

Tout d'abord, nous allons décomposer la matrice A avec son SVD : $\|Ax - b\| = \|UDV^T - b\|$.

Sachant que les matrices U et V sont orthogonales, et que les matrices orthogonales préservent la norme, tel que pour une matrice orthogonale H et une matrice quelconque A, $\|HA\| = \|A\|$, nous pouvons transformer cette équation de cette manière : $\|UDV^T - b\| = \|DV^T x - U^T b\|$.

Notons $y = V^T x$ et $b' = U^T b$ et nous avons une nouvelle équation dont il faut minimiser le résultat $\|Dy - b'\|$.

La formule se présente de la manière suivante :

$$\begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ & & \dots & \\ 0 & 0 & \dots & \sigma_n \\ \hline & & & 0 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} b'_1 \\ b'_2 \\ \dots \\ b'_n \\ \hline b'_{n+1} \\ \dots \\ b'_m \end{pmatrix}$$

Il est possible de calculer la matrice y telle que $y_i = b'_i/\sigma_i$ pour tout $i \leq n$.
Après avoir calculé la matrice y, nous pouvons trouver la matrice x tel que $x = Vy$.

4.3.2 Erreur de reprojection :

Pour déterminer la précision de la triangulation effectuée, nous calculons les erreurs de reprojection. C'est la différence entre la projection d'un point 3D sur les images et ses repères. Plus le résultat est bas, plus la triangulation a pu être précise.

$$\sum_i d(x_i, \hat{x}_i)^2$$

Figure 4,6 : formule de l'erreur de reprojection

Cela permet d'avoir un certain intervalle de confiance sur le résultat triangulé. Et c'est également une métrique utilisée pour déterminer la précision de la calibration d'une caméra par rapport à la scène.

La triangulation permet d'estimer des points 3D qui serviront, par exemple, dans des calculs de distances, d'angles... Puisque Sphaeroptica se veut simple, nous allons nous limiter à la comparaison de calcul de distances.

4.4 Comparaisons

Nous avons comparé les mesures prises par Sphaeroptica et le modèle de référence calculé par un scanner Artec Micro.

Pour cela, nous avons pris 3 arthropodes :

- Un **Geonomus goeffroi** (taille : 22,8577 mm)
- Un **Eudicella gralli** (taille : 33,672 mm)
- Une **mouche** (taille 27,34 mm)

Nous avons placé 6 points sur chacun de ces arthropodes :

- 2 points à l'avant et l'arrière pour pouvoir calculer la longueur
- 2 points sur les côtés pour pouvoir calculer la largeur
- 2 points sur le dos et le ventre pour pouvoir calculer la hauteur

Pour chaque arthropode, nous avons calculé les distances entre chacun de ces points sur le modèle de référence (15 distances) et les avons comparées aux mêmes mesures prises sur Sphaeroptica après une triangulation à 2-vues, 3-vues et 5-vues. Nous avons ainsi voulu déterminer si mettre plus de repères permettait d'avoir plus de précision ou non.

Il est important de noter que cette comparaison comporte un risque d'erreur lié à la sélection manuelle des points.

Les mesures de comparaisons nous ont permis de :

- valider la précision de Sphaeroptica ;
- définir l'importance du nombre de points de repère à poser pour obtenir des résultats plus précis ;
- voir en quoi la calibration de caméra influence le résultat des calculs ;
- démontrer l'importance d'une bonne identification de points 3D pour assurer la fiabilité de la triangulation et des mesures effectuées, et donc, la nécessité d'une bonne qualité d'image, en particulier pour les arthropodes à structure complexe.

Les résultats de comparaison sont présentés dans les deux diagrammes suivants, l'un représentant les différences de mesures en valeur absolue et l'autre en valeur proportionnelle pour les 3 arthropodes sélectionnés, ainsi que pour l'ensemble des mesures prises ("aggregation").

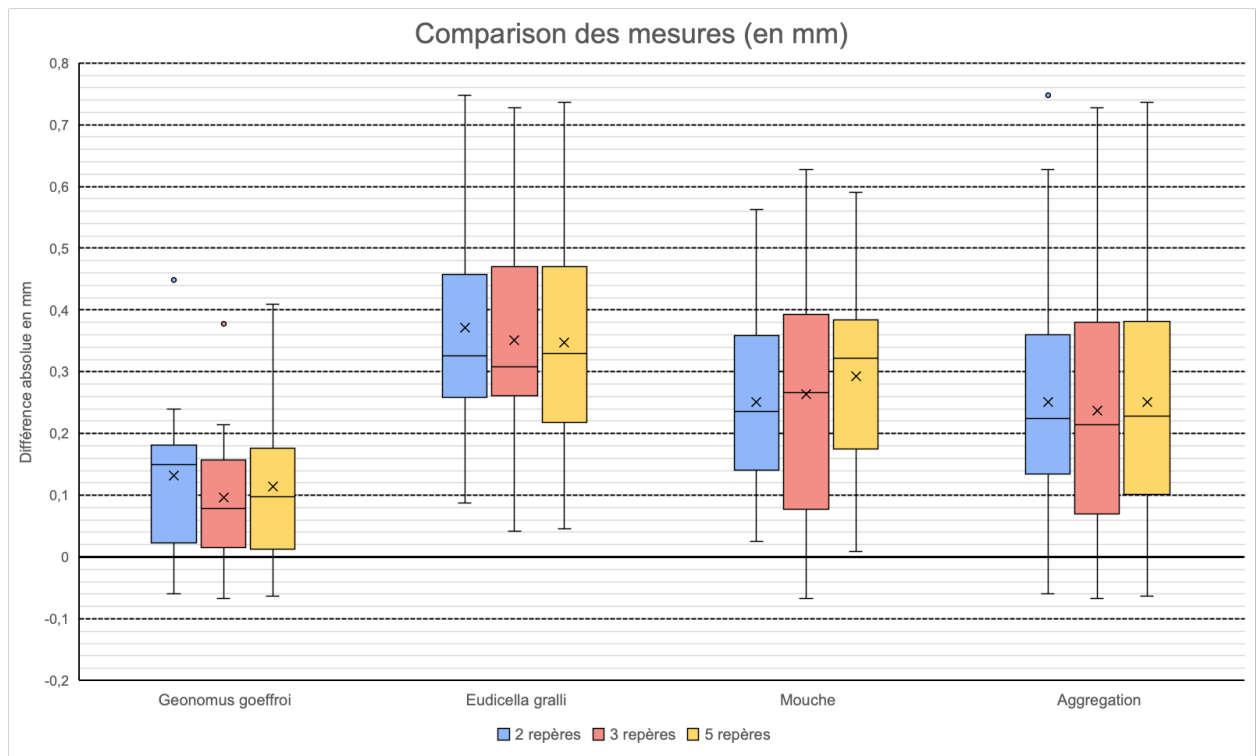


Figure 21: différences de mesures en mm

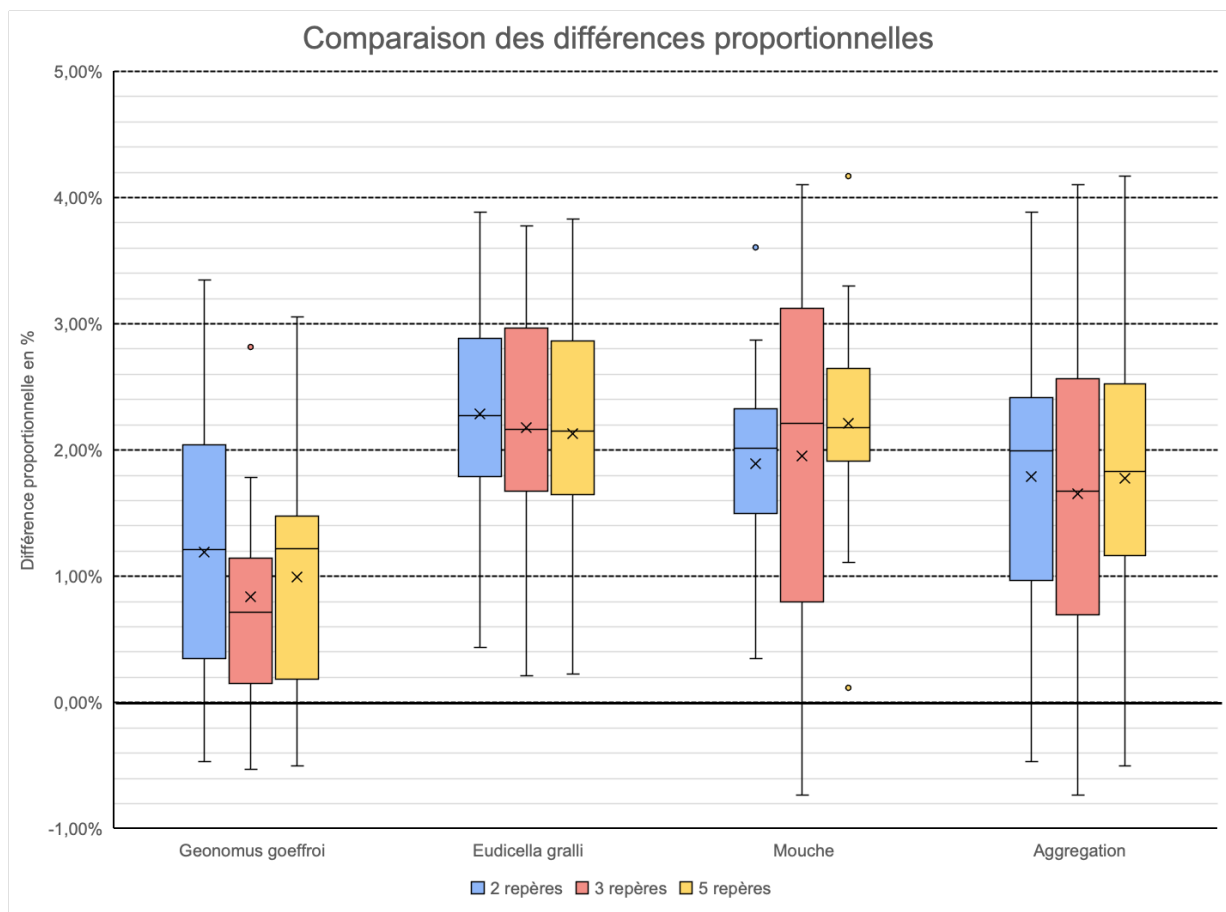


Figure 22: différences de mesures en %

Le tableau suivant offre un résumé des comparaisons effectuées :

Table 1: moyenne, médiane et écart type

	Différence mm 2 repères	Différence mm 3 repères	Différence mm 5 repères	Différence % 2 repères	Différence % 3 repères	Différence % 5 repères
Moyenne	0.2515503	0.2369903	0.2515919	1.79%	1.66%	1.78%
Médiane	0.2240929	0.2144973	0.2283230	1.99%	1.67%	1.83%
Ecart-type	0.1796953	0.1946265	0.1820768	1.02%	1.20%	1.06%

Sur la figure 4.6, nous voyons que la différence des mesures entre Sphaeroptica et les modèles de référence est aux alentours de 2,2 dixième de mm (valeur médiane).

Sur la figure 4.7, représentant ces mêmes mesures en valeur proportionnelle, on peut voir que Sphaeroptica offre une précision proche de 98% (valeur médiane) par rapport aux modèles de référence .

Nous observons également que la quasi-totalité des mesures sont plus grandes que leurs références. Notre hypothèse est que cela s'explique par la calibration du scanner qui n'a pas été optimale vu l'absence d'une calibration avec une mire 3D. En l'absence de paramètres extrinsèques exacts, la calibration calculée est juste à une échelle près. Cette échelle donne une imprécision qui est constante.

Pour valider cette hypothèse, il a fallu gommer les erreurs liées à l'échelle. Pour ce faire, nous avons travaillé avec des ratios. Par ratio, nous entendons les proportions "longueur - largeur", "longueur - hauteur", "largeur - hauteur". Ensuite, nous avons comparé les différences de ratio entre les mesures prises par Sphaeroptica et celles prises par le modèle de référence.

La table suivante montre les résultats de cette comparaison :

Table 2: différences de ratio

	Difference ratio 2 repères	Difference ratio 3 repères	Difference ratio 5 repères
Moyenne	0.97%	1.03%	0.55%
Médiane	0.85%	0.33%	0.57%
Ecart-Type	0.63%	1.25%	0.34%

En retirant l'échelle de l'équation, nous constatons que Sphaeroptica offre une précision supérieure à 99% (valeur médiane) par rapport aux modèles de référence.

Sur les quatre figures qui précèdent, l'étude des mesures prises sur les trois arthropodes montre que poser plus de repères est à double tranchant et par conséquent, n'équivaut pas forcément à une augmentation significative de la précision. En effet, un seul repère posé de manière très approximative peut rapidement fausser le calcul de triangulation.

Nous pouvons voir que la précision de Sphaeroptica a été meilleure pour le *Geonomus goeffroi* que pour la mouche. Cela s'explique par le fait que la mouche ayant une texture complexe, il a été difficile de trouver des points facilement identifiables sur le modèle de référence et sur les images, empêchant une pose très précise des repères.

Le calcul des erreurs de projections permet d'évaluer la confiance qu'on peut avoir dans les mesures prises. Le tableau suivant donne les erreurs de reprojection après la triangulation des points 3D. Plus l'erreur est basse, plus la triangulation a pu être précise.

Table 3: erreur de reprojection sur les arthropodes mesurés

	<i>Geonomus goeffroi</i>	<i>Eudicella gralli</i>	Mouche	Moyenne
rep error 2 repères	1.404679614	2.163723868	3.290482152	2.286295211
rep error 3 repères	2.009250319	2.413773756	6.603617319	3.675547131
rep error 5 repères	2.248383858	2.486028169	6.685717499	3.806709842

Nous remarquons que pour la mouche, plus il y a de repères posés, plus l'erreur de reprojection est conséquente. Elle est presque trois fois plus grande que pour les autres arthropodes. Ceci est dû à l'imprécision des repères posés à cause de la texture de la mouche.

Ce dernier phénomène a été encore plus visible lors de la prise de mesure sur une deuxième mouche. Cette mouche ayant des poils plus longs, il fut encore plus difficile de distinguer des points uniques, à la fois sur les images prises, et sur le modèle.

Les figures suivantes montrent les résultats obtenus en comparant les mesures prises sur le modèle de référence et sur Sphaeroptica :

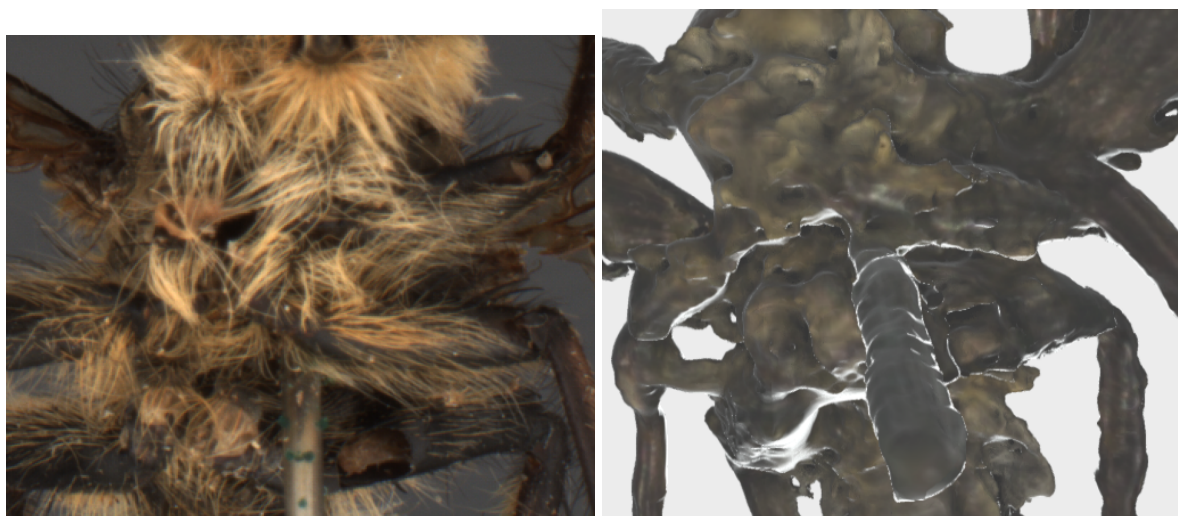


Figure 23: image et modèle 3D de la deuxième mouche

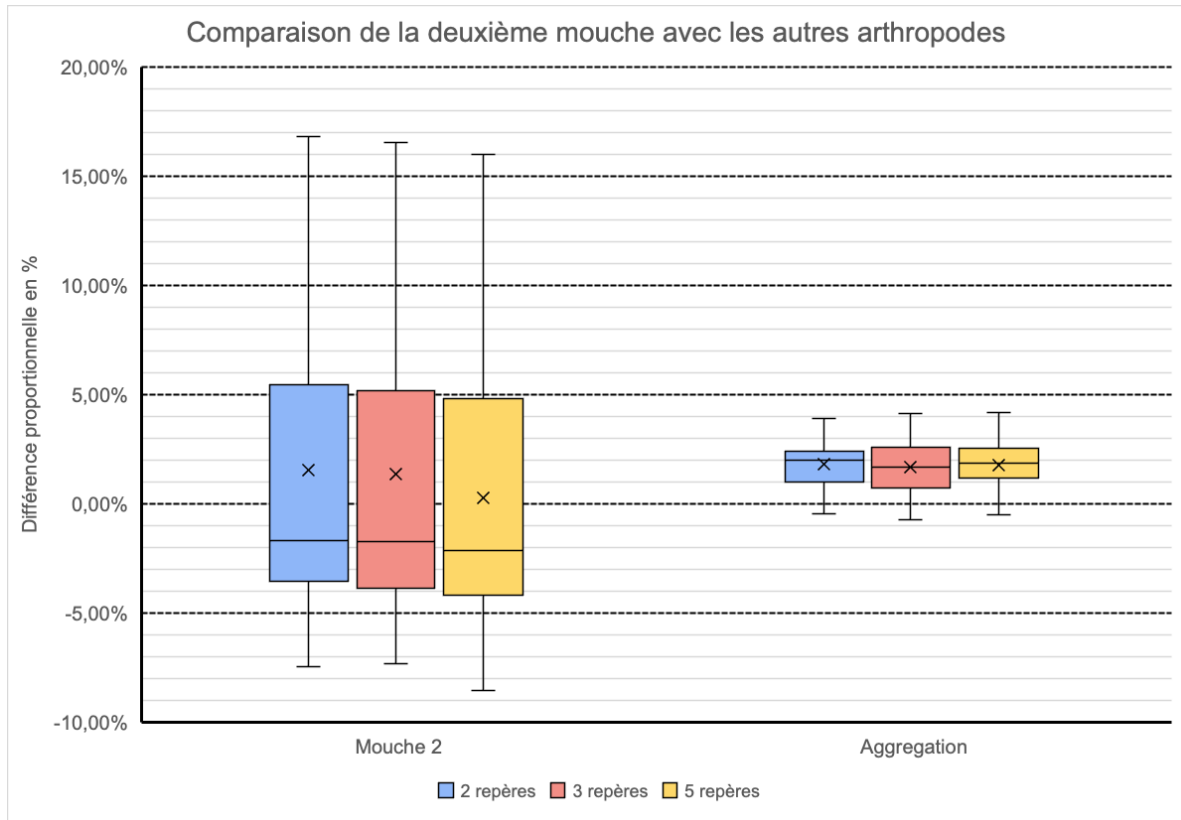


Figure 24: comparaison des différences observées entre la deuxième mouche et le restes des arthropodes

Au vu des valeurs aberrantes, ces dernières n'ont pas été prises en compte dans les estimations de précisions. L'étendue des différences trouvées est trois à cinq fois plus grande que pour les autres arthropodes choisis.

Cette statistique peut s'expliquer par les raisons suivantes :

- Le modèle de référence n'étant pas lui-même suffisamment précis (voir figure 23), certains points n'ont pas pu être correctement sélectionnés ;
- Plusieurs points n'étaient pas identifiables de façon unique, ce qui a rendu la pose des repères plus approximative. Comme explicité plus haut, une pose approximative peut causer une grande erreur dans la triangulation.

En vérifiant l'erreur de reprojection, nous avons observé que celle-ci est très élevée, confirmant que les mesures prises par Sphaeroptica pour cette deuxième mouche ne devaient pas être prises en considération pour la comparaison avec le modèle de référence.

Nous pouvons conclure de ces calculs l'importance d'avoir une qualité impeccable pour les images pour les arthropodes ayant beaucoup de détails, comme les poils de cette mouche.

5 Caméra Virtuelle

Nous allons maintenant parler de la caméra virtuelle qui a été développée au sein de Sphaeroptica. Elle a pour but de pouvoir se déplacer de vue en vue de la manière la plus fluide et agréable possible :

- La caméra virtuelle est un plus par rapport au projet de base
- Elle n'a qu'un but visuel, recopiant au mieux le mouvement qu'on peut faire avec un modèle 3D
- Son but est en effet de pouvoir synthétiser de nouvelles vues pour donner un rendu comme pour des visualisations en 3D

Il est important de noter que toute image générée par la caméra virtuelle ne pourra pas être utilisée pour la triangulation. En effet cela amènerait des risques d'erreurs géométriques considérables.

La contrainte de ce projet a été de pouvoir avoir la meilleure précision possible grâce aux images et à leur grande résolution.

Tout d'abord, il a fallu trouver un moyen pour toujours sélectionner la bonne image en fonction de la position de la caméra virtuelle donnée.

Pour cela, grâce à la contrainte de la prise d'images autour d'un point central, nous avons une méthode utilisant des mesures de longitude et latitude.

Parce que la sphère d'images n'est pas totalement parfaite, certaines généralisations ont été effectuées pour contrer ces légères imperfections.

Le centre de la sphère a été calculé par une triangulation des points principaux de chacune des images. L'erreur de projection est d'environ 100 pixels, ce qui peut être conséquent lors de mesures géométriques. Cependant, puisque le but de cette caméra n'est que visuel, ce genre d'erreur est négligeable.

Pour décider des déplacements de la caméra virtuelle, nous allons utiliser les positions géodésiques (longitude et latitudes) autour d'une sphère.

Ensuite, une image de référence sera sélectionnée grâce à un calcul du plus proche voisin.

Enfin, il va falloir effectuer une rotation de cette image afin de donner un visuel approximatif de ce que verrait une caméra hypothétique. Cette rotation est effectuée grâce à un calcul de l'homographie entre la caméra calibrée de l'image de référence et la caméra virtuelle.

5.1 Calcul de la position géodésique

Pour générer des positions géodésiques, des n-vectors ont été utilisés [34].

Un n-vector est un vecteur unitaire représentant le vecteur sortant pour toute position sur un ellipsoïde donné. Sachant que l'ellipsoïde généré sera une sphère parfaite, le n-vector de chaque caméra sera simplement le vecteur normalisé entre la caméra et le centre de la sphère.

Le calcul du n-vector grâce aux positions géodésiques se fait de la manière suivante :

$$\mathbf{n}^e = \begin{bmatrix} \cos(\text{latitude}) \cos(\text{longitude}) \\ \cos(\text{latitude}) \sin(\text{longitude}) \\ \sin(\text{latitude}) \end{bmatrix}$$

Au contraire, pour calculer la latitude, voici deux équations presque équivalentes :

$$\text{latitude} = \arcsin(n_z^e) = \arctan\left(\frac{n_z^e}{\sqrt{n_x^{e2} + n_y^{e2}}}\right)$$

Parce que pour les ordinateurs, le calcul d'arcsin au niveau des pôles peut être imprécis, il est préférable d'utiliser l'équation de droite.

Pour la longitude, voici l'équation :

$$\text{longitude} = \arctan\left(\frac{n_y^e}{n_x^e}\right)$$

5.2 Sélection de l'image de référence

Maintenant que nous avons les positions des caméras autour d'une sphère et un moyen de calculer celle d'une caméra virtuelle, il faut trouver l'image de la caméra la plus proche de la position hypothétique.

Pour cela, nous allons calculer l'angle central entre la position virtuelle et chacune des caméras pour choisir la plus proche.

La formule est la suivante [35, p. 16]:

$$\Delta\sigma = \arccos(\sin\phi_1 \sin\phi_2 + \cos\phi_1 \cos\phi_2 \cos(\Delta\lambda))$$

avec ϕ_1 et ϕ_2 étant respectivement les latitudes des points 1 et 2 et $\Delta\lambda$ étant la différence absolue des longitudes de ces points.

5.3 Calibration de la caméra virtuelle

Pour calibrer une caméra virtuelle, il faut déterminer sa position et son orientation.

Ces deux valeurs vont être déterminées par les positions géodésiques calculées auparavant.

Pour trouver la position, nous utilisons la formule du calcul du n-vector pour trouver le vecteur de direction et multiplions celui-ci par la distance entre l'image de référence et le centre de la sphère

Concernant l'orientation de la caméra, nous devons tout d'abord effectuer une rotation initiale du système de coordonnées.

Les matrices de rotations utilisées, avec les angles ω , ϕ , κ , sont :

$$R_x = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\omega) & -\sin(\omega) \\ 0 & \sin(\omega) & \cos(\omega) \end{pmatrix} \quad R_y = \begin{pmatrix} \cos(\phi) & 0 & \sin(\phi) \\ 0 & 1 & 0 \\ -\sin(\phi) & 0 & \cos(\phi) \end{pmatrix}$$

$$R_z = \begin{pmatrix} \cos(\kappa) & -\sin(\kappa) & 0 \\ \sin(\kappa) & \cos(\kappa) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

avec R_x la matrice de rotation sur l'axe des X, R_y la matrice de rotation sur l'axe des Y et R_z celle sur l'axe des Z

Comme déjà dit dans le chapitre de la calibration de caméra, l'orientation de la caméra est la suivante : l'axe Z pointe vers la direction de la caméra, l'axe X pointe vers la droite et l'axe Y vers le bas. M

Pour avoir la rotation à la position (0,0), il faut d'abord effectuer une rotation de 90° sur l'axe des Y et ensuite une rotation de -90° sur l'axe des Z.

Nous appellerons cela la position de base, car cette rotation servira pour toutes les orientations autour de la sphère.

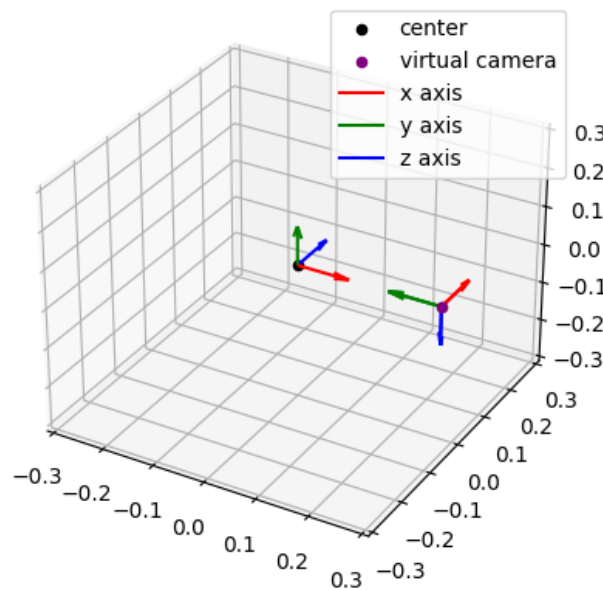


Figure 25: position de base de la caméra

Les rotations à effectuer pour arriver à cette position de bases sont les suivantes :

$$base_cam = R_z(-90).R_y(90)$$

Par après, la caméra subira une rotation sur l'axe des Y en fonction de la longitude puis une sur l'axe des X en fonction de sa latitude.

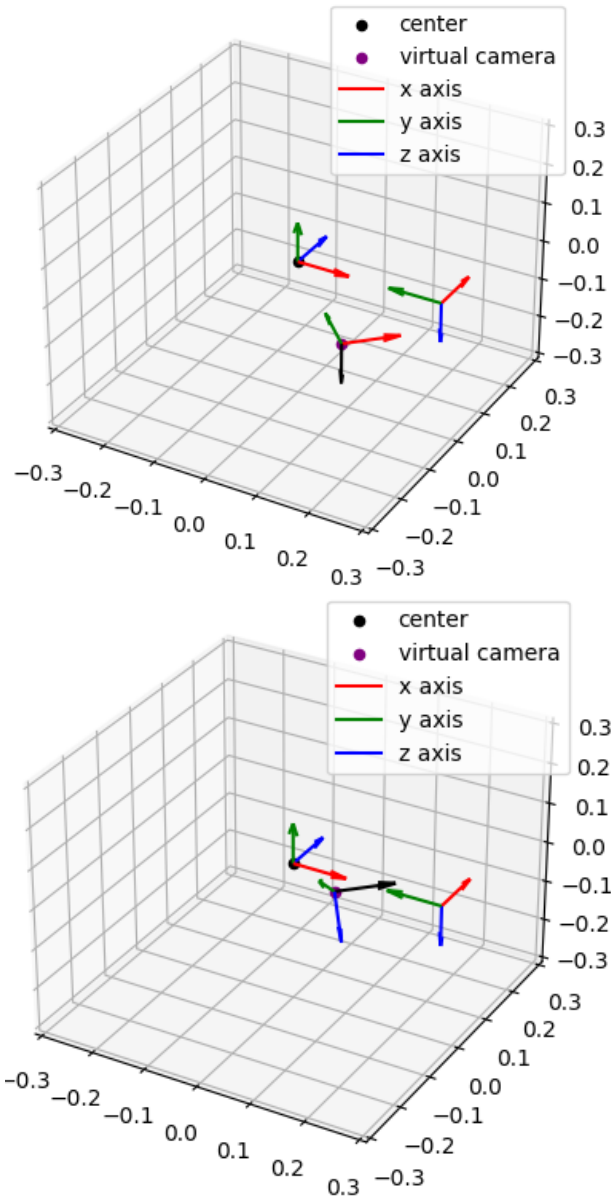


Figure 26: rotations effectuées pour calibrer la caméra virtuelle

Les dernières rotations nécessaires pour calculer la calibration extrinsèque de la caméra virtuelle sont :

$$cam = R_x(lat).R_y(long).base_cam$$

avec long et lat étant les valeurs de longitude et latitude.

5.4 Calcul de la matrice d'homographie H

Une homographie est une transformation inversible des points et lignes qui sont sur un plan de projection P^2 [36].

Par Hartley et Zisserman dans "Multiple View Geometry in Computer Vision", nous pouvons retrouver le théorème algébrique suivant :

"Un mapping $h : P^2 \rightarrow P^2$ est une projectivité si et seulement s'il existe une matrice non singulière 3×3 H tel que pour chaque point dans P^2 représenté par un vecteur de coordonnées homogènes x , il est vrai que $h(x) = Hx$." [15, p. 33]

Pour interpréter ce théorème, tout point présent dans P^2 peut être représenté par un vecteur de coordonnées homogènes x . La matrice d'homographie H permet de représenter tout x sur un nouveau plan de projection P^2 .

Pour calculer la matrice H, nous avons besoin de quatre points présents sur le même plan (qui ne contient aucun des deux centres de caméra) et de leurs correspondances sur les deux plans d'images.

Pour cela, nous allons prendre l'hypothèse que l'image de référence est simplement un plan dans le monde, centré au centre de la sphère et dont la normale est la direction de la caméra de référence.

Cela permet de retirer la variable de profondeur du monde en 3D et d'utiliser ce plan comme plan 2D pour les correspondances.

Pour trouver les correspondances des quatre points sur les deux images, nous devons :

1. Choisir arbitrairement quatre pixels sur l'image de référence.
2. Trouver pour chaque pixel, son rayon de projection
3. Trouver l'intersection de chaque rayon de projection sur le plan hypothétique pour trouver des points dans l'espace 3D
4. Faire une projection de ces 4 intersections sur le plan d'image de la caméra virtuelle.

Pour le point 2, selon Zisserman et Hartley, le rayon de projection d'un pixel passe par deux points distincts, C, le centre de caméra et P^+x avec x , le pixel désigné, et P^+ , la matrice pseudo-inverse de P, la matrice de projection de la caméra de référence.

Nous calculons la matrice pseudo-inverse $P^+ = P^T (PP^T)^{-1}$ sachant que $PP^+ = I$, la matrice identité.

Nous pouvons savoir avec certitude que P^+x est projeté sur le pixel x car $P(P^+x) = Ix = x$.

Enfin, nous pouvons simplement normaliser le vecteur entre ces 2 points pour trouver le vecteur de direction unitaire.

Pour le point 3, nous devons trouver les points d'intersections entre des rayons et un plan [37, p. 17].

Sachant que pour tout point p sur un plan ayant comme normal n , $(p - p_0) \cdot n = 0$.

avec p_0 , le centre du plan

Et que la fonction pour un rayon est : $x(t) = l_0 + l * t$

avec $x(t)$ la position sur la ligne à distance t , l_0 l'origine de la ligne et l le vecteur de direction.

Nous devons calculer la distance t , tel que $(l_0 + l * t - p_0) \cdot n = 0$.

Pour résoudre t , nous avons : $(l * t) \cdot n + (l_0 - p_0) \cdot n = 0$

suivi de : $t = -\frac{(l_0 - p_0) \cdot n}{l \cdot n} = \frac{(p_0 - l_0) \cdot n}{l \cdot n}$

Maintenant que la distance t a pu être trouvée, nous pouvons calculer le point d'intersection entre la ligne et le plan : $p_{int} = l_0 + l * t$

Pour le point 4, nous allons utiliser la formule de calibration déjà démontrée plus haut. Nous utilisons la calibration de la caméra virtuelle calculée grâce aux positions géodésiques pour déterminer ce que serait la projection de ces quatre pixels sur la caméra.

Enfin, maintenant qu'on a une correspondance d'au moins quatre pixels entre les deux images, nous pouvons calculer la matrice H .

On peut utiliser l'algorithme du DLT.

Pour une transformation $x \leftrightarrow x'$, son équation est $x' = Hx$. Sachant qu'on utilise des coordonnées homogènes, les résultats seront équivalents mais à une échelle non nulle.

Nous pouvons alors donner l'égalité suivante : $x' \times Hx = 0$.

Notons tout d'abord cette équation :

$$Hx_i = \begin{pmatrix} h^{1\top} x_i \\ h^{2\top} x_i \\ h^{3\top} x_i \end{pmatrix}$$

avec h^i , la i -ème ligne de la matrice H .

Pour tout point x_i noté (u_i, v_i, w_i) , la formule équivaut à :

$$x' \times Hx = \begin{pmatrix} v'_i h^{3\top} x_i - w'_i h^{2\top} x_i \\ w'_i h^{1\top} x_i - u'_i h^{3\top} x_i \\ u'_i h^{2\top} x_i - v'_i h^{1\top} x_i \end{pmatrix}$$

Puisque la dernière équation de la matrice n'est pas entièrement indépendante car obtenue par une somme des 2 premières équations. Cela nous permet d'avoir l'équation finale de type $A_i h = 0$ qui peut être résolu grâce à un SVD :

$$\begin{bmatrix} 0^\top & -w'_i x_i^\top & v'_i x_i^\top \\ w'_i x_i^\top & 0^\top & -u'_i x_i^\top \end{bmatrix} \begin{pmatrix} h^1 \\ h^2 \\ h^3 \end{pmatrix} = 0$$

Où A_i est une matrice 2x8 qui correspond à la correspondance de 1 seul point 2D et son image.

La matrice A entière sera donc une matrice 8x9, ce qui est suffisant pour répondre aux huit degrés de liberté de la matrice H, puisqu'elle est équivalente même quand son échelle change.

Une dernière contrainte est souvent rajoutée pour normaliser la matrice : $\|h\| = 1$.

Pour que cet algorithme fonctionne, il est nécessaire qu'aucun set de 3 points ne soit colinéaire.

Une alternative pour calculer la matrice d'homographie est de changer l'égalité en un set non homogène d'équations linéaires en mettant comme contrainte que $h^9 = 1$.

Cet algorithme ne fonctionne pas si un des points désignés pointe vers l'infini, mais comme le calcul des pixels sur l'image prévient ce cas-ci, cette contrainte est négligeable.

Nous terminons avec une fonction de type $Ax = b$ en dérivant l'équation démontrée plus haut :

$$\begin{bmatrix} 0 & 0 & 0 & -u_i w'_i & -v_i w'_i & -w_i w'_i & u_i v'_i & v_i v'_i \\ u_i w'_i & u_i w'_i & u_i w'_i & 0 & 0 & 0 & -u_i u'_i & -v_i u'_i \end{bmatrix} h = \begin{pmatrix} -w_i v'_i \\ w_i u'_i \end{pmatrix}$$

Pour trouver la solution la plus proche, nous allons utiliser le DLT comme expliqué dans le chapitre 4.3.1.

C'est cette dernière fonction qui va être utilisée dans Sphaeroptica.

6 Conclusion

6.1 Contributions

Sphaeroptica a été développé pour répondre aux limitations auxquelles sont confrontés les scientifiques lorsqu'ils utilisent les solutions de numérisation actuelles, notamment pour l'étude des arthropodes sur une copie digitale.

Pour étudier un arthropode, les entomologistes prennent en compte sa morphologie. Certaines caractéristiques, comme la taille et les proportions, sont flagrantes mais d'autres, tout aussi importantes, peuvent être beaucoup plus subtiles, comme la présence ou l'absence de soie par exemple. C'est pourquoi il est primordial de leur proposer une version numérique avec une résolution aussi précise que possible. Sans cela, il est impossible d'effectuer une identification complète de l'arthropode.

Nous avons assemblé un dispositif scAnt, avec l'aide des infrastructures de l'IRSNB, pour avoir une prise d'images 360° automatique des arthropodes.

Après une calibration des images en utilisant les outils déjà à disposition, Sphaeroptica offre une visualisation du spécimen en 3D en utilisant les images 2D ainsi qu'une possibilité d'effectuer des mesures sur le spécimen grâce à une triangulation sur des repères posés.

Sphaeroptica, couplé à un scanner, permet une numérisation des spécimens qui est rapide, automatisée et moins coûteuse ainsi qu'une étude sur cette version numérique.

De plus, grâce à l'utilisation des images brutes, tous les détails de la morphologie restent visibles pour les scientifiques utilisant le logiciel. Ceci adresse la problématique de la reconnaissance de toutes les caractéristiques de l'arthropode pour une identification complète de ce dernier.

Nous avons également développé au sein de Sphaeroptica une caméra virtuelle ayant pour but d'offrir une visualisation plus fluide des spécimens en 3D.

6.2 Discussions

En quoi le logiciel Sphaeroptica apporte-t-il une réelle solution pour les entomologistes ?

Les solutions de numérisation actuelles ne sont pas optimales, en particulier pour l'étude des spécimens petits et/ou fins, comme les arthropodes, avec des structures complexes. En perdant de la résolution lors du calcul du modèle 3D, celles-ci ne permettent pas l'identification de l'entièreté des caractéristiques nécessaires pour l'étude d'un arthropode.



Figure 27: comparaison rapprochée entre une image prise par scAnt et le modèle de référence calculée par Artec

De plus, chaque solution actuelle comporte ses propres limitations.

Par exemple, les scanners à lumières structurées ont du mal à reconstruire les surfaces brillantes, sombres ou poilues, surfaces qui sont très communes parmi les arthropodes. Il est possible de résoudre cette problématique en ajoutant un spray mat. Mais cette opération risque d'endommager les spécimens et surtout de provoquer une perte d'information quant à la texture. C'est un risque inacceptable pour les arthropodes holotypes, les spécimens qui ont servi à la description de leurs espèces. Les éléments très fins des petits arthropodes peuvent aussi compliquer l'utilisation d'un scanner déjà très coûteux.

Par comparaison, la photogrammétrie classique permet, pour chaque cas et besoin du scan, de simplement changer l'objectif de la caméra. Cela est donc plus facilement adaptable si nécessaire. C'est la raison pour laquelle Sphaeroptica est basée sur les principes de cette dernière.

Mais, la création d'un modèle 3D par photogrammétrie classique est coûteuse en temps et en ressources. Un seul modèle prend plusieurs heures à être construit sur une station de travail performante (Intel Xeon, 2xNvidia 3080GTX, 64Gb RAM SSD).

Le logiciel Sphaeroptica ne crée pas de modèle 3D ; il utilise simplement les images 2D calibrées.

Malheureusement, ce dernier n'est pas lui-même exempt de limitations.

La plus importante est l'espace en stockage que peut prendre un seul modèle. Un set de 180 images brutes à 20 MPX occupe 4 Go de mémoire. Cela est 80 fois supérieur à l'espace requis pour un modèle 3D, qui est dans les alentours de 50MB. Même si le coût du stockage est relativement bon marché, la quantité nécessaire pour numériser la collection complète d'arthropodes de l'IRSNB (38 millions de spécimens) est non négligeable. De plus, cela peut impacter l'utilisation de la bande passante.

Pour surmonter ce problème de stockage, il est possible de compresser les images. Une compression JPEG à 90% permet de réduire l'espace de stockage à 200 Mb. Il n'a pas encore été établi s'il fallait effectuer cette compression avant ou après la calibration et si elle permettait de garder la précision du modèle.

Un autre aspect de Sphaeroptica est que le logiciel ne prend en compte que les images 2D et leur position dans la scène 3D. Sphaeroptica ne connaît pas la forme de l'arthropode en question. Par conséquent, le calcul de distance effectué par le logiciel ne permet qu'un calcul de la droite entre deux points 3D. Il n'est pas possible de calculer une distance en tenant compte de la forme du spécimen vu que le logiciel ne la connaît pas.

Une dernière limitation est que Sphaeroptica dépend de la qualité des images et de leur calibration pour pouvoir sélectionner les points de repère de façon précise, sachant que cette sélection est manuelle. Sphaeroptica est donc sujet à des erreurs de calculs dus à une sélection potentiellement imprécise, en particulier lorsque la texture de l'arthropode est complexe.

6.3 Évolution future

Sphaeroptica est un logiciel toujours en développement. Il est donc possible d'ajouter ou d'améliorer des fonctionnalités.

En premier lieu, il serait intéressant d'intégrer un pseudo-focus stacking. Pour un même angle de vue, il serait possible d'avoir plusieurs images et de passer de l'une à l'autre. Cela permettrait de combiner les avantages du focus stacking, la netteté du spécimen, et la contrainte de l'utilisation d'images brutes, la qualité de la résolution des images étant la première préoccupation des entomologistes.

Une deuxième amélioration serait d'adresser les limitations de Sphaeroptica, en particulier celle liée au calcul des distances entre deux points 3D. Pour adresser celle-ci, il serait intéressant d'utiliser le "sparse point cloud" créé lors de l'alignement des caméras, pour pouvoir mettre des checkpoints et mettre du volume entre 2 distances. Cette approche permettrait de calculer des semi-repères forts utilisés en morphométrie géométrique.

Concernant la caméra virtuelle, il est possible d'améliorer le confort de l'utilisateur. Pour cela, nous proposons d'utiliser une IA Neural Radiance Fields (NeRF). Cette IA permet, après entraînement, de calculer assez efficacement la vue en lui donnant la position et orientation de la caméra. En effet, n'utiliser qu'une seule image de référence ne permet pas une bonne synthèse des nouvelles vues.

La dernière amélioration ne consiste pas à modifier Sphaeroptica, mais concerne la calibration du scanner utilisé. Nous n'avons pas eu la possibilité d'effectuer un scan entier avec une mire 3D, la qualité de la résolution des images étant la première préoccupation des entomologistes.

Une dernière piste de réflexion concerne la possibilité d'utiliser des images EDOF. Pour ce faire, il serait intéressant de pouvoir quantifier les erreurs géométriques apportées par l'utilisation de telles images. Si celles-ci sont négligeables, la netteté des images du spécimen pourrait être un atout majeur pour les entomologistes.

Nous avons pu poser les premières bases de Sphaeroptica et créer un prototype convaincant. Les résultats obtenus pendant ce mémoire sont encourageants et ce logiciel possède encore

un potentiel d'évolution pour s'accorder au mieux aux besoins des entomologistes pour la photogrammétrie des arthropodes.

Bibliographie

- [1] F. Plum and D. Labonte, 'scAnt —an open-source platform for the creation of 3D models of arthropods (and other small objects)', *PeerJ*, vol. 9, p. e11155, Apr. 2021, doi: 10.7717/peerj.11155.
- [2] J. Brecko and A. Mathys, 'Handbook of best practice and standards for 2D+ and 3D imaging of natural history collections', *Eur. J. Taxon.*, no. 623, Apr. 2020, doi: 10.5852/ejt.2020.623.
- [3] 'Welcome to the DIGIT-KEY application (Delivery 7.2 of the Synthesys+ project) — Digit'. <https://digit.naturalheritage.be/> (accessed Jun. 04, 2023).
- [4] Pluta, Martin, Glöckler, Falko, Kroupa, Alexander, and Schurian, Bernhard, 'ZooSphere - Development of a software for automated spheric image capturing and interactive 3D visualization of biological collection objects', presented at the Missouri Botanical Garden Open Conference Systems, TDWG 2014 ANNUAL CONFERENCE, Elmia Congress Centre, Jönköping, Oct. 2014.
- [5] J. Wilm, O. V. Olesen, and R. Larsen, 'SLStudio: Open-source framework for real-time structured light', in *2014 4th International Conference on Image Processing Theory, Tools and Applications (IPTA)*, Paris, France: IEEE, Oct. 2014, pp. 1–4. doi: 10.1109/IPTA.2014.7002001.
- [6] McMillion, Matthew, 'Comment fonctionne le scan 3D à lumière structurée ? | Solutions de scan professionnelles | Artec3D', *Comment fonctionne le scan 3D à lumière structurée ?*, Nov. 25, 2022. <https://www.artec3d.com/fr/learning-center/structured-light-3d-scanning> (accessed Jun. 04, 2023).
- [7] J. Brecko, A. Mathys, W. Dekoninck, M. Leponce, D. VandenSpiegel, and P. Semal, 'Focus stacking: Comparing commercial top-end set-ups with a semi-automatic low budget approach. A possible solution for mass digitization of type specimens', *ZooKeys*, vol. 464, pp. 1–23, Dec. 2014, doi: 10.3897/zookeys.464.8615.
- [8] Littlefield, Rik, 'An Introduction to Focus Stacking --- Extended Depth of Field Digital Photography', *An Introduction to Focus Stacking — Extended Depth of Field Digital Photography*, Nov. 26, 2009. http://www.janrik.net/insects/ExtendedDOF/LepSocNewsFinal/EDOF_NewsLepSoc_2005summer.htm (accessed Jun. 04, 2023).
- [9] B. Ströbel, S. Schmelzle, N. Blüthgen, and M. Heethoff, 'An automated device for the digitization and 3D modelling of insects, combining extended-depth-of-field and all-side multi-view imaging', *ZooKeys*, vol. 759, pp. 1–27, May 2018, doi: 10.3897/zookeys.759.24584.
- [10] J. L. Schonberger and J.-M. Frahm, 'Structure-from-Motion Revisited', in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 4104–4113. doi: 10.1109/CVPR.2016.445.
- [11] Kitani, Kris, 'Camera Matrix', Carnegie Mellon University, 2016. [Online]. Available: https://www.cs.cmu.edu/~16385/s17/Slides/11.1_Camera_matrix.pdf
- [12] Gartia, Ansiha, 'Computer Vision Project', *Project 3: Camera Calibration and Fundamental Matrix Estimation with RANSAC*, 2016.

https://www.cc.gatech.edu/classes/AY2016/cs4476_fall/results/proj3/html/agartia3/index.html (accessed Jun. 05, 2023).

- [13] A. Anwar, 'What are Intrinsic and Extrinsic Camera Parameters in Computer Vision?', *Medium*, Mar. 01, 2022. <https://towardsdatascience.com/what-are-intrinsic-and-extrinsic-camera-parameters-in-computer-vision-7071b72fb8ec> (accessed Jun. 04, 2023).
- [14] J. Bloomenthal and J. Rokne, 'Homogeneous coordinates', *Vis. Comput.*, vol. 11, no. 1, pp. 15–26, Jan. 1994, doi: 10.1007/BF01900696.
- [15] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*, 2nd ed. Cambridge, UK ; New York: Cambridge University Press, 2003.
- [16] L. D. Kirkpatrick and G. E. Francis, *Physics: a world view*, 6th ed. Belmont, CA: Thomson Brooks/Cole, 2007.
- [17] OpenCV, 'OpenCV: Camera Calibration and 3D Reconstruction'. https://docs.opencv.org/4.x/d9/d0c/group__calib3d.html (accessed Jun. 04, 2023).
- [18] J. Weng, P. Cohen, and M. Herniou, 'Camera calibration with distortion models and accuracy evaluation', *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 10, pp. 965–980, Oct. 1992, doi: 10.1109/34.159901.
- [19] S. White, M. Satran, and A. Koren, 'Systèmes de coordonnées - UWP applications', Sep. 24, 2022. <https://learn.microsoft.com/fr-fr/windows/uwp/graphics-concepts/coordinate-systems> (accessed Jun. 05, 2023).
- [20] Rousseau, Jean-Jacques, 'Aberrations géométriques', *Aberrations Géométriques*. <http://ressources.univ-lemans.fr/AccesLibre/UM/Pedago/physique/02b/coursoptique/abergeo.html> (accessed Jun. 05, 2023).
- [21] Chiou Geoffrey Nelson, 'Reducing the variance of intrinsic camera calibration results in the ROS camera_calibration package', MS thesis, University of Texas, 2017. [Online]. Available: https://pab47.github.io/reports/Geoffrey_Chiou_MS_Thesis.pdf
- [22] Z. Zhang, 'A flexible new technique for camera calibration', *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000, doi: 10.1109/34.888718.
- [23] R. Toldo, R. Gherardi, M. Farenzena, and A. Fusiello, 'Hierarchical structure-and-motion recovery from uncalibrated images', *Comput. Vis. Image Underst.*, vol. 140, pp. 127–143, Nov. 2015, doi: 10.1016/j.cviu.2015.05.011.
- [24] Z. Cui and P. Tan, 'Global Structure-from-Motion by Similarity Averaging', in *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile: IEEE, Dec. 2015, pp. 864–872. doi: 10.1109/ICCV.2015.105.
- [25] E. Karami, S. Prasad, and M. Shehata, 'Image Matching Using SIFT, SURF, BRIEF and ORB: Performance Comparison for Distorted Images'. arXiv, Oct. 07, 2017. Accessed: Jun. 04, 2023. [Online]. Available: <http://arxiv.org/abs/1710.02726>
- [26] D. G. Lowe, 'Object recognition from local scale-invariant features', in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Kerkyra, Greece: IEEE, 1999, pp. 1150–1157 vol.2. doi: 10.1109/ICCV.1999.790410.

- [27] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, 'ORB: An efficient alternative to SIFT or SURF', in *2011 International Conference on Computer Vision*, Barcelona, Spain: IEEE, Nov. 2011, pp. 2564–2571. doi: 10.1109/ICCV.2011.6126544.
- [28] M. A. Fischler and R. C. Bolles, 'Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography', *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981, doi: 10.1145/358669.358692.
- [29] S. Li, C. Xu, and M. Xie, 'A Robust $O(n)$ Solution to the Perspective-n-Point Problem', *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1444–1450, Jul. 2012, doi: 10.1109/TPAMI.2012.41.
- [30] R. I. Hartley and P. Sturm, 'Triangulation', *Comput. Vis. Image Underst.*, vol. 68, no. 2, pp. 146–157, Nov. 1997, doi: 10.1006/cviu.1997.0547.
- [31] Solà, Joan, 'Towards Visual Localization, Mapping and Moving Objects Tracking by a Mobile Robot: a Geometric and Probabilistic Approach', PhD thesis, Institut National Polytechnique de Toulouse, 2017. Accessed: Jun. 04, 2023. [Online]. Available: <https://theses.hal.science/tel-00136307>
- [32] Tabb, Amy, 'N-view triangulation: DLT homogeneous method 1 (1).', *Amy Tabb*, Oct. 30, 2021. <https://amytabb.com/tips/tutorials/2021/10/30/triangulation-DLT-homogenous1/> (accessed Jun. 04, 2023).
- [33] T. Batpurev, 'Direct Linear Transforms (DLT)', *Temuge's webpage*, Feb. 06, 2021. https://temugeb.github.io/computer_vision/2021/02/06/direct-linear-transorms.html (accessed Jun. 04, 2023).
- [34] K. Gade, 'A Non-singular Horizontal Position Representation', *J. Navig.*, vol. 63, no. 3, pp. 395–417, Jul. 2010, doi: 10.1017/S0373463309990415.
- [35] E. Porcu, M. Bevilacqua, and M. G. Genton, 'Spatio-Temporal Covariance and Cross-Covariance Functions of the Great Circle Distance on a Sphere', *J. Am. Stat. Assoc.*, vol. 111, no. 514, pp. 888–898, Apr. 2016, doi: 10.1080/01621459.2015.1072541.
- [36] E. Dubrofsky, 'Homography Estimation', Master's essay, The University of British Columbia, Carleton University, 2007. [Online]. Available: https://www.cs.ubc.ca/sites/default/files/2022-12/Dubrofsky_Elan.pdf
- [37] Funkhouser, Thomas, 'Ray Casting', Princeton University, 2020. Accessed: Jun. 04, 2023. [Online]. Available: <https://www.cs.princeton.edu/courses/archive/fall00/cs426/lectures/raycast/sld001.htm>

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
École polytechnique de Louvain

Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve, Belgique | www.uclouvain.be/epl