

Master 1 - Ingénierie Informatique

Travaux d'Etude et de Recherche

-

GraphoScan

Mémoire Final

Thibault CHARPIGNON
Benoît GALLET
Emmanuel HERRMANN
Martin RÉTY

Encadré par : Matthieu EXBRAYAT

6 Février - 22 Mai 2017

Contents

1 Résumé du projet	3
1.1 Présentation	3
1.2 But du projet	3
2 Analyse de l'existant	3
2.1 Fonctionnement de l'application	3
2.2 Langages et frameworks utilisés	8
2.3 Fonctionnalités déjà implémentées	8
2.4 Fréquence d'acquisition limitée	8
2.5 Impossibilité de bouger la feuille	9
2.6 Problèmes divers	9
3 Besoins fonctionnels et non fonctionnels	9
3.1 Augmentation de la cadence d'acquisition	9
3.1.1 Programmation parallèle	10
3.1.2 Complexité	10
3.1.3 Modularité	10
3.1.4 Généralisation	10
3.2 Tracking de la plume	11
3.2.1 Analyse de la complexité	11
3.2.2 Gestion mouvements	11
3.3 Autres axes de travail	12
3.3.1 Zone de capture	12
3.3.2 Changement de la structure	12
3.3.3 Compatibilité Windows - Linux - macOS	12
4 Prototypes et résultats de tests préparatoires	12
5 Architecture	13
5.1 Acquisition	13
5.2 Tracking et reconstruction	13
6 Algorithmes et structures de données	15
6.1 Parallélisation OpenMP	15
6.2 Export des paramètres de la caméra	15
6.3 Tracking	16
6.4 Fonctionnement de la reconstruction 3D	17
6.4.1 Initialisation de la fenêtre et du contexte OpenGL	17
6.4.2 Dessin des points dans le repère 3D	17
7 Complexité de l'acquisition	18
8 Tests de fonctionnement et de validation	19
8.1 Latence	19
8.2 FPS	20

9 Bilan	20
10 Extensions et améliorations possibles	22
10.1 Droits d'auteur	22
10.2 Interface graphique	22
10.3 Bouger la feuille	23
10.4 Synchronisation native	23
10.5 Reconstruction 3D à n caméras	23
11 Planning	23
12 Manuel d'utilisation	27
12.1 Installation	27
12.1.1 FlyCap	27
12.1.2 OpenCV	27
12.1.3 OpenGL, glew, glut et glm	28
12.2 Utilisation	28
12.2.1 Acquisition	29
12.2.2 Reconstruction	29
References	30

1 Résumé du projet

1.1 Présentation

Ce projet de TER prolonge un travail déjà entamé l'année dernière par deux étudiants de Polytech, consistant à enregistrer en vidéo l'écriture d'un calligraphie, pour pouvoir reconstruire un modèle en 3D du mouvement de la plume. Une structure en bois supporte deux caméras, que l'on peut bouger le long de rails puis fixer à l'aide de vis. Le calligraphie écrit sous cette structure et la feuille est éclairée par des spots lumineux. Il faut alors associer l'image des deux caméras, ce qui n'est pas possible nativement avec le logiciel fourni par le fabricant (FlyCapture de PointGrey) pour faire de l'acquisition vidéo en stéréo, puis reconstituer via OpenGL les mouvements de la plume. Ces mouvements ont été sauvegardés grâce à des algorithmes de tracking, travaillant sur les vidéos enregistrées auparavant.

1.2 But du projet

Ce projet avait pour objectif de réaliser une reconstitution 3D des mouvements du calligraphie. Celui-ci peut recopier plusieurs textes, provenant de différents lieux et différentes époques, afin de pouvoir comparer les styles d'écriture, définir s'il existait différentes écoles d'écriture, différents styles, etc. De manière plus générale, le projet pourra servir pour beaucoup d'applications par la suite, car le code final se veut le plus généraliste possible.

La plupart des travaux actuels sur l'écriture utilisent des tablettes tactiles ou des capteurs placés à l'intérieur des stylos. Mais comme dans notre cas on des outils d'écriture spécifiques, une feuille et une plume ou un calame (morceau de roseau taillé en pointe), la captation du mouvement ne peut se faire a priori qu'indirectement, par le biais de capture de vidéos et de traitement d'images.

2 Analyse de l'existant

2.1 Fonctionnement de l'application

Cette section présente le fonctionnement global de l'application, afin de mieux cerner le sujet sur lequel nous avons travaillé. Il faut cependant préciser que ce n'est pas sous cette forme que nous avons reçu le projet, il ne fonctionnait que partiellement, la version présentée ici est la version finale post-TER.

Le dispositif montré sur la figure 1 nous permet de capturer les vidéos en direct. Ces deux vidéos sont enregistrées sur ordinateur, pour pouvoir faire de la stéréo : on encode les images en même temps pour avoir exactement les mêmes frames au même moment. On applique ensuite des paramètres d'undistortion pour lisser les bords de l'image. En effet, la caméra ayant un grand angle, l'extérieur de l'image est déformé, ce qui peut apporter des problèmes lors de la reconstruction 3D.

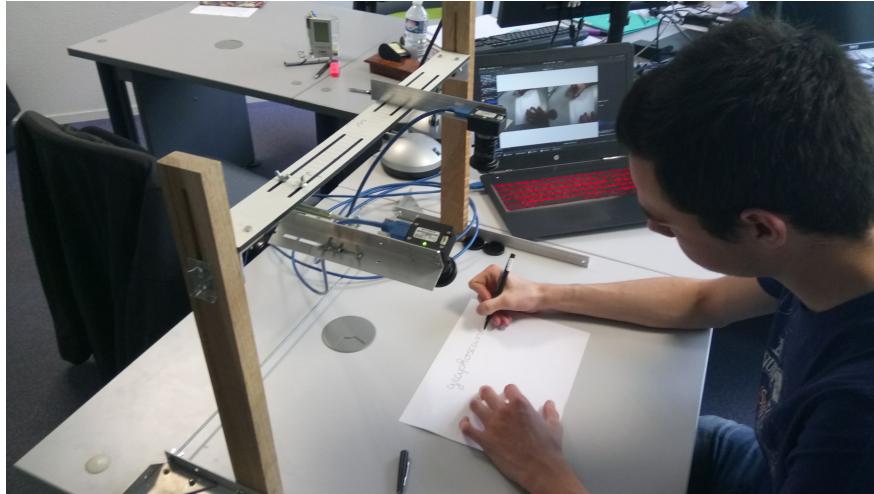


Figure 1: Ecriture et acquisition

Une fois les deux vidéos récupérées, on peut lancer la reconstruction. La première étape est de retrouver l'écriture sur l'image, pour cela différentes méthodes sont disponibles. On peut par exemple faire du tracking. Après avoir sélectionné une zone d'intérêt grâce à un ROI (Region Of Interest) selector (figure 2), on essaye de suivre le mouvement. Il suffit d'enregistrer les coordonnées du centre du point pour retrouver le mouvement dessiné (figure 3).

La seconde technique est celle du HOG (Histogramme de gradient orienté) qui permet de détecter des formes dans une image. Ce qui donne au départ une image comme sur la figure 4, et après nettoyage 5.

L'une ou l'autre de ces techniques peuvent être utilisées pour avoir une suite de coordonnées correspondant au dessin de l'écriture. Une fois que ces deux coordonnées sont récupérées, il existe des fonctions dans OpenCV pour faire des points 3D à partir de points 2D. Il faut pour cela récupérer les paramètres extrinsèques (matrices de rotation et vecteurs de translation, pour passer du repère lié à l'espace de travail au repère lié à la caméra) de la caméra.

Une fois que les coordonnées sont récupérées par l'une ou l'autre méthode, une fonction nous permet de rajouter deux points entre chaque coordonnée pour fluidifier la ligne. Ces coordonnées sont ensuite interprétées par OpenGL pour permettre de les afficher dans un repère en trois dimensions (figures 6 et 7).

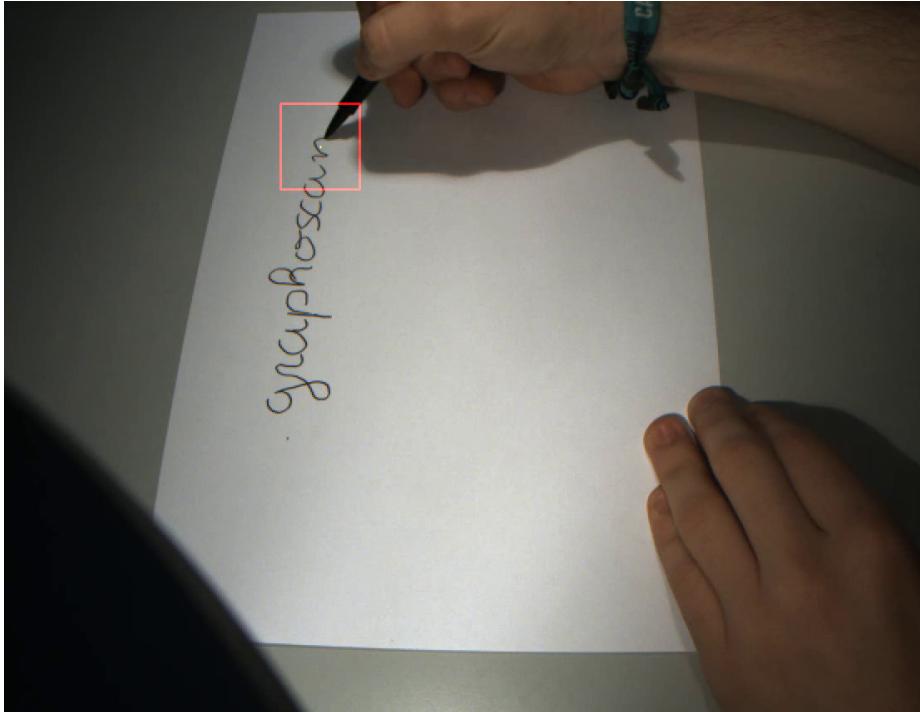


Figure 2: ROI selector

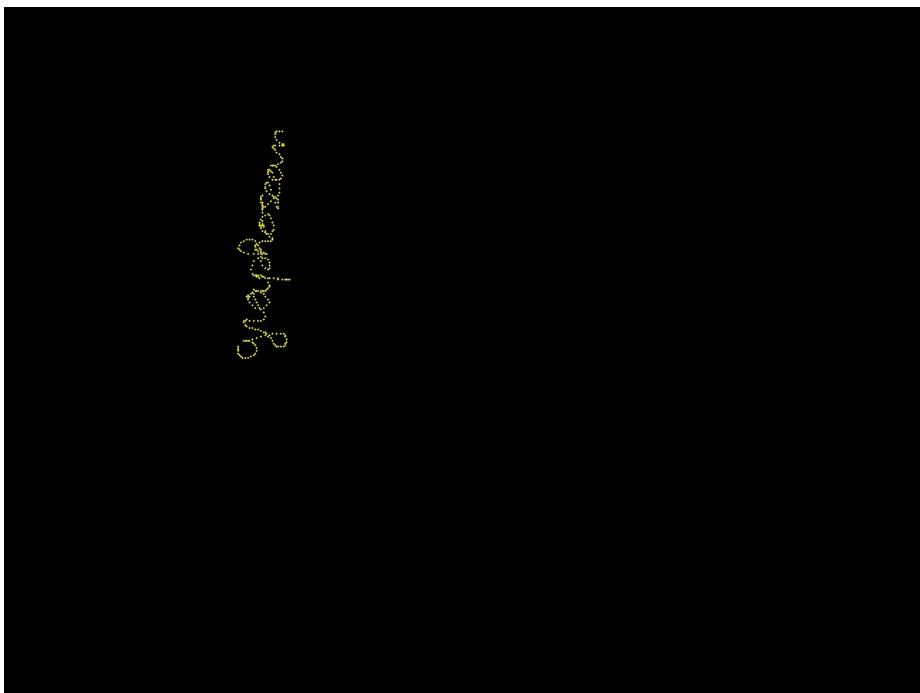
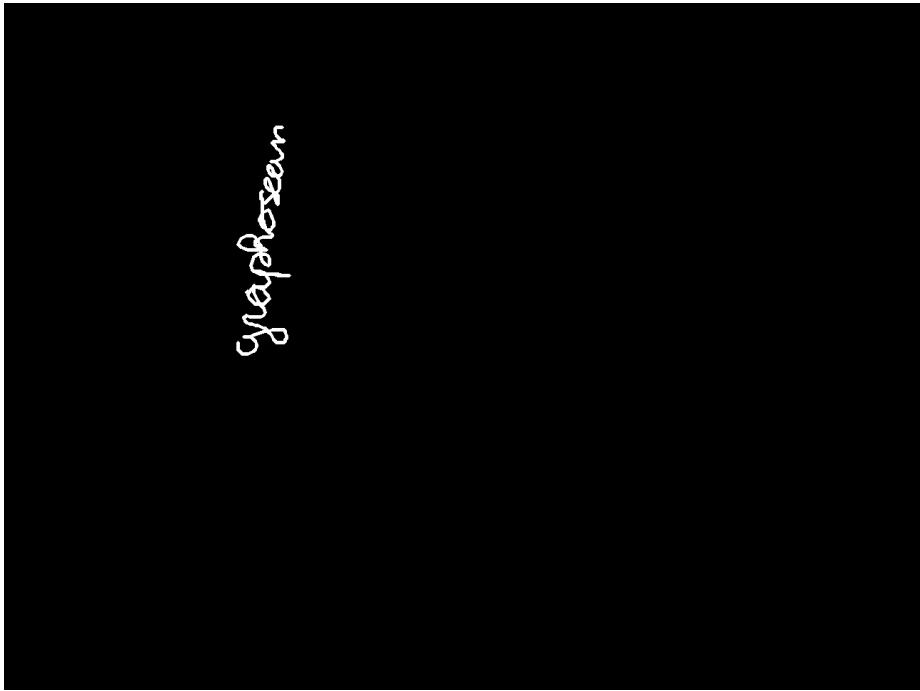


Figure 3: Reconnaissance avec le tracking



Figure 4: HOG brut



6

Figure 5: HOG nettoyé

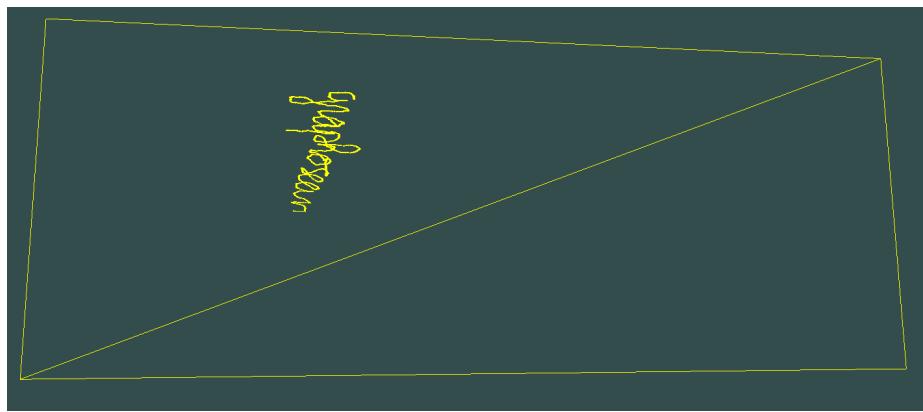


Figure 6: Image 3D - 1



Figure 7: Image 3D - 2

2.2 Langages et frameworks utilisés

Le programme en lui-même est entièrement réalisé en C++, et différentes bibliothèques graphiques sont utilisées dans ce projet pour le traitement du flux vidéo, dans le but de faire du tracking sur le résultat, comme OpenGL et OpenCV. OpenGL est utilisé pour la reconstruction du mouvement de la plume en 3D, tandis qu'OpenCV est plus utilisé pour le traitement de l'image, notamment toutes les opérations faites dessus. De plus, Matlab est utilisé pour effectuer diverses opérations mathématiques sur les images, comme par exemple pour la calibration originelle servant à l'alignement pour la stéréo, ainsi que pour l'enlèvement de la distorsion. En effet, comme les caméras ont un grand angle, les éléments sur le bord de l'image sont courbés, il faut donc les remettre droits pour pouvoir travailler dessus.

2.3 Fonctionnalités déjà implémentées

Le programme tel qu'il nous a été fourni dispose de plusieurs fonctionnalités élémentaires permettant son bon fonctionnement. Parmi celles-ci, nous pouvons compter :

- Le calibrage du dispositif dans sa globalité (caméras + surface d'écriture)
- La synchronisation des deux caméras pour une reconstitution en trois dimensions des gestes lors de l'écriture

En plus de ces fonctionnalités, existe un dispositif matériel de capture. Ce dernier (Figure 8) est composé d'une structure en bois sur laquelle sont montées deux caméras. Ces dernières sont connectées à l'ordinateur par le biais d'un câble USB 3.0. Il est également possible de les bouger afin d'en ajuster les réglages.

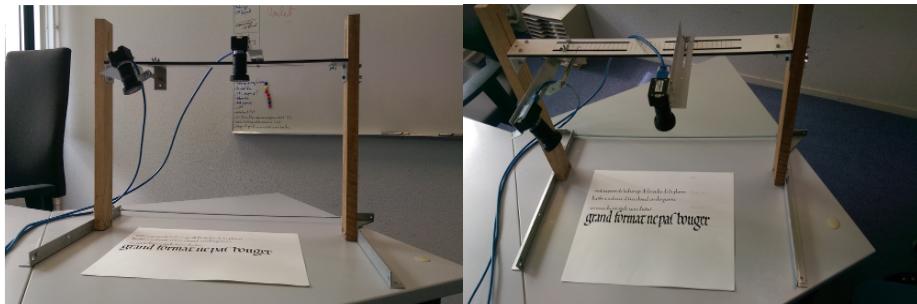


Figure 8: Dispositif de capture stéréo

2.4 Fréquence d'acquisition limitée

Jusqu'alors, le programme ne possédait pas une fréquence d'acquisition suffisante de l'image. En effet, cette dernière n'était que de l'ordre de huit images

par seconde. De ce fait, cette valeur ne permet pas une reconstitution précise en trois dimensions des gestes du calligraphe. L'objectif ici est donc de se rapprocher le plus possible de la fréquence maximale d'acquisition des caméras, soit trente images par seconde et ainsi gagner en précision lors du traitement.

2.5 Impossibilité de bouger la feuille

Avec la configuration actuelle, le calligraphe a l'impossibilité de bouger la feuille sur laquelle il écrit sous peine de perdre les réglages définis auparavant. Instinctivement, la personne qui écrit peut souhaiter bouger cette feuille et ainsi gagner en confort. Le but serait donc de trouver un moyen de gérer un changement de position de la feuille sans que cela n'affecte les résultats, en recalculant les réglages par exemple.

2.6 Problèmes divers

Initialement, l'éclairage du dispositif se faisait à l'aide de deux spots disposés de part et d'autre du calligraphe. Le problème majeur d'un tel moyen est la présence d'ombres à certains endroits rendant le traitement des images difficile, ainsi que la chaleur des lampes pouvant se révéler gênante à la longue. Un autre problème à gérer sont les gestes "inutiles" à l'acquisition que le calligraphe peut faire. En effet ce dernier peut par exemple vouloir étendre son bras pour se relaxer, geste qui sera pris en compte par le programme dans la reconstitution.

3 Besoins fonctionnels et non fonctionnels

Une séparation des tâches était déjà effective dans le projet original, sur lequel travaillaient deux personnes : une personne s'occupait de l'acquisition stéréo, pendant que la seconde était sur la modélisation 3D de la plume. Cette séparation a été conservée dans ce TER : GALLET Benoît et HERRMANN Emmanuel s'occupent de la première partie, tandis que CHARPIGNON Thibault et RÉTY Martin ont pris la deuxième.

3.1 Augmentation de la cadence d'acquisition

Naturellement, différentes idées nous sont venues pour augmenter la cadence d'acquisition de la vidéo en stéréo. Le programme fonctionne suivant plusieurs étapes : Tout d'abord, une image est capturée à partir des deux caméras, puis les deux images sont encodées dans une vidéo. Une fois la capture finie, la vidéo est traitée afin de régler les problèmes de distorsion. Grâce à cette grande boucle qui capture les images deux par deux (une image par caméra), les vidéos des caméras sont synchronisées (la xième image de chaque vidéo est datée du même instant, à une latence près), et permettent donc d'avoir simultanément la feuille d'écriture filmée sous deux angles différents. La modélisation en 3D sous OpenGL est alors possible. La seule variable est que plus la cadence d'acquisition est élevée, plus il

y aura de FPS (Frame Per Second) sur les vidéos finales, et plus la modélisation 3D de la plume sera précise. De plus, notre architecture et notre code devaient être assez robustes, pour que si un jour une troisième voire une quatrième caméra sont rajoutées, le nombre de FPS ne redescende pas drastiquement.

3.1.1 Programmation parallèle

Grâce à la programmation parallèle, qu'elle soit au niveau du CPU avec de l'OpenMP ou des threads, ou au niveau du GPU avec CUDA, nous pensions pouvoir accélérer l'acquisition des images, et donc des FPS sur les vidéos finales. Nous pensions regarder quelles parties pouvaient être faites en parallèle, peut-être est-il possible d'uniquement récupérer une image tous les trois centièmes de secondes (pour les 30 fps) dans le programme principal, et de faire tous les autres traitements dans des régions parallèles, avec par exemple un thread qui s'occupe d'ajouter la prochaine image à la vidéo, un autre thread qui enlève la distorsion de l'image, etc.

3.1.2 Complexité

Reprendre le code pour en examiner sa complexité était une autre piste envisagée pour augmenter les FPS. Nous pensions séparer cette idée en deux étapes : Tout d'abord regarder la complexité de l'algorithme dans sa globalité, pour se rendre compte s'il y a problème ou non à ce niveau là, et voir les morceaux posant plus problème que le reste, puis faire des tests plus précisément sur ces parties pour voir ce qui ne va pas. Une analyse légèrement différente pouvait être effectuée, avec des tests fonctionnels calculant quelle partie prend le plus de temps. Ces tests sont très complémentaires de ceux de complexité, à eux deux ils devaient mettre en exergue les problèmes principaux du code.

3.1.3 Modularité

Outre cette analyse de la complexité, une mise au propre du code devait être effectuée. En effet, tout se trouve dans la fonction `main`, dans deux grandes boucles. Une partie de notre travail était donc de modulariser cette fonction, de la séparer en plusieurs méthodes afin de gagner en clarté. Nous avons par la suite utilisé l'option de compilation `-O3` pour faire de l'inlining, ce qui permet de remplacer l'appel de la fonction par le corps de celle-ci, permettant de minimiser le coût de cette modularisation du code lors de son exécution.

3.1.4 Généralisation

Sinon, le dernier axe sur lequel nous avons travaillé a été la généralisation du nombre de caméras. A l'origine, tout dans le code était fait pour deux caméras, avec du code dupliqué pour chaque action. La généralisation pour n caméras sera facilitée par la modularisation du code, et permettra par la suite de rajouter une ou plusieurs caméras sans modification majeure, uniquement en changeant quelques `#define`.

3.2 Tracking de la plume

Comme pour la partie sur l'augmentation de la cadence d'acquisition, différents problèmes sont à résoudre pour le tracking. Cette partie permet de traiter les vidéos produites par les caméras et d'en ressortir une trace des mouvements effectués par le calligraphe. L'étudiant de Polytech qui a travaillé sur cette partie a recherché différents algorithmes permettant d'effectuer ce tracking. Son étude se focalise sur deux algorithmes basés sur l'apprentissage de toutes les apparences observées de l'objet, et d'une estimation des erreurs pour ensuite les éviter :

- Tracking Learning Detection (TLD)
- Kernelized Correlation Filters (KCF)

3.2.1 Analyse de la complexité

Heureusement, l'analyse des deux algorithmes a déjà été faite par l'étudiant, ce qui a montré que dans notre cas l'algorithme KCF est le plus efficace. Son étude est basée sur plusieurs critères, la déviation moyenne des deux vidéos, le nombre de *frames* et le temps de calcul. Seul le premier critère est réellement différent entre les deux méthodes. C'est cette différence qui a orienté son choix vers l'algorithme KCF.

Ici, notre premier axe de recherche sera orienté vers une étude complémentaire de ces algorithmes pour vérifier la véracité de l'analyse précédente. Pour cela nous allons réutiliser les critères d'étude et ensuite essayer d'en trouver d'autres pour confirmer le choix. Dans un second temps il nous faudra rechercher d'autres algorithmes ou méthodes de programmation pour améliorer le tracking.

3.2.2 Gestion mouvements

Bien entendu, le choix des algorithmes n'est pas la seule difficulté. Nous faisons face également à des contraintes physiques liées aux mouvements du calligraphe. Par exemple il doit prendre des temps de repos afin de garder sa fluidité d'écriture en faisant des gestes de relaxation du poignet. Ces mouvements ne doivent pas être pris en compte par l'algorithme de tracking afin d'éviter des erreurs sur la représentation du mouvement.

Résoudre ce problème pourrait passer par la sauvegarde à un temps T et à un temps T+1 d'une image de la partie suivie, puis analyser la différence entre les deux images et en ressortir un résultat positif ou négatif. Cela revient à prendre la dernière image où le calligraphe écrit et une autre image qui permettra de voir si le mouvement est la continuité de l'écriture ou un mouvement parasite.

3.3 Autres axes de travail

3.3.1 Zone de capture

En écrivant, le calligraphe doit de temps en temps bouger la feuille pour se repositionner et continuer sa rédaction. L'algorithme actuel ne gère pas ce mouvement, ce qui nécessitait après chaque mouvement de la feuille un nouveau calibrage des caméras et de la zone de capture. Une solution possible pour résoudre ce problème est la mise en place d'un système de cadre pour que le calligraphe sache la zone dans laquelle il peut écrire. Ce cadre pourrait être un marquage sur la feuille qui délimitera la zone de capture. Nous souhaitons également rechercher d'autre solutions possibles pour résoudre ce problème.

3.3.2 Changement de la structure

Pour le moment le dispositif de capture ne comporte que deux caméras et des angles de prises de vue bien définis. L'ajout d'une caméra et le repositionnement des deux premières peut permettre de rendre plus précise l'acquisition. De cette manière nous aurions à notre disposition des informations supplémentaires pour améliorer la reconstitution du mouvement. Il nous faut donc tester différentes configurations et choisir la meilleure.

Un des facteurs majeurs de la capture d'image est la lumière. En effet, il est important que la feuille soit bien éclairée pour le confort et l'écriture du calligraphe. La structure actuelle ne comporte pas d'éclairage du tout, il était nécessaire d'avoir une lampe d'appoint. Une solution simple est l'ajout d'un panneau LED pour avoir une luminosité uniforme sur toute la feuille.

Tous ces changements devront peut-être être accompagnés d'une refonte totale du dispositif.

3.3.3 Compatibilité Windows - Linux - macOS

A l'origine, les étudiants ont développé tout le code sur Windows et plus particulièrement sur l'IDE Visual Studio (C++). Pour rendre le code réutilisable à l'avenir nous avons comme objectif de pouvoir l'utiliser sur tous les systèmes d'exploitation (Linux/macOS en plus). Pour cela il est nécessaire d'uniformiser le code et de se servir de bibliothèques communes pour standardiser au mieux le projet.

4 Prototypes et résultats de tests préparatoires

Il nous fallait, pour bien prendre en main le projet, tester réellement le dispositif de lancement du logiciel jusqu'à la capture vidéo. Nous avons dû procéder à l'installation de tout l'environnement de travail nécessaire (FlyCap2, OpenCV, OpenGL) et l'acquisition des premières vidéos avec les caméras mises

à notre disposition.

Notre première tâche a été de transférer le code initial sous Linux et ainsi le tester directement. Les tests ont été concluants et le premier groupe a pu commencer directement à améliorer le système. Le second, quant à lui, a récupéré le code concernant le tracking mais a rencontré de gros problèmes lors de son passage de Windows vers Linux. Le code n'utilisant pas des fonctions standards, important des bibliothèques en "dur" et étant peu commenté, il a été un temps impossible de tester le code de l'étudiant qui travaillait sur le tracking l'année précédente. Il aura fallu passer plusieurs semaines de lecture, compréhension, débogage, etc pour réussir à faire enfin marcher le tracking et la reconstruction 3D. Le code n'était pas du tout commenté, il nous a fallu le déchiffrer variable par variable, fonction par fonction, classe par classe. De plus, les shaders OpenGL n'étaient pas présents, il nous a fallu les recréer en comprenant ce qu'il fallait passer à la carte graphique, et ce dont on avait besoin dans le code CPU.

5 Architecture

Le projet est scindé en deux grosses parties, presque indépendantes l'une de l'autre. Comme montré dans la figure 9 , toute l'acquisition des vidéos se fait en parallèle dans la partie acquisition. Cette partie gère aussi l'undistortion des vidéos. Une fois les vidéos créées et traitées correctement, on peut les envoyer à la seconde grosse partie qui est celle du tracking/reconstruction 3D. Cette partie permet de gérer différentes choses, comme des types de tracking ou de la reconstruction en OpenGL. Cette classe fonctionne comme une boîte noire, que l'on peut paramétriser à loisir pour réaliser ce dont on a besoin.

5.1 Acquisition

L'acquisition est elle-même séparée en deux parties. La première est l'acquisition à proprement parler, elle récupère le flux vidéo des deux caméras et permet de les synchroniser. Les vidéos sont encodées en avi, image par image, au rythme de 30 fps, ce qui est la cadence maximum des caméras.

Une fois les vidéos synchrones créées, une correction des distorsions est appliquée dessus. En effet, comme les caméras possèdent un grand angle, les bords sont déformés, il faut donc les remettre droits.

5.2 Tracking et reconstruction

Les vidéos créées sont ensuite passées au deuxième programme. Celui-ci peut exécuter différentes fonctions sur les vidéos. Une classe HOG permet de faire de la reconnaissance d'image, et une classe GraphoScan permet de lancer les différents trackings. Pour l'OpenGL, la classe Shader permet de compiler des shaders tandis qu'OpenGL permet de dessiner les points dans le repère. Enfin,

Camera sert à exécuter les différents mouvements de la caméra virtuelle de la vue 3D (zoom, direction, ...).

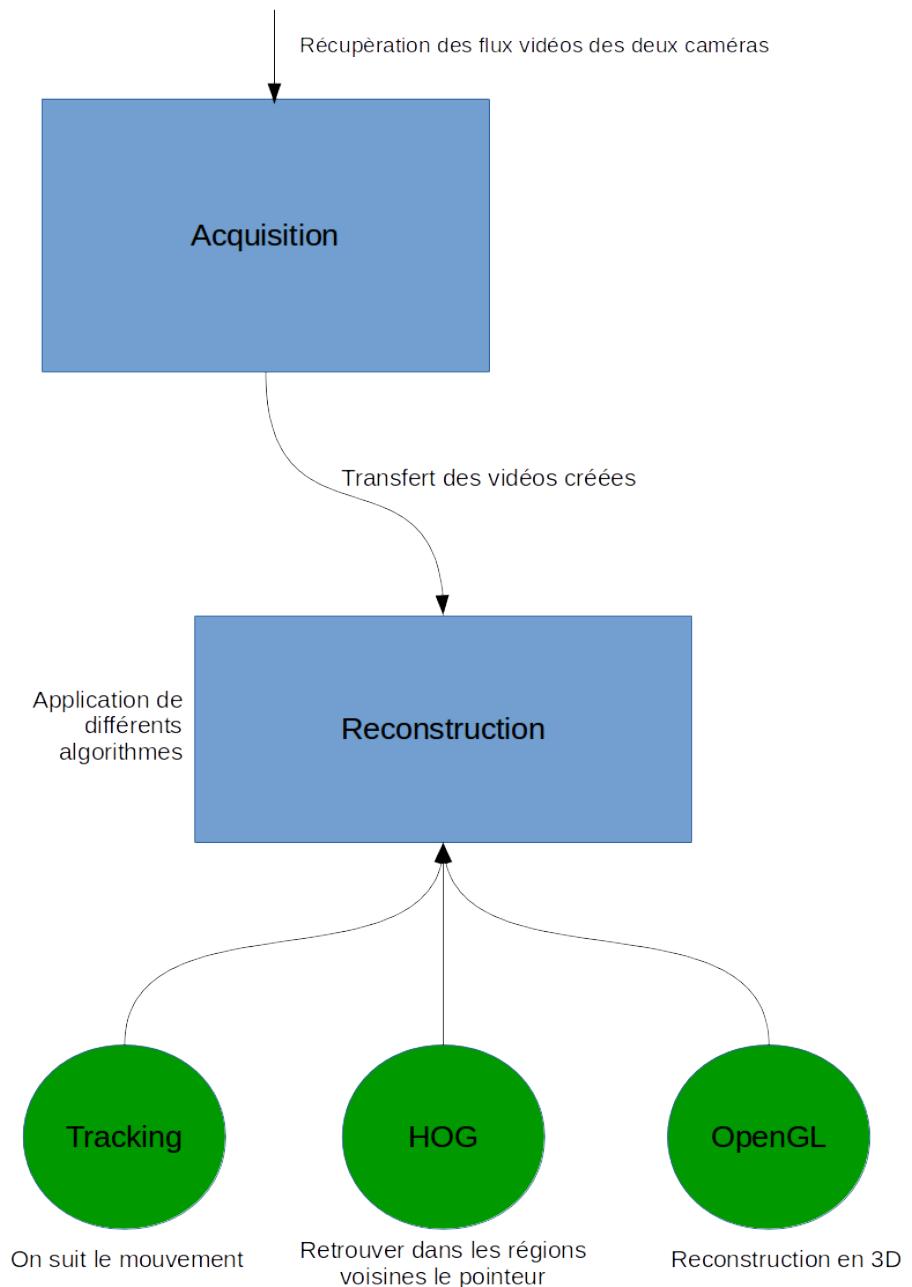


Figure 9: architecture

6 Algorithmes et structures de données

6.1 Parallélisation OpenMP

Afin de pouvoir faire de l'acquisition vidéo en simultané sur plusieurs caméras tout en gardant le plus possible d'images par seconde, nous avons opté pour une parallélisation de l'algorithme d'acquisition grâce à OpenMP. Ainsi nous affectons - dans la mesure du possible - une caméra à un thread en lançant la région parallèle de cette sorte :

```
#pragma omp parallel num_threads(numCameras)
```

Avec `numCameras` le nombre de caméras détectées.

Juste avant de commencer la capture, nous posons une barrière à l'aide de

```
#pragma omp barrier
```

dans le but de déclencher la capture des caméras de la manière la plus synchronisée possible. À l'intérieur de la boucle d'acquisition, une nouvelle barrière est mise avant chaque récupération du buffer de la caméra. Dans le cas où l'utilisateur souhaite faire un affichage de la capture qu'il est en train d'effectuer, un des threads est choisi via

```
#pragma omp single
```

afin de s'occuper de cet affichage. Enfin, chaque thread écrit la frame dans la vidéo correspondant à sa caméra.

6.2 Export des paramètres de la caméra

La récupération des paramètres de la caméra, calculés via MatLab, va nous permettre de faire l'undistortion et la reconstruction 3D. Jusqu'à présent, ces paramètres étaient rentrés en dur dans les programmes, nous avons donc décidé de faire des imports(exports de ces données pour plus de simplicité et de réutilisabilité. Pour ce faire, une fois les paramètres de calibration calculés dans MatLab , on récupère un objet CameraParameters par caméra. Afin de générer les fichiers de configuration nécessaires, il suffit de rentrer deux commandes par caméra :

```
dlmwrite(    \
'*PATH_TO_ACQUISITION*/Calib_camera_*NUM_CAMERA*_Matlab.txt', \
camera*NUM_CAMERA*.IntrinsicMatrix,'delimiter', ' ', \
'precision', 5)
dlmwrite(    \
'*PATH_TO_ACQUISITION*/Calib_camera_*NUM_CAMERA*_Matlab.txt', \
horzcat(camera*NUM_CAMERA*.RadialDistortion, \
camera*NUM_CAMERA*.TangentialDistortion), \
'-append', 'delimiter', ' ', 'precision',5)
```

Il faut remplacer `*NUM_CAMERA*` par le nom de l'objet de la caméra correspondante.

6.3 Tracking

De nombreux algorithmes de tracking existent, dans ce projet nous en utilisons plusieurs, tous fournis par OpenCV:

- KCF (High-Speed Tracking with Kernilized Correlation Filters) utilise les propriétés des matrices circulantes pour faire son tracking. (cf [7])
- TLD (Tracking Learning Detection) suit l'objet voulu image par image et grâce à une comparaison avec toutes les apparences observées réajuste le tracker de la cible.
- MIL (Multiple instance learning) crée et enrichie un classifieur pour séparer l'objet suivi et le reste de l'image.
- Boosting, comme MIL, utilise des classifieurs pour savoir où est l'objet et où sont les éléments extérieurs.

Il y a également d'autres algorithmes mais nous n'avons pas eu le temps de les tester. L'étudiant précédent a fait une comparaison entre l'algorithme KCF et TLD, à l'aide du comparatif des méthodes de tracking existantes, proposé par Joao F.Henriques et de quelques tests, il en est ressorti que KCF était plus efficace que TLD. Toutefois nous nous sommes également penché sur ce sujet. Pour savoir si un algorithme de tracking est plus efficace qu'un autre nous les avons comparés sur différents critères:

- La perte ou non du tracker sur la cible.
- La taille du tracker nécessaire pour avoir un tracking.
- Le suivi de certains mouvements.
- La rapidité d'exécution de l'algorithme.

Après avoir défini ces critères nous avons commencé les tests sur les mêmes vidéos à 30 fps. Ces tests ont donné des résultats complètement différents pour chaque algorithme. Il faut toutefois relativiser ces tests à la taille de la ROI. En effet, plus celle-ci est grande, et plus l'algorithme sera lent à l'exécution.

KCF est l'algorithme le plus stable, le tracker suit parfaitement la cible tant que celle-ci n'effectue pas un mouvement trop rapide. C'est également celui qui a le temps d'exécution le plus rapide. Après analyse de son code (OpenSource), nous avons ressorti une complexité d'environ $O(n^4)$, toutefois cette analyse fut rapide et nous n'avons pas eu le temps de l'approfondir. Boosting et MIL, qui utilisent la même technique d'apprentissage et d'enrichissement de classifieurs, donnent les mêmes résultats. Le choix de la région d'intérêt (ROI) a un impact significatif sur le tracking. L'utilisation d'un multi-tracker favorise l'analyse de l'image et ressort un tracking efficace, de plus la sélection d'une grande ROI n'a pas de réel impact. TLD donne des résultats complètement différents selon la situation. Lors d'un tracking sur des mouvements simples, celui-ci est le moins efficace de tous. En effet, il recalcule la ROI à chaque frame ce qui est très

coûteux et modifie le résultat final du tracking. Toutefois lors de mouvements rapides, qui ne sont pas gérés par les autres algorithmes, TLD arrive à retrouver l'objet grâce à sa base de connaissance.

A partir de tous ces tests nous en avons conclu que le choix de l'algorithme dépend de l'objet que l'on veut suivre. Si celui-ci est simple à observer, KCF est le plus approprié, mais lorsqu'il commence à être plus compliqué à suivre, il est préférable de se tourner vers les autres algorithmes qui utilisent leurs bases de connaissances pour retrouver l'objet.

6.4 Fonctionnement de la reconstruction 3D

Pour faire la reconstruction nous avons gardé les différentes librairies utilisées dans le code fourni. Nous avons utilisé GLFW 3 comme librairie de gestion de fenêtre et de contexte OpenGL, GLM comme librairie mathématique permettant de manipuler des matrices de toute taille et GLEW comme librairie de gestion des extensions utilisées et surtout pour gérer la plateforme sur laquelle on utilise OpenGL.

La reconstruction 3D de ce qui est écrit dans la vidéo capturée se fait en fin de programme. Le tracking produit un fichier de points 3D qui sont utilisés pour la reconstruction avec OpenGL. Ce programme s'effectue en deux phases.

6.4.1 Initialisation de la fenêtre et du contexte OpenGL

La première phase consiste à initialiser les outils OpenGL utilisés. Dans un premier temps il a fallu initialiser la fenêtre OpenGL, c'est à dire attribuer la version d'OpenGL utilisée, une taille, et même lier les fonctions d'événement OpenGL (mouvement de souris, touches clavier, etc) à la fenêtre. Ensuite il y a une initialisation du contexte, c'est à dire créer les outils qui permettent de lier les points 3D, la fenêtre et d'autres valeurs à la carte graphique (GPU). Pour cela on utilise des Vertex Array Object (VAO) qui permettent de lier à la carte graphique un programme nommé Shader, composé de FragmentShader qui sont utilisés pour dessiner dans la fenêtre. On utilise également des Vertex Buffer Object (VBO) qui nous permettent de lier des variables (ou structures) stockées dans le processeur (CPU) avec la GPU, qui permettront d'influer sur ce qui est dessiné dans la fenêtre par le Shader.

6.4.2 Dessin des points dans le repère 3D

La deuxième phase est la phase de traitement des données. Dans un premier temps il a fallu *parser* le fichier dans lequel les points 3D sont contenus. Ensuite nous avons chargé dans deux VAO différents deux Shaders, un qui sera utilisé pour construire un plan dans le repère 3D et un second qui sera utilisé pour effectuer le dessin des points 3D du tracking. Pour le dessin des points lors

du *parsing* nous plaçons chaque point dans des matrices à quatre lignes et une colonne qui contiendront les coordonnées du repère x,y,z et w. Une fois la liste remplie nous plaçons le point dans la fenêtre OpenGL à l'aide d'un savant calcul de placement de point qu'on nommera PVM. Ce PVM n'est rien d'autre qu'une suite de multiplications de matrices ordonnées : matrice de projection * matrice de vue * matrice de modèle qui permettent de rendre visible un dessin dans une fenêtre en fonction de la position de la camera et des coordonnées du point à dessiner. La matrice du modèle (M) est obtenue par la translation T , la rotation R appliquée sur l'objet tel que : $R*T*v = M*v$ où v est un vecteur de l'objet. La matrice de vue (V) est obtenue par la multiplication de M et un alignement des objets de la fenêtre par rapport à la vue humaine : $v' = V*M*v$ où v est un vecteur de l'objet. La matrice de projection (P) est obtenue par la multiplication de V, M, et une projection dans la fenêtre des objets : $v' = P*V*M*v$ où v est un vecteur de l'objet. Une fois ces valeurs de PVM envoyées à la GPU, le point est alors dessiné et est visible sur la fenêtre. Pour le dessin du plan nous avons repris les paramètres PVM calculés précédemment, puis effectué le dessin des points placés dans le repère. Nous avons alors en visuel le plan et ce qui a été reconstruit en 3D, une représentation de l'écriture effectuée dans la vidéo. Il est alors possible de bouger dans la fenêtre avec des touches et d'orienter sa vue en fonction de la position de la souris.

7 Complexité de l'acquisition

Nous prenons comme hypothèse le fait qu'il y ait n lignes et m colonnes dans une image. Nous proposons deux complexités, une borne min si on considère qu'un pixel de l'image se traite en temps $O(1)$, et une borne max si l'on considère que le pixel se traite en temps $O(3)$ à cause de ses caractéristiques RGB.

Borne min

Récupération de l'image brute (RetrieveBuffer) : $O(n * m)$

Conversion de l'image brute vers RGB (rawImage.Convert) : $O(n * m)$

Création de la matrice de l'image RGB (imageRGB=cv::Mat) : $O(n * m)$

Ecriture d'une image dans la vidéo (outputVideo.write) : $O(n * m)$

Ce qui nous donne un total de temps d'exécution en $O(4(n * m))$.

Borne max

Il suffit de multiplier par trois les valeurs précédentes, ce qui nous donne au total $O(12(n * m))$

Analyse

Nous pouvons donc en conclure que le temps d'exécution de l'algorithme C est : $O(4(n * m)) \leq C \leq O(12(n * m))$

C'est-à-dire qu'il s'exécute globalement en temps linéaire en la taille de l'entrée. Toutefois, cela reste une supposition car nous ne connaissons pas la complexité de toutes les méthodes OpenCV utilisées lors de l'acquisition. Une compression est également effectuée lors de l'enregistrement de la vidéo, car nous encodons en XVID, ce qui apporte bien sûr un delta sur la complexité totale de l'algorithme.

De plus, il faut l'exécuter à chaque tour de boucle (à chaque acquisition d'image), donc on pourrait dire que sa complexité serait alors de $O(n * m * f)$, f étant le nombre de frames que nous enregistrons. Evidemment, cela vaut pour la version parallèle du code, si on prend en compte la version séquentielle, il faut bien sûr multiplier cette complexité par le nombre de caméras que l'on a.

8 Tests de fonctionnement et de validation

Différents tests de validation ont été effectués tout au long du projet afin de s'assurer du bon fonctionnement de l'application. Ces tests pourront être refaits par la suite en décommentant les `#define` nécessaires, et pourront donc être réutilisés par des étudiants reprenant ce projet.

8.1 Latence

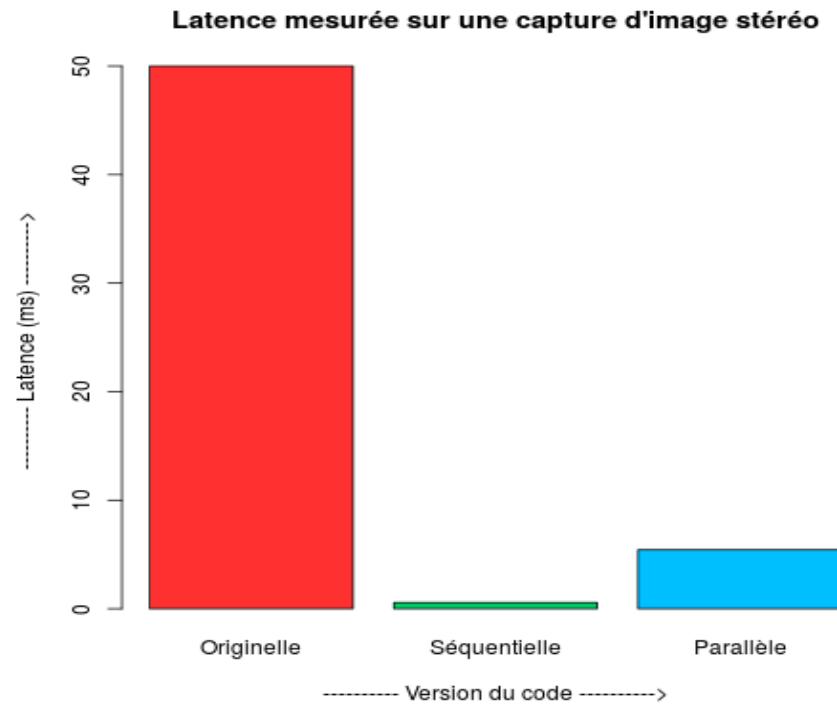


Figure 10: Latence

Version du code	Min	Max	Moyenne	Médiane	Écart-type
Originelle	NA	50	NA	NA	NA
Séquentielle	0.554	0.58	0.568	0.5688	0.0079
Parallèle	4.754	5.962	5.377	5.433	0.3932

Figure 11: Comparaison de la latence

La latence est une caractéristique incontournable de notre projet, le client nous ayant précisé plusieurs fois que c’était très important qu’il y ait le moins de décalage possible entre les deux vidéos. Plus le décalage est grand et moins la reconstruction en 3D par la suite sera précise. Nous avons donc mesuré précisément le temps qui sépare deux frames dans le code.

Ce diagramme en bâton (Figure 10) montre cette latence. Elle est plus importante avec le parallélisme qu’avec le programme séquentiel. Il faudra donc que le client pèse le pour et le contre sur la méthode qu’il préfère utiliser. Avec le parallelisme, le nombre de caméra pourra être augmenté (de façon raisonnable) sans chute de fps, car chaque cœur de la machine s’occupera d’une caméra particulière, par contre, la latence est plus importante : 5 à 6 ms. Avec la version séquentielle du code, la latence est plus que minime : 1/2 millième de seconde environ. Par contre, séquentiellement, comme on lance les acquisitions les unes après les autres, plus il y a de caméras, plus celles qui sont lancées en dernier auront de décalage avec la première.

8.2 FPS

L’augmentation du nombre de FPS était un des objectifs principaux de notre TER. En effet, le groupe précédent n’arrivaient qu’à enregistrer des images à la vitesse de 7/8 FPS, ce qui ne permettait pas une reconstruction optimale. Comme on peut le voir sur la Figure 12, nous faisons de l’acquisition à 30 FPS constants, en parallèle comme en séquentiel. Cependant, nous pensons qu’il faut privilégier le code fait en parallèle, car il est plus robuste à l’ajout de caméras futures.

9 Bilan

Nous avions différents objectifs au départ, concernant les deux parties de notre TER, acquisition et tracking/reconstruction.

Le premier était d’avoir une vidéo à 30 fps. Précédemment, le programme capturait les images à la vitesse de 7 ou 8 images/secondes, et actuellement on crée des vidéos en 30 fps constants, et en plus avec de la parallélisation, ce qui nous laisse à penser que le nombre de caméras peut augmenter sans baisser ce chiffre.

Il fallait ensuite généraliser le code de l’acquisition pour qu’il puisse marcher pour n caméras. Au départ, tout était fait pour deux caméras, avec du code

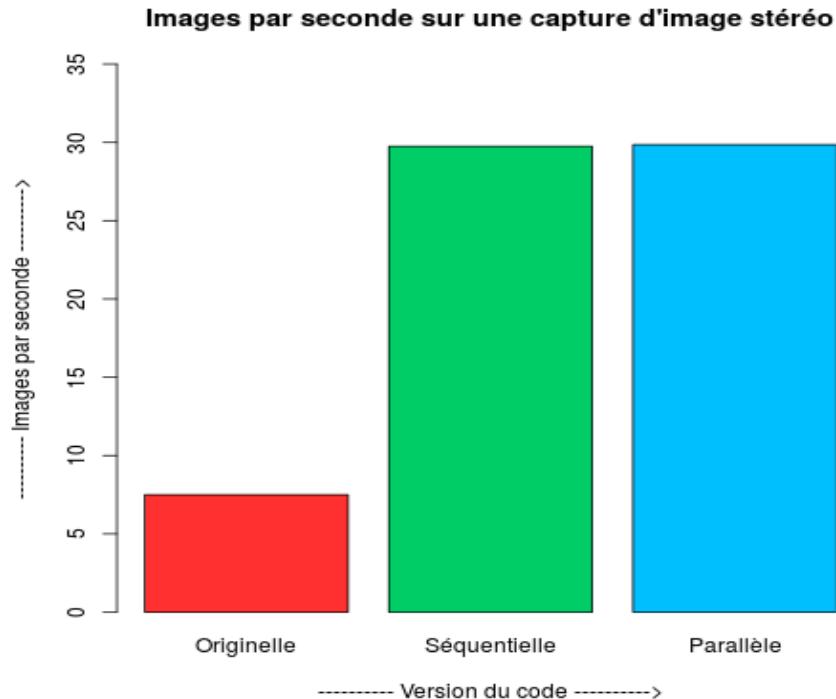


Figure 12: FPS

Version du code	Min	Max	Moyenne	Médiane	Écart-type
Originelle	7	8	NA	NA	
Séquentielle	29.11	29.85	29.62	29.75	0.2564
Parallèle	29.41	30.3	29.86	29.86	0.4697

Figure 13: Comparaison des FPS

copié-collé en deux fois pour chacune des caméras. Tout cela a été modifié, avec des tableaux de taille n (pour n caméras). Par contre, pour la partie reconstruction 3D, cela marche toujours avec deux caméras.

Le code devait être compilable et exécutable sous environnement UNIX, les étudiants précédents l'ayant fait marcher uniquement sous Windows. Ce point a été également réalisé, une partie de ce rapport (section Manuel) a même été faite pour permettre aux prochains étudiants qui travailleront sur ce projet d'installer les frameworks et bibliothèques nécessaires à la compilation du code, ainsi que des explications sur l'utilisation de l'application.

Pouvoir bouger la feuille n'a malheureusement pas pu être fait par manque

de temps, cependant nous avons réfléchi à une solution pour le faire, qui est détaillée dans la partie Extensions et améliorations possibles.

Enfin, il fallait rendre le code plus lisible. Ceci a également été fait, grâce à un gros travail de documentation et de compréhension du code. En effet, la majeure partie n'était pas commentée, avec des noms de variables non explicites, etc. Cela a été amélioré, avec une documentation Doxygen, afin que le code puisse être compris par tous.

10 Extensions et améliorations possibles

10.1 Droits d'auteur

Au fur et à mesure que nous essayions de comprendre le code de la reconstruction 3D (ce qui a été long car il a été fourni sans doc, commentaires et explications d'utilisation, avec une syntaxe et des imports pour Windows), nous nous sommes rendu compte que sur cette partie plusieurs morceaux de code avaient été copiés sur Internet.

La classe Shader, ainsi que la moitié de la classe Camera provient de http://blog.csdn.net/sinat_26989191/article/details/51205149, un outil pour faire un système solaire en 3D.

La classe HOG est presque entièrement copié-collée de la fin de ce programme (<http://lib.csdn.net/snippet/cplusplus/28974>), qui est sous copyright.

Diverses autres parties du code proviennent de copiés-collés de différentes documentations ou exemples, ce qui est dans une certaine mesure moins dérangeant.

Il y a certainement d'autres endroits du projet qui proviennent d'Internet, tout n'a pas été vérifié par manque de temps. Au total, nous estimons qu'au moins 50% du code n'a pas été écrit par le groupe précédent. Cela implique que si ce projet est repris par un groupe futur, il faudra réécrire ces parties du code, ou du moins indiquer clairement les sources et les différents copyrights sur les classes copiées.

10.2 Interface graphique

Une extension possible serait de réaliser une interface graphique, par exemple en Qt. Cela permettrait un maniement beaucoup plus facile de l'application, et rendrait possible son utilisation par des personnes extérieures. En effet, en ce moment tout s'exécute dans le terminal, ce qui peut être moins intuitif. Avec cette interface graphique, on pourrait par exemple choisir de sélectionner une caméra pour voir ce qu'elle enregistre, définir différentes options comme choisir l'algorithme de tracking à utiliser, les sorties vidéos ou images à effectuer, ...

10.3 Bouger la feuille

La dernière amélioration à apporter au code serait de pouvoir bouger la feuille. Une technique possible serait de faire du multi-tracking sur deux points. En effet, en dessinant juste deux points noirs sur la feuille et en faisant du tracking dessus, on serait capable de détecter tout mouvement ou rotation de la feuille. Il suffira donc par la suite d'appliquer une translation aux coordonnées de l'écriture par rapport à celle de ces deux points pour pouvoir avoir une image qui colle plus à la réalité.

10.4 Synchronisation native

En naviguant sur la documentation de PointGrey (le fabricant des caméras), nous nous sommes rendu compte qu'il existait déjà un moyen de faire de l'acquisition stéréo ([5]), grâce à des fonctions dans l'API fournie, et à un cable reliant les caméras. Peut-être faudrait-il envisager aussi cette solution dans le futur, même si notre travail d'amélioration du code de l'étudiant précédent s'étant occupé de cette partie est désormais fini. Il y aurait donc peut-être moyen dans le futur de supprimer la partie de l'acquisition vidéos, et de juste utiliser la deuxième phase du code, avec un peu de modifications pour intégrer cette acquisition native avant. Cela permettrait d'avoir une application en un bloc, plutôt que séparée comme actuellement.

10.5 Reconstruction 3D à n caméras

Comme dit dans le bilan, la fonction qui met en relation des points en 2D pour faire des points 3D ne prend que deux caméras. Pour modifier cela, il faudrait changer la méthode de reconstruction 3D de points qui actuellement ne prend les points que de deux caméras. Une autre solution serait de lancer le programme sur toutes les paires de caméras et de faire ensuite la moyenne des points récoltés pour plus de réalisme dans la reconstruction.

11 Planning

Les petites barres sur les diagrammes de Gantt (Figure 14 et suivantes) correspondent aux différentes réunions que l'on a eu, suivies des noms des participants. Les plus grandes correspondent quant à elles aux tâches effectuées. Il y en a peu pour le moment car la phase de compréhension et d'installation des composants a été relativement longue.

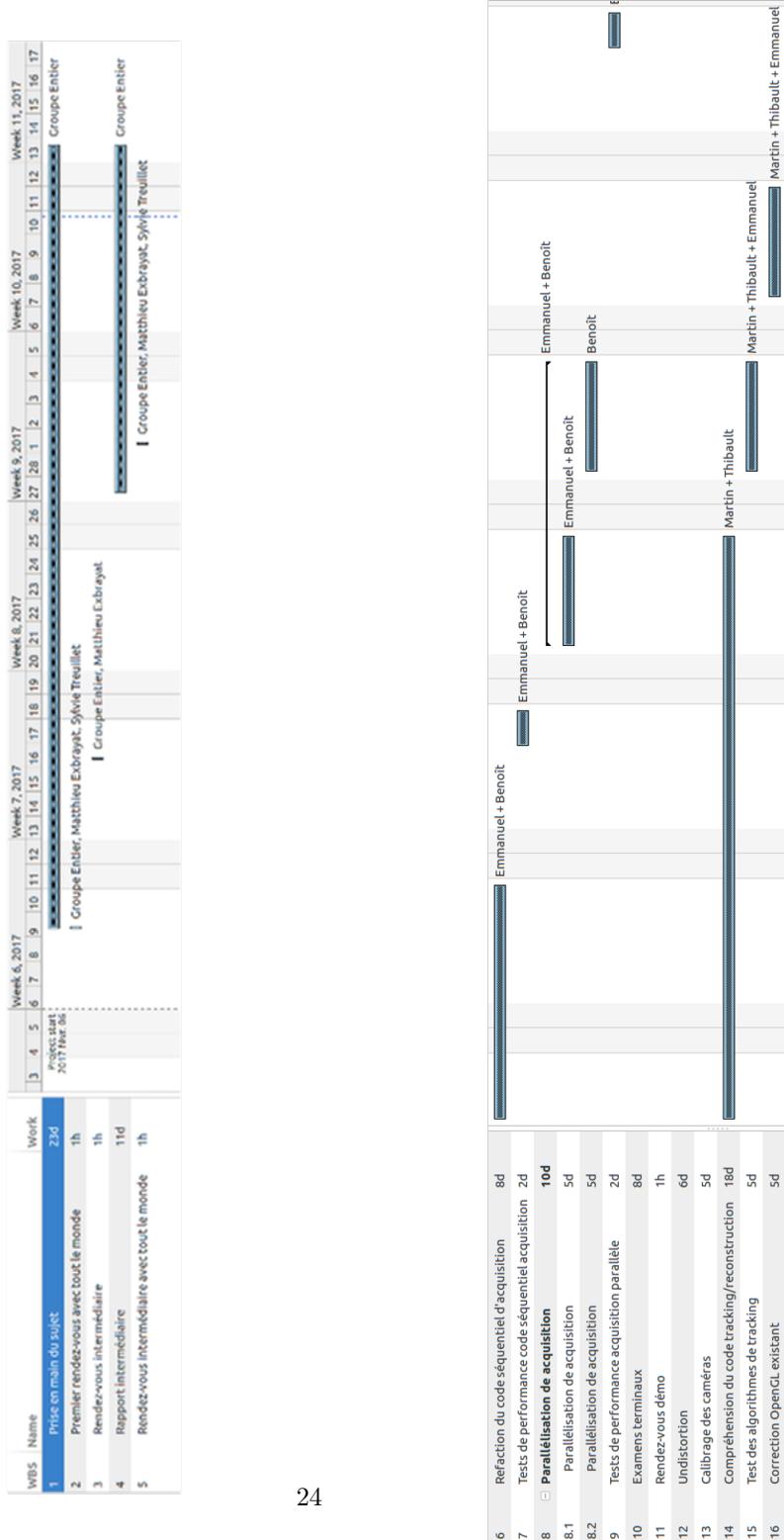


Figure 14: Diagramme de Gantt 1/2



Figure 15: Diagramme de Gantt 2/2

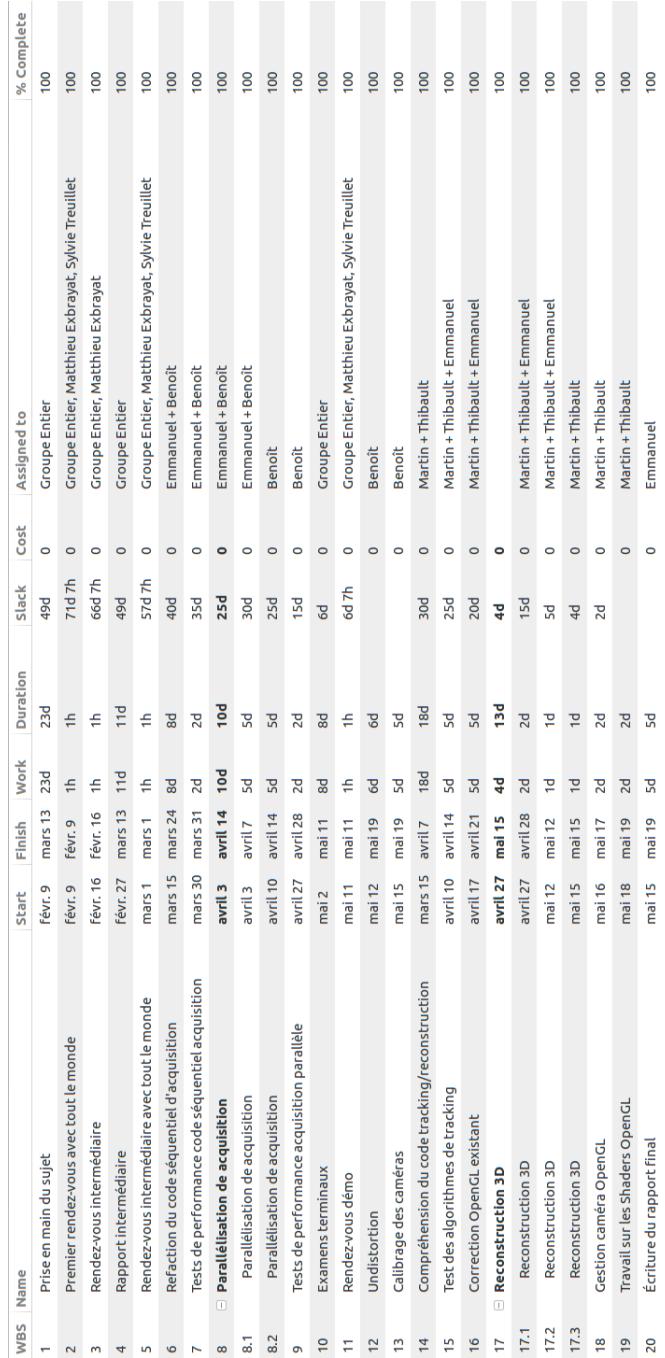


Figure 16: Diagramme de Gantt - Total

12 Manuel d'utilisation

Pour pouvoir lancer et utiliser cette application, différentes bibliothèques doivent être installées sur l'ordinateur. Toutes les manipulations sont décrites pour un environnement Linux (ou Mac), mais peuvent être transposées assez facilement pour Windows.

12.1 Installation

12.1.1 FlyCap

FlyCap est le framework de PointGrey, le fabricant des deux caméras, permettant de communiquer dans un programme avec les caméras. L'installation est relativement aisée, il suffit d'aller sur le site du constructeur (<https://www.ptgrey.com/support/downloads>) pour télécharger le framework. Il faut alors renseigner le type de caméra (Chameleon3), le modèle (CM3-U3-13S2C-CS) puis le système d'exploitation utilisé. Il faut ensuite sélectionner "Latest FlyCapture2 Full SDK" dans la partie software. Une fois le téléchargement terminé, un README détaille toutes les étapes nécessaires à l'installation, il suffit de les suivre une à une pour avoir FlyCap en état de marche. Pour vérifier si le logiciel s'est bien installé, il est possible de le lancer dans un terminal via la commande *flycap* pour voir la vue d'une des caméras.

12.1.2 OpenCV

OpenCV doit être installé également, afin de profiter de son traitement d'image et de ses algorithmes de tracking. Il a été décidé d'utiliser OpenCV 3, car celui-ci contient un module optionnel OpenCV_contrib. Ce dernier contient tous les algorithmes de tracking, c'est un extra-module indispensable pour pouvoir exécuter le projet.

Pour installer OpenCV, voici la marche à suivre :

- On se place dans le dossier de travail et on télécharge les dernières versions OpenCV et d'OpenCV_contrib.

```
cd ~/<my_working_directory>
git clone https://github.com/opencv/opencv.git
git clone https://github.com/opencv/opencv_contrib.git
```

- On se place dans le dossier OpenCV téléchargé et on crée un répertoire temporaire pour le build.

```
cd ~/opencv
mkdir build
cd build
```

- On lance la configuration en n'oubliant pas d'inclure l'extra-module. Cette étape peut prendre un certain temps, jusqu'à 1h30 suivant les machines.

```
cmake -D CMAKE_BUILD_TYPE=Release  
-DOPENCV_EXTRA_MODULES_PATH=*PATH_TO_FOLDER*/opencv_contrib/modules  
-D CMAKE_INSTALL_PREFIX=/usr/local ..
```

- On démarre l'installation. Si on ne possède que quatre cœur, mieux vaut marquer -j4.

```
make -j7
```

- Dans certains cas, il faut rajouter la ligne suivante dans le .bashrc pour indiquer où se trouvent les librairies OpenCV.

```
export LD_LIBRARY_PATH=\$LD_LIBRARY_PATH:/usr/local/lib
```

12.1.3 OpenGL, glew, glut et glm

- Il faut commencer par installer OpenGL et Mesa :

```
sudo apt-get install cmake xorg-dev libglu1-mesa-dev
```

- Pour savoir si l'opération s'est bien passée, il faut regarder que les deux fichiers suivants sont bien présents :

```
/usr/include/GL  
/usr/lib/x86_64-linux-gnu/libGL.so
```

- A présent, il faut retourner dans le dossier de travail pour installer GLFW. Commencer par télécharger le code source (<https://sourceforge.net/projects/glfw/files/glfw/3.0.4/glfw-3.0.4.zip/download>), puis exécuter les instructions suivantes :

```
cd glfw-3.0.4  
rehash  
cmake -G "Unix Makefiles"  
make  
sudo make install
```

- Les fichiers suivants doivent maintenant être présents :

```
/usr/local/include/GLFW  
/usr/local/lib/libglfw3.a
```

Pour installer glew : Dans synaptic, installer les packages glew-utils, libglew-dev, libglew1.13, libglewmmx1.13

Pour installer glm : Dans synaptic, installer le package libglm-dev.

12.2 Utilisation

Le travail est séparé en deux parties : acquisition et reconstruction, que l'on retrouve dans deux dossiers différents.

12.2.1 Acquisition

Un Makefile est fait pour compiler cette partie. Il permet de compiler le code avec les différentes librairies opencv et flycap. Une fois compilé, il suffit de lancer l'exécutable créé pour lancer l'acquisition stéréo.

12.2.2 Reconstruction

Cette partie nécessite plus de bibliothèques, à cause de la reconstruction OpenGL. Le mode de fonctionnement est donc légèrement différent. Il faut se placer à l'intérieur du répertoire build, et lancer la commande `cmake ..` pour construire dans le dossier parent l'environnement d'exécution. Il faut ensuite faire la commande `make` pour compiler le programme. Il suffit alors pour lancer l'application de passer en paramètre de l'exécutable les deux vidéos à traiter pour la reconstruction 3D (dans l'ordre : la vidéo gauche puis droite), puis de se laisser guider par les instructions dans le terminal.

Si on a besoin de faire du tracking, on doit décommenter le `#define OPENCV` qui permettra de lancer le tracking sur les deux vidéos. Après lancement du programme, il y a une fenêtre de choix permettant de parcourir la vidéo frame par frame, afin de savoir à laquelle commencer et finir le tracking. Il faut ensuite quitter cette fenêtre en appuyant sur `echap`, puis revenir sur le terminal et choisir si on enregistre la vidéo et si on veut sélectionner la partie de la vidéo à traiter. Une fenêtre de sélection de la zone d'intérêt apparaît, il faut cliquer avec la souris sur l'objet à suivre, puis agrandir la zone à la souris. Celle zone de sélection peut être modifiée autant de fois que nécessaire, et un appui sur la touche espace permet de lancer le tracking. Il se passe ensuite la même chose pour la seconde vidéo. Il faut bien penser à appuyer sur `echap` dans les fenêtres OpenCV pour passer aux parties suivantes.

A partir de cela, des fichiers de points sont créés, fichiers qui seront nécessaires pour faire la reconstruction 3D avec OpenGL. Tant que les fichiers de points sont présents dans le dossier de l'exécutable, il suffit de décommenter le `#define OPENGL` pour lancer la reconstruction 3D (le `#define OPENCV` n'est alors plus nécessaire). Une fenêtre OpenGL s'ouvre, on peut déplacer l'objet dans le repère en 3D grâce à la souris et aux touches `zqsd`.

References

- [1] Julien Bossut, Myrkvid, damien200, and volfoni54. Opencv : Installation in linux (doc ubuntu). <https://doc.ubuntu-fr.org/opencv>.
- [2] GLFW dev team. Cmake avec glfw. <http://www.glfw.org/docs/latest/build.html>.
- [3] Joao Faro, Christian Bailer, and Joao F. Henriques. Tracker kcf. <https://github.com/joaofaro/KCFcpp/blob/master/src/kcftracker.cpp>.
- [4] Bernat Gabot. Opencv : Installation in windows. http://docs.opencv.org/2.4/doc/tutorials/introduction/windows_install/windows_install.html#windows-installation.
- [5] Point Grey. Synchronisation des caméras native. <https://eu.ptgrey.com/tan/11052>.
- [6] João F. Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista. High-speed tracking with kernelized correlation filters. <https://arxiv.org/pdf/1404.7584.pdf>.
- [7] João F. Henriques, Rui Caseiro, Pedro Martins, and Jorge Batista. Kernelized correlation filters. <http://www.robots.ox.ac.uk/~joao/circulant/>.
- [8] Ana Huaman. Opencv : Installation in linux. http://docs.opencv.org/2.4/doc/tutorials/introduction/linux_install/linux_install.html#linux-installation.
- [9] Sugih Jamin. Installation de glfw. <http://web.eecs.umich.edu/~sugih/courses/eecs487/glfw-howto/>.
- [10] Opengl-tutorial.org. Gestion de la caméra opengl. <http://www.opengl-tutorial.org/beginners-tutorials/tutorial-6-keyboard-and-mouse/>.
- [11] PointGrey. Flycapture sdk. <https://www.ptgrey.com/flycapture-sdk>.
- [12] OpenCV Dev Team. Opencv 2.4 documentation. <http://docs.opencv.org/2.4/>.
- [13] OpenCV Dev Team. Opencv 3.2 documentation. <http://docs.opencv.org/master/>.
- [14] OpenGL Dev Team. Opengl 4.x documentation. <https://www.khronos.org/registry/OpenGL-Refpages/gl4/>.