

UNIVERSITÉ DE BORDEAUX - REALITY TECH

**Mémoire de stage
Master 2 Informatique
Image et son**

Auteur :
Nicolas PALARD

Client :
Jérémy LAVIOLE
Référent :
Vincent LEPETIT



Résumé

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed non risus. Suspendisse lectus tortor, dignissim sit amet, adipiscing nec, ultricies sed, dolor. Cras elementum ultrices diam. Maecenas ligula massa, varius a, semper congue, euismod non, mi. Proin porttitor, orci nec nonummy molestie, enim est eleifend mi, non fermentum diam nisl sit amet erat. Duis semper. Duis arcu massa, scelerisque vitae, consequat in, pretium a, enim. Pellentesque congue. Ut in risus volutpat libero pharetra tempor. Cras vestibulum bibendum augue. Praesent egestas leo in pede. Praesent blandit odio eu enim. Pellentesque sed dui ut augue blandit sodales. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Aliquam nibh. Mauris ac mauris sed pede pellentesque fermentum. Maecenas adipiscing ante non diam sodales hendrerit. Ut velit mauris, egestas sed, gravida nec, ornare ut, mi. Aenean ut orci vel massa suscipit pulvinar. Nulla sollicitudin. Fusce varius, ligula non tempus aliquam, nunc turpis ullamcorper nibh, in tempus sapien eros vitae ligula. Pellentesque rhoncus nunc et augue. Integer id felis. Curabitur aliquet pellentesque diam. Integer quis metus vitae elit lobortis egestas. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi vel erat non mauris convallis vehicula. Nulla et sapien. Integer tortor tellus, aliquam faucibus, convallis id, congue eu, quam. Mauris ullamcorper felis vitae erat. Proin feugiat, augue non elementum posuere, metus purus iaculis lectus, et tristique ligula justo vitae magna. Aliquam convallis sollicitudin purus. Praesent aliquam, enim at fermentum mollis, ligula massa adipiscing nisl, ac euismod nibh nisl eu lectus. Fusce vulputate sem at sapien. Vivamus leo. Aliquam euismod libero eu enim. Nulla nec felis sed leo placerat imperdiet. Aenean suscipit nulla in justo. Suspendisse cursus rutrum augue. Nulla tincidunt tincidunt mi. Curabitur iaculis, lorem vel rhoncus faucibus, felis magna fermentum augue, et ultricies lacus lorem varius purus. Curabitur eu amet.

Table des matières

1	Introduction	1
1.1	Contexte et cadre	1
1.2	Objectifs	2
2	Notions	3
3	État de l'art	7
3.1	PapARt	7
3.2	Systèmes de réalité augmentée spatiale	8
3.3	Bilan	8
4	Développement d'application	9
4.1	ReARTable	9
4.1.1	Besoins de l'application	10
4.1.2	Choix et implémentation	10
4.2	Extraction de document	11
4.2.1	Besoins de l'application	11
4.2.2	Choix et implémentation	12
5	Prototype haute performance	15
5.1	Logiciel	15
5.1.1	Convolution - Théorie	15
5.1.2	Convolution - Implémentation sur GPU	16
5.2	Matériel	17
5.3	Nectar - Architecture micro services	17
Annexes		19
Annexe 1		19
1	Partie 1	19
1.1	Sous-partie 1	19
1.2	Sous-partie 2	19
1.3	Sous-partie 3	19
2	Partie 2	19
2.1	Sous-partie 1	19
2.2	Sous-partie 2	19
2.3	Sous-partie 3	19
Annexe 2		20
Prérequis	20
1	Partie 1	20
1.1	Sous-parie 1	20
1.2	Sous-parie 2	20
2	Partie 2	20
3	Partie 3	21

Introduction

Ce mémoire retracera les missions réalisées durant mon stage de Master 2 Informatique pour l'Image et le Son à l'Université de Bordeaux 1 effectué entre Avril et Septembre 2018 (6 mois) dans la société RealityTech. Ce rapport ne couvrira cependant que les 5 premiers mois du stage car la date de rendu de ce dernier précède d'un mois la date de fin du stage.

Le stage a donc été effectué chez RealityTech une jeune start-up de réalité augmentée spatiale. Issue de l'Inria de Bordeaux, l'institut national de la recherche en information et en automatique, cette dernière est la continuité d'un projet de recherche mené par Jérémie Laviole, ex ingénieur de recherche à l'Inria. Ce projet, PapARt¹ Paper Augmented Reality ToolKit, est un kit de développement (SDK) permettant de créer des expériences de réalité augmentée sous forme d'applications de projection interactive dans des feuilles de papier. Pour créer ces expériences, du matériel spécifique est nécessaire, des caméra couleurs et des caméra de profondeur utilisées pour capter le monde réel, un projecteur pour pouvoir projeter le contenu numérique et un ordinateur. Ce matériel est fourni par RealityTech qui réalise tout le travail de calibration et d'installation de l'environnement nécessaire. En parallèle de la partie matériel, la société continue de développer activement le SDK et de créer des applications, afin d'étendre les possibilités de la technologie. Le but est de pouvoir ouvrir la technologie à d'autres domaines que celui de la recherche pour mettre les outils collaboratifs, les systèmes interactifs aux services de l'éducation, du vidéo ludique, de la vulgarisation etc...

1.1 Contexte et cadre

Actuellement, RealityTech se développe dans un incubateur de start-up appelé La Banquiz² situé 4 rue Eugène et Marc Dulout, à Pessac Centre. L'objectif de La Banquiz est de promouvoir des start-up Open Source³ et innovantes en leur apportant à leur dirigeant des formations, du coaching individuel et collectif, de l'aide pour la recherche de financement et tout ce qui gravite autour de l'accompagnement de jeunes entreprises.

A ce jour RealityTech ne travail qu'avec des laboratoires de recherche (comme ...) et cherche à étendre son secteur d'activité. Les systèmes proposés par la société fournissent les résultats espérés et la dynamique de celle-ci s'oriente donc vers une commercialisation du produit. Dans cette optique, l'entreprise souhaite innover du point de vue matériel et logiciel en développant une plateforme haute performance mais aussi du point de vue utilisabilité. Actuellement le développement d'application clients se fait via un SDK Processing⁴ cependant ce dernier ne répond pas totalement aux besoins lorsqu'il s'agit de développer des applications utilisant par exemple la 3D.

Pour débuter ce stage, j'ai eu l'occasion de partir au Laval Virtual, le plus grand salon international sur la réalité augmentée et virtuelle, pendant une semaine en tant qu'exposant. J'ai pu développer une bonne connaissance du produit et ai pu observer de nombreuses technologies à l'état de l'art dans ces domaines. C'est aussi ce salon qui m'a permis de comprendre les besoins auxquels pouvait répondre une technologie comme celle de RealityTech et par la même occasion les enjeux et les apports de celle-ci. Ça a aussi été une très bonne expérience au niveau relationnel car elle a permis de créer rapidement une relation avec Jeremy Laviole, mon tuteur de stage.

Le reste de mon stage a été réalisé à la Banquiz où j'ai été au contact de toutes les entreprises officiant en son sein. J'ai notamment pu rencontrer d'autres stagiaires tel que Rémi Kressmann, développeur, et

1. <https://project.inria.fr/papart/fr/>

2. <http://labanquiz.com>

3. https://fr.wikipedia.org/wiki/Open_source

4.

Gabin Andrieux, commercial chez LockEmail, une entreprise de cyber-sécurité, mais aussi des dirigeants comme Jean François Schaff, un brillant physicien ayant créer une plateforme de télé expertise pour les professionnels de santé, avec qui j'ai énormément échangé aussi bien sur des concepts de programmation, que sur la culture de l'informatique en générale.

Problématique du sujet

1.2 Objectifs

Le déroulement du stage a été fortement guidé par les besoins de la société.

Applications de démonstration Le premier gros objectif du stage était le développement d'applications de démonstration en utilisant le produit de l'entreprise. Le but était de comprendre l'essence, le fonctionnement global du produit et ce qu'il était possible/impossible de réaliser avec celui ci. Cet objectif m'a permis d'acquérir à la fois une vision globale de l'architecture logiciel et du fonctionnement interne du kit de développement, et de l'architecture matérielle nécessaire a l'utilisation du kit. En développant ces applications de démonstration, j'ai acquis une vision globale du projet qui m'a permis d'avoir une certaine autonomie assez rapidement

Plateforme haute performance Le deuxième objectif était de réaliser une preuve de concept haute performance du produit. En effet, comme expliqué dans la partie sur le contexte (sec. 1.1), l'entreprise se lançait dans le développement d'une nouvelle plateforme haute performance. Aussi bien au niveau matériel, ordinateur, caméra, projecteur, que logiciel, algorithmes, communication inter processus, accès au matériel, il a fallut réaliser des tests complets.

Kit de développement Le dernier objectif était le développement d'un kit de développement (SDK) permettant de créer des applications de réalité augmentée spatiale ou en vue au travers utilisant la technologie de RealityTech sous Unity. Le but était de pouvoir profiter de la puissance d'Unity en tant que moteur de jeu / moteur 3D.

Notions

Le domaine d'activité qui entoure ce stage est très riche en termes de notions et de vocabulaire. Afin de mieux comprendre de quoi il va être question tout au long de ce rapport, il est nécessaire d'en définir les notions de base.

Réalité virtuelle La réalité virtuelle plus communément appelé *Virtual Reality (VR)* désigne l'ensemble des environnements purement numériques (fig 2.1), qu'ils soient réalistes ou non, dans lesquels aucune interaction avec l'environnement réel n'est possible et inversement. Cette réalité se base très généralement sur un casque *Head Mounted Display (HMD)* dont l'utilisateur doit se munir afin d'être immergé dans un monde numérique avec lequel il peut interagir. Dans la réalité virtuelle, l'immersion est une notion importante lorsqu'il s'agit de la différenciée d'un simple programme informatique.



FIGURE 2.1 – Représentation de continuum de la virtualité par Milgram et Kishino, 1995[5]

Réalité augmentée La réalité augmentée plus communément appelé *Augmented Reality (AR)* quant à elle est un sous domaine de la réalité virtuelle. L'idée de la réalité augmentée est de venir superposer à l'environnement réel des éléments virtuels. Ces éléments vont alors venir "augmenter" notre monde en apportant le plus souvent des compléments d'informations. Elle est donc qualifiée de sous domaine de la réalité virtuelle car l'utilisateur n'est plus immergé dans un environnement complètement numérique mais du contenu virtuel est ajouté en contexte à la vision réelle. Par abus de langage le terme de réalité augmentée est souvent utilisé pour parler de réalité mixte dont la notion est détaillée ci dessous. Il faut noter que ce type de réalité ne se base pas uniquement sur des *HMD* mais peut être aussi apprécié à l'aide d'un téléphone par exemple (fig 2.2).



FIGURE 2.2 – Réalité augmentée vu au travers d'un téléphone¹

Réalité mixte La réalité mixte, ou hybride, plus communément appelé *Mixed Reality (MR)*, ou *Crossed Reality (XR)*, est la fusion parfaite de l'environnement numérique et de l'environnement physique (fig 2.1). Dans ce "nouvel" environnement, les objets physiques et numériques coexistent et peuvent interagir entre eux, par exemple une table peut devenir une plateforme pour un personnage virtuel (fig 2.3).

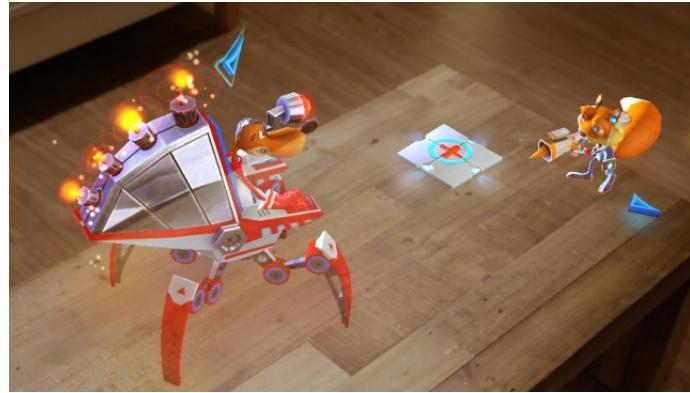


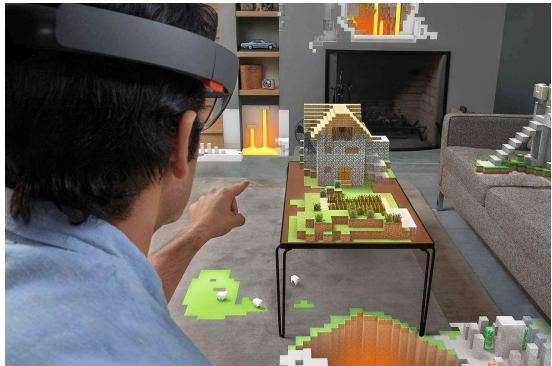
FIGURE 2.3 – Asobo StudioTM- Young Conker©

Réalité augmentée vue au travers La réalité augmentée vue au travers, plus communément appelé *See Through Augmented Reality (STAR)* est une technique de visualisation de la réalité augmentée où les éléments numériques sont vu au travers d'un écran (fig 2.4a) ou d'un *HMD* (fig 2.4b). C'est le type de visualisation le plus utilisé actuellement. L'un des défauts majeur de ce type de visualisation est que la plus part du temps, chaque utilisateur a besoin de son propre écran ou casque pour pouvoir en profiter pleinement ce qui limite grandement les expériences collaboratives. Aussi les principaux défauts liés aux écrans s'appliquent aussi, à savoir fatigue visuel etc.

1. Source : <https://www.engadget.com>



(a) Pokémons GO - Vue au travers téléphone²



(b) Microsoft HoloLens - Vue au travers casque³

FIGURE 2.4 – Réalité augmentée vue au travers

Réalité augmentée spatiale La réalité augmentée spatiale, plus communément appelé *Spatial Augmented Reality (SAR)* est une technique de visualisation qui se base sur un dispositif de projection. Les éléments virtuels qui viennent "augmenter" le monde réel sont projetés dans l'espace (fig 2.5), d'où le terme spatial. L'avantage de cette technologie est qu'elle est naturellement collaborative car les projections sont visibles sans l'aide d'aucun dispositif. Un autre atout de cette technologie est la visualisation directement sur les objets physiques manipulable ce qui permet de développer facilement des interfaces tangibles.



FIGURE 2.5 – Présentation d'une voiture en utilisant la réalité augmentée spatiale⁴

Interface tangible Une interface utilisateur tangible ou *Tangible User Interface (TUI)* est une interface utilisateur via laquelle des objets physiques, ou encore le toucher, permettent de manipuler des données numériques (fig 2.6b). Les interfaces utilisateurs tangibles remplacent très souvent les interfaces utilisateur graphiques (fig 2.6a) où *Graphical User Interface (GUI)* dans la plupart des applications de réalité augmentée car elles fournissent un contrôle direct à l'utilisateur sur ce qu'il souhaite manipuler (par opposition au contrôle indirect, comme la souris, nécessaire à la manipulation des GUI).

-
- 3. Source : Pokemon GO
 - 3. Source : Microsoft HoloLens
 - 4. Source : Google Image

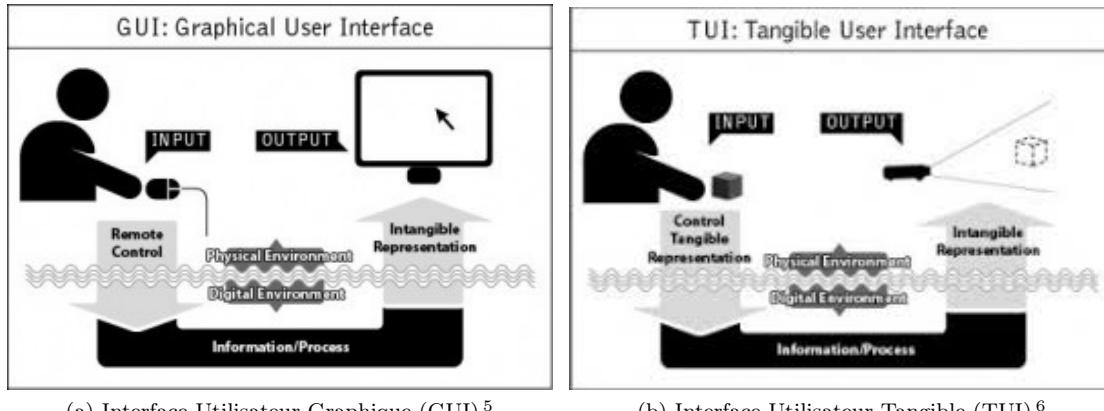


FIGURE 2.6 – Différences des interfaces utilisateurs

Calcul haute performance Le calcul haute performance ou *General-Purpose computing on Graphics Processing Units (GPGPU)* désigne une méthode de calcul utilisant la carte graphique (GPU) plutôt que le processeur (CPU). Cette technique permet de bénéficier de la puissance de la carte graphique afin de réaliser du calcul en parallèle et est très souvent utilisée pour la plupart des traitement lourd comme par exemple le rendu d'une scène 3D, l'encodage de vidéo, les simulations physiques (particules) etc. Cette technique repose sur le grand nombre de coeurs présent dans les cartes graphiques (contrairement aux processeurs) et sur la capacité de chacun de ces coeurs à effectuer des opérations simples de manière très efficace. Le calcul haute performance ne peut cependant pas se passer du CPU qui va être principalement utilisé pour récolter et transférer les données traitées ou à traiter.

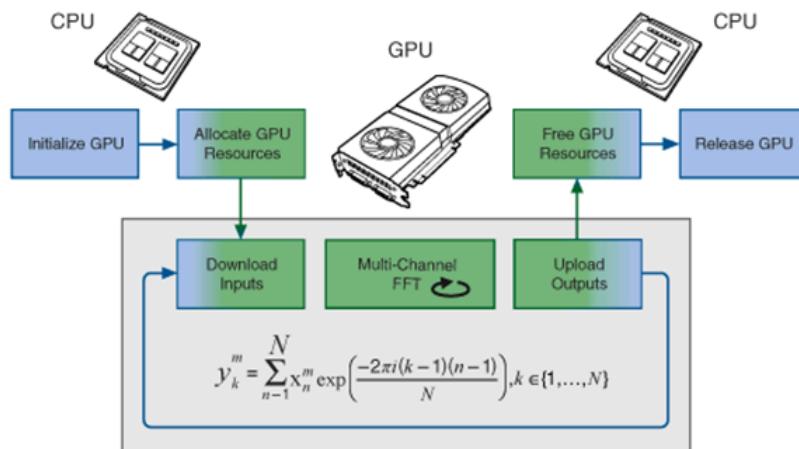


FIGURE 2.7 – Exemple de calcul de la FFT sur GPU⁷

6. Source : Icon Library - From GUI to TUI

6. Source : Icon Library - From GUI to TUI

7. Source : National Instruments

État de l'art

Intro

3.1 PapARt

Comme décrit dans l'introduction chapitre 1, PapARt ou Paper Augmented Reality Toolkit se présente sous la forme d'un kit de développement permettant de créer des applications interactives en réalité augmentée de création de dessins ou de peinture (fig 3.1). L'idée est de proposer une technique numérique non intrusive pour faciliter une tâche complexe, tel que le dessin tout en permettant à l'utilisateur de s'exprimer.



FIGURE 3.1 – Jeremy Laviole utilisant PapARt pour dessiner en réalité augmentée¹

Le système interactif (fig 3.2) permettant d'utiliser tout le potentiel de PapARt est très spécifique. Il se compose de 2 dispositifs d'acquisitions, une caméra couleur observant la zone de travail pour venir y détecter les feuilles de papier qui serviront de base à la projection. Les feuilles de papier détectées par PapARt sont ornés de marqueurs ARToolKitPlus², est une bibliothèque de détection de marqueurs fiduciaires³. Ces marqueurs une fois détectés permettent d'estimer assez précisément la position de la feuille. Une fois la feuille détectée, PapARt se charge ensuite d'interpréter les marqueurs pour y projeter le contenu adéquat comme par exemple un dessin ou un menu.

Le deuxième dispositif d'acquisition est une caméra de profondeur qui a pour rôle de détecter les différents utilisateurs et les potentiels interactions. Grâce aux informations de profondeur, les interactions peuvent être détecté soit sur le plan de la zone de travail, ce sont des interactions qualifiées de "touch", soit dans l'espace au dessus de la zone de travail, qu'on qualifiera de "pointage 3D". En plus de ces deux caméra, un projecteur est présent pour gérer toute la partie visualisation. Son rôle est de projeter dans les zones adéquates (i.e détectées via des feuilles de marqueur) le contenu numérique désiré.

Pour avoir une représentation à l'échelle comme mentionné, il est nécessaire d'avoir un calibration caméras/projecteur très précise. Cette calibration très précise permet aux système complet d'avoir des

1. Source : Inria - PapARt

2. <https://github.com/paroj/ar toolkitplus>

3. Un marqueur fiduciaire est un objet placé dans le champ de vision le plus souvent de système d'imagerie qui apparaît sur l'image produite et qui va servir de point de repère ou de référence.

capacités d'interaction et de manipulation (toucher, balayage, balayage à deux doigts) d'une tablette tactile.

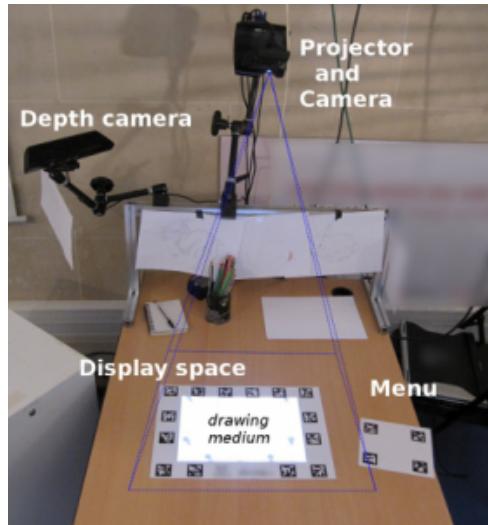


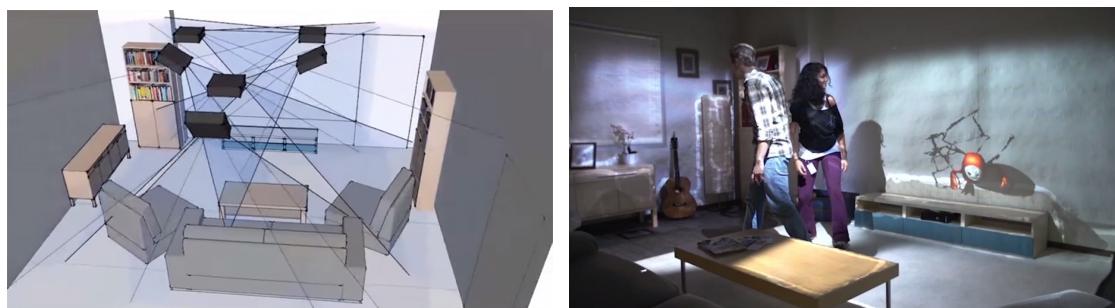
FIGURE 3.2 – Système interactif utilisant PapARt⁴

Au-delà des applications d'aide au dessin qui sont extrêmement nombreuses, PapARt ouvre un champ des possibles assez large. Le rôle de RealityTech est d'explorer ce champ des possibles en améliorant PapARt et en fournissant de nouveaux cas d'utilisation toujours plus innovants.

3.2 Systèmes de réalité augmentée spatiale

RoomAliveToolKit RoomAliveToolKit[4] est un projet tout droit sorti des laboratoires de recherche de Microsoft. RoomAliveToolKit est un kit de développement créé en 2013 par Nikunj Raghuvanshi, Eyal Ofek et Andy Wilson qui permet, à l'instar de PapARt, de créer des expériences de projection interactive. La principale différence réside dans le fait que RoomAliveToolKit a pour but de donner vie à des pièces entières en utilisant plusieurs projecteurs et plusieurs caméras qui fonctionnent en unisson.

RoomAliveToolKit a permis entre autres de développer de nombreux projets basés sur la projection interactive tel que RoomAlive, Room2Room, IllumiRoom et bien d'autre.



(a) Système de projection interactif nécessaire à l'utilisation de RoomAliveToolkit⁵

(b) RoomAlive - Démonstration⁶

FIGURE 3.3 – Microsoft Research : RoomAliveToolkit et RoomAlive

3.3 Bilan

4. Source : Inria - PapARt

6. Source : RoomAliveToolkit

6. Source : Microsoft HoloLens

Développement d'application

Comme expliqué dans l'introduction chapitre 1 le premier objectif du stage était le développement d'applications de réalité augmentée spatiale. Il était important pour commencer, d'évaluer les possibilités mais aussi les contraintes qu'offrait le kit de développement. Ainsi un travail d'analyse et de critique de l'API à été effectué en parallèle du développement d'applications.

4.1 ReARTable

D'après le contexte et le public ciblé par l'entreprise, il m'a paru intéressant de développer une démonstration a but à la fois ludique et éducatif. J'ai donc choisi de recréer une Reactable[7] proposé par la société du même nom en réalité augmentée spatiale.

La Reactable est un instrument de musique électronique permettant la génération de son en direct développé depuis 2003. Présenté sous forme d'une table interactive, le son est généré via des éléments tangibles (fig. 4.1) placés a sa surface.



FIGURE 4.1 – Élément tangible utilisé pour la génération d'un élément de synthétiseur sur la Reactable¹

Chaque élément tangible représente un élément de synthétiseur qu'il est possible de contrôler de plusieurs façon :

- La distance de l'élément par rapport a un autre élément. Cette propriété peut être utilisée pour contrôler, par exemple, l'interaction entre deux éléments.
- L'orientation de l'élément sur la table. Cette propriété peut être utilisée pour contrôler, par exemple, la fréquence de l'élément ce qui va avoir pour effet par exemple pour un battement de ralentir ou d'accélérer ce dernier.
- La disposition de l'élément. Cette propriété permet entre autre de combiner des éléments pour créer des nouveaux son plus riches et plus complexes.
- La position du doigt de l'utilisateur par rapport a un élément. On peut venir contrôler divers paramètre comme l'amplitude par exemple en venant faire graviter son doigt autour d'un élément. Ainsi, c'est en combinant plusieurs éléments entre eux avec différentes orientation et différentes dispositions que l'utilisateur va pouvoir peu a peu "construire" sa musique. Au delà de la détection des éléments tangible, la table est rétro éclairée et permet donc la visualisation en directe de la musique générée (fig. 4.2).

1. Source : Reactable : Elements tangibles



FIGURE 4.2 – Visualisation du son sur la Reactable²

4.1.1 Besoins de l'application

Le but de l'application était de proposer une démonstration ce de qu'est capable de faire le système proposé par RealityTech et non pas de créer une simulateur de musique en direct fini reprenant tous les points de la Reactable. Un tel développement pourrai faire l'objet d'un stage entier et ce n'était pas le cas ici.

Pour être en adéquation avec l'idéologie de l'entreprise, l'interface tangible et les modes d'interactions avec la musique était le point le plus cruciale. En gardant ça en tête nous avons défini les besoins fonctionnels principaux :

- Générer du son en direct.
- Créer une représentation physique du son. Chaque son ou élément sonore devait avoir une représentation physique qui lui était associé, c'est à dire, un élément ou groupement d'éléments tangibles représentant ce son.
- Déetecter des éléments physiques représentant les éléments sonores dans une image. L'application devait pouvoir détecter dans une image de caméra les divers éléments physiques présents de façon a ce qu'ils soient utilisés pour identifier les éléments sonores.
- Identifier les représentation physique des sons. Chaque élément sonore étant représenter par un ou plusieurs éléments physique, l'application devait être capable, à partir des résultats de la détection, d'identifier et de différencier des éléments sonores entre eux.
- Modifier un élément sonore. L'application devait pouvoir contrôler certains paramètre défini a l'avance de chaque élément sonore généré. Ces paramètres ont pour but d'apporter a l'utilisateur a un niveau de contrôle supérieur lors de la création de musique en direct.
- Déetecter des événements liés au toucher. Dans le cas du contrôle d'un son, l'utilisateur peut être amener a toucher des zones interactives pour déclencher divers effets.
- Crée une visualisation basique d'un son. L'application devait proposer une visualisation du son généré pour guider l'utilisateur dans son expérimentation.

4.1.2 Choix et implémentation

L'application a donc été développé avec Processing en utilisant PapARt pour la partie visualisation, détection et projection et Sonic Pi[1] pour la génération de musique en direct. Sonic Pi est un synthétiseur temps réel qui permet très facilement de générer des sons de manière cohérente. Le gros avantage de Sonic Pi est qu'il résout tout seul énormément de problèmes posé par la génération dynamique de musique comme par exemple la synchronisation des boucles, les effets d'entrée et de sortie des instruments et bien d'autre ce qui décomplexifie énormément la synthèse de musique en direct et la rend très accessible.

Comme on peut le voir sur le schéma explicatif (fig 4.3), les éléments tangibles représentants des sons se présentent sous forme de regroupement d'éléments rond de petite taille (des aimants dans notre cas). L'idée derrière ce choix est d'encourager la manipulation d'élément physique pour garder le contenu

2. Source : Reactable

numérique en contexte et favoriser la création. On peut différencier deux sons en fonction du contenu du regroupement (nombre, position et couleur des éléments regroupés).

FIGURE 4.3 – ReARtable : Schéma général

Une fois les éléments détectés, regroupés et identifiés, l'élément sonore associé peut être créé. La création d'élément sonore se fait simplement via la transmission d'un message OSC³ à un serveur Sonic Pi préalablement démarré. Ce message contient l'identifiant unique de la boucle que Sonic Pi doit démarrer. Pour chaque éléments sonores que l'application peut créer Sonic Pi possède une fonction à exécuter que nous avons préalablement créé. Toutes les communications entre l'application et Sonic Pi utilisent ce protocole ce qui permet de démarrer/arrêter/modifier certaines parties du son en directe.

Pour ce qui est du contrôle du son, une zone autour du composant est défini dans laquelle soit un élément tangible, soit une interaction physique (avec le doigt) vont être détecté et converti en interaction avec le contenu numérique (fig. 4.4).

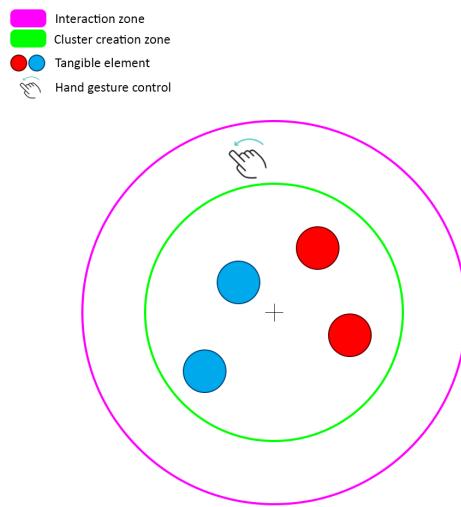


FIGURE 4.4 – Schéma représentant la création d'un son avec zone d'interaction

La dernière étape du développement de l'application était la visualisation de la musique générée (section 4.1.1). Cette étape n'a finalement pas été aboutie par manque de temps. L'idée était d'utiliser le spectre du son et les différentes fréquences qui le compose récupérable à l'aide d'une transformée de Fourier⁴ pour créer une visualisation globale basée sur les fréquences avec des variations visuelles en fonction de la hauteur, du tempo du son et tous les autres paramètres du son qu'on peut extraire.

4.2 Extraction de document

Plus tard au cours de mon stage des cas d'utilisation où l'extraction et la numérisation de document ont été abordés comme par exemple et ainsi il était judicieux de développer une preuve de concept de cette fonctionnalité sous la forme d'une application de SAR.

Le but de ce développement était d'expérimenter diverses techniques de détection de document en temps réel se basant ou non sur des connaissances à priori comme la taille du document, sa couleur, la couleur du fond (duquel il faut extraire le document), la présence d'éléments distinctifs (comme des marqueurs fiduciaires ou des ronds colorés de petite taille).

4.2.1 Besoins de l'application

- Accéder au flux vidéo d'une caméra. L'application devait avoir accès au flux vidéo d'une caméra filmant le document à détecter.

3. Le protocole OSC où OpenSoundControl est un format de transmission de données conçu pour le contrôle en temps réel

4. Opération mathématique permettant de décomposer un signal en la somme des signaux qui le compose Wikipédia - Transformation de Fourier.

FIGURE 4.5 – Document à détecter muni de marqueurs colorés

- Déetecter un document dans une image. Des images extraites du flux vidéo, l'application devait être capable, avec ou sans connaissance a priori, de détecter un document se trouvant dans cette image.
- Extraire un document d'une image. Grâce au résultat de la détection, l'application devait être capable d'extraire ce document de l'image afin d'obtenir une image ne contenant que le dit document.

4.2.2 Choix et implémentation

La détection de document est un problème connu en traitement d'image sur lequel j'avais déjà eu l'occasion de travailler lors de mon projet de fin d'études durant le deuxième semestre de mon année de Master 2.

Dans cette application nous avons mis en place plusieurs détection différentes pour essayer de trouver une solution à ce problème.

Détection de document basé sur des marqueurs colorés La première détection utilise une connaissance a priori sur le document : Le document cible est muni de lignes d'éléments ronds colorés de petite taille dans un ou plusieurs de ses coins (fig. 4.6a). Ainsi tout l'enjeu de cette détection se base dans la détection de ces éléments colorés dont il est question plus tard dans ce rapport (voir section ??).

Une fois les éléments détectés, ils sont regroupés en différentes lignes (fig. 4.6b). Une ligne est défini par un regroupement d'éléments dont l'écart entre chaque éléments ne dépasse pas une certaine distance verticale ou horizontale. L'angle de la ligne est défini par les deux premiers éléments qui la compose. Si un autre élément "dérive" il est rejeté et la ligne est créée. Cette ligne est ensuite utilisé pour calculer deux vecteurs, dont un est confondu avec la ligne et le deuxième est perpendiculaire au premier (fig 4.6c). La détection finale du document se fait en calculant l'intersection des différents vecteurs verticaux et horizontaux (fig. 4.6d)

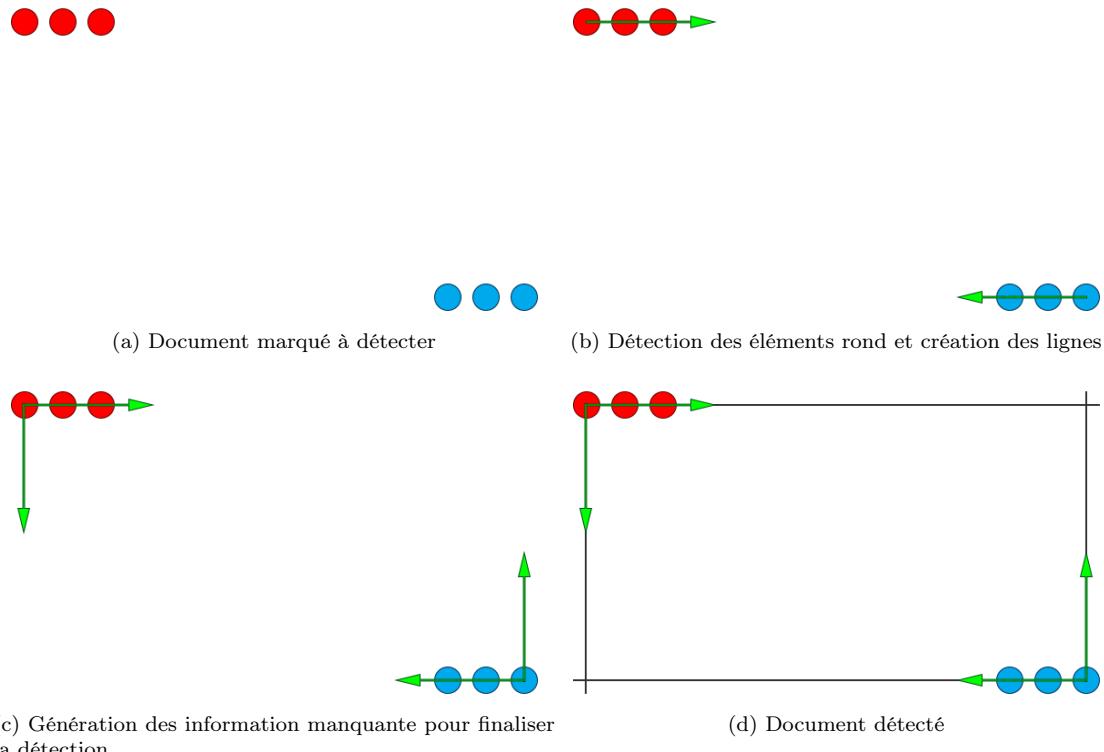


FIGURE 4.6 – Détection de document étape par étape

Comme on peut le voir figure 4.6d, lors de cette détection les bords du document sont rognés. Généralement ces zones ne contiennent aucune information car elles correspondent généralement aux

marges verticales et horizontales que chaque document possède. Cependant comme évoqué dans les cas d'utilisation, cette détection ne sera pas utiliser seulement pour des documents de type A4, on pourra s'en servir comme outil pour suivre une feuille de papier à la manière des marqueurs ARToolKit, ou pour détecter des post-it par exemple. Ainsi ces marges ne peuvent pas être ignoré dans notre cas.

Une simple connaissance a priori de la distance des éléments rond par rapport au coin permet de résoudre ce problème ou alors pour avoir une détection bien plus précise et robuste, il est possible d'utiliser des algorithmes de détection de contour couplé à des algorithmes d'extraction de lignes pour retrouver le vrai coin du document. C'est ce dont nous allons parler dans la deuxième partie.

Détection de document - Canny et transformée de Hough Sans connaissance a priori la détection de document devient un problème compliqué et bien connu du monde de l'informatique surtout lorsqu'il y a un besoin de temps réel vient s'ajouter la tâche, comme par exemple, dans une application de scan de document.

L'algorithme de Canny[2] est un filtre de détection de contour permettant d'extraire d'une image des contours (fig 4.7b) très précis respectant trois critères : La bonne détection, la bonne localisation et la clarté de la réponse. Ces trois critères en font un très bon choix dans le cadre de la détection de document ou la qualité et surtout la précision de la détection est importante pour ne pas rogner des bouts de document par exemple. Cet algorithme a cependant tendance à laisser beaucoup de bruit issue de faible contour tout de même détectés c'est pourquoi une étape de floutage visant à lisser les zones de faible contour est fortement conseillée.

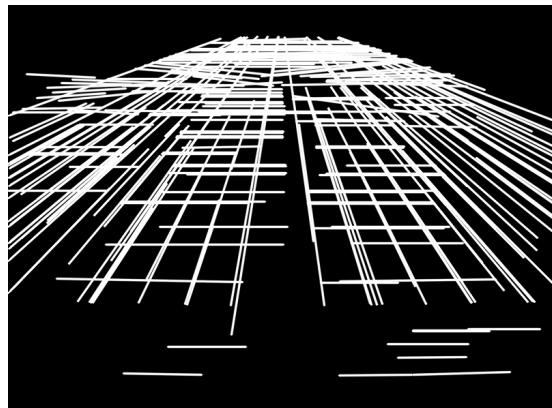
Une fois les contours détectés, il est possible d'essayer d'extraire directement le document mais c'est une tâche plus complexe, car elle requiert d'analyser les contours, qu'il est possible de faciliter en utilisant la transformée de Hough[3]. La transformée de Hough permet d'extraire n'importe quelle forme à partir d'une image contour en utilisant les propriétés mathématiques de la forme. Dans notre cas où nous souhaitons extraire des lignes droites les propriétés mathématiques utilisées correspondent aux coordonnées polaires considérés comme plus robuste que l'équation de la droite (fig 4.7c)



(a) Image originale⁵



(b) Canny - Détection de contours⁶



(c) Transformée de Hough - Détection de lignes⁷

FIGURE 4.7 – Détection de lignes : Canny + Hough

7. Source : <http://funvision.blogspot.com/2016/01/hough-lines-and-canny-edge-sobel.html>

7. Source : <http://funvision.blogspot.com/2016/01/hough-lines-and-canny-edge-sobel.html>

7. Source : <http://funvision.blogspot.com/2016/01/hough-lines-and-canny-edge-sobel.html>

Il ne reste qu'à filtrer les lignes pour trouver des potentiels documents dans une image.

Cette succession de traitement est cependant lourde et peut difficilement être effectué en temps réel sur des images de haute résolution. Dans notre cas, nous nous sommes servis de ces algorithmes seulement sur des parties d'image (de petite résolution) de façon à améliorer une première détection grossière effectuée par exemple, à l'aide de marqueurs colorés. Une fois la première détection effectuée, nous obtenons une position plus ou moins précise des quatre coins nécessaires à l'extraction du document. Nous utilisons cette information sur la position potentiel des coins pour extraire dans des sous images centrées sur ces coins (fig. 4.8a). Ensuite nous appliquons les deux algorithmes mentionnés plus tôt, à savoir, la détection de contour (fig. 4.8b) puis la détection de ligne (fig. 4.8c). Nous filtrons ensuite le résultat de la détection de ligne pour obtenir exactement une ligne verticale et une ligne horizontale. Une fois ces deux lignes trouvées, nous calculons les équations de droite (pente, et constante) associées pour pouvoir en calculer l'intersection et finalement trouver le coin dans cette sous image (fig. 4.8d).

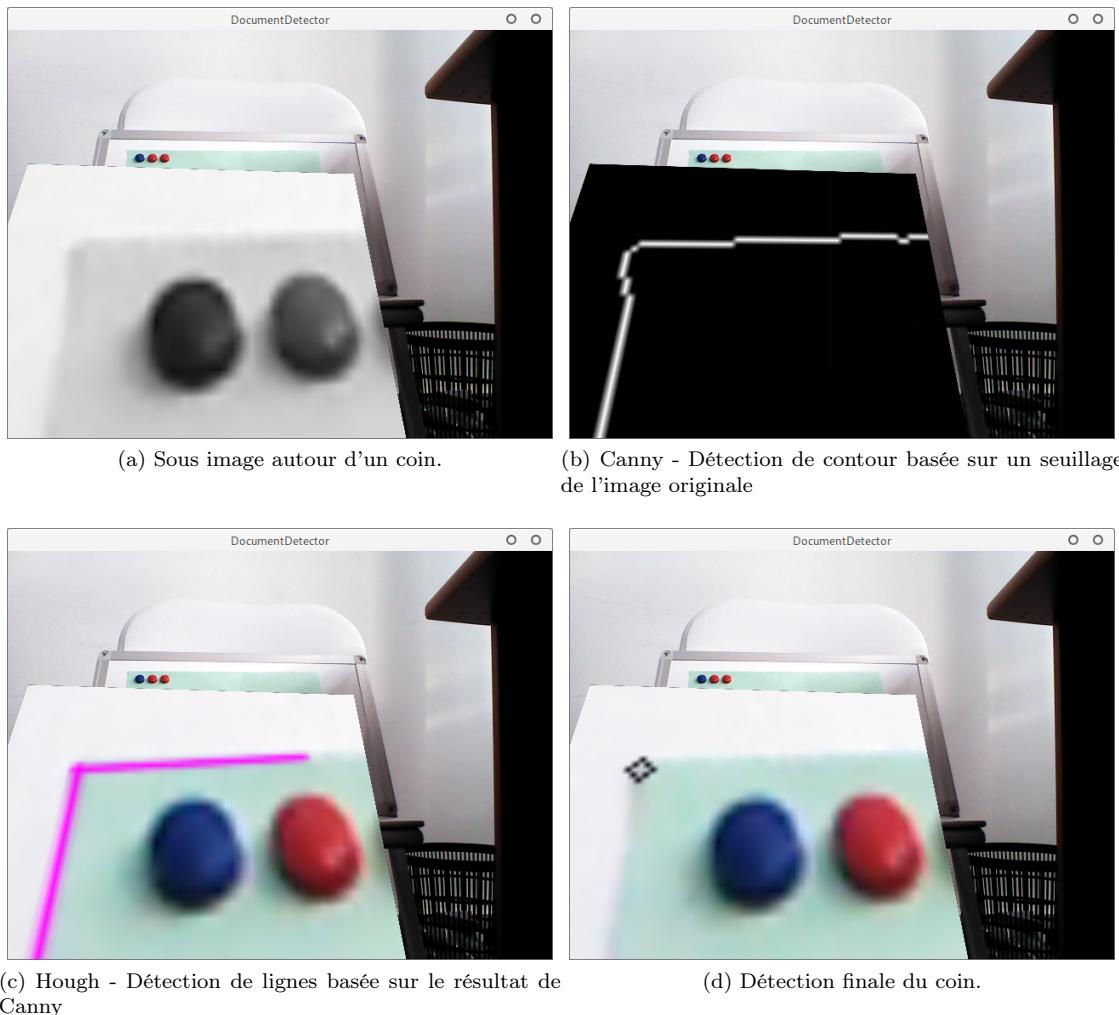


FIGURE 4.8 – Affinage de la détection d'un coin du document

Bilan Pour finir, vous trouverez Figure un rapide comparatif des différents résultats obtenus. On peut observer quelques points notables : En utilisant Hough et Canny pour améliorer la détection d'un coin, l'extraction est plus fidèle au document réel. En effet, les coins détectés sont plus précis que des hypothèses effectuées à priori. Cette méthode est cependant moins rapide car elle requiert de nombreux calculs en plus. Lorsqu'il est utilisé en temps réel, on peut cependant observer que la détection avec connaissance à priori du modèle est bien plus robuste, car elle ne dépend pas d'une deuxième détection qui a beaucoup de chance d'échouer (seuillage, détection de contour, détection de ligne, intersection entre deux droites puis obtention finale du coin) ainsi la méthode à utiliser variera avec les cas d'utilisations. Par exemple dans le cadre d'une application de scan de document, une méthode d'extraction de document plus précise sera envisagée, mais dans le cadre d'une estimation de pose (3D)⁸ on pourra préférer une détection rapide et robuste.

8. Pose : Wikipédia

Prototype haute performance

Durant mon stage, j'ai été amené à développer un prototype haute performance des outils que propose RealityTech. Pour atteindre cet objectif, le développement de ce prototype s'est déroulé selon plusieurs axes. Il y a eu tout d'abord une partie logiciel où le but était d'optimiser les algorithmes de traitements d'image et donc avoir des gains de performances, la partie matériel où là nous nous sommes intéressés de plus près à la latence des caméra, la rapidité de l'encodage et la puissance de calcul théorique des ordinateurs. Pour finir nous nous sommes attaqués à l'architecture générale du logiciel que propose RealityTech, et nous avons prototypé un système de micro services.

5.1 Logiciel

L'optimisation logiciel, de nos jours, passe souvent par des développements très spécifiques visant à améliorer un point crucial d'une application. Dans notre cas, la plupart des algorithmes gourmands en ressources sont ceux que nous utilisons pour le traitement des images obtenues grâce au flux vidéo de la caméra couleur présente dans le matériel. Le traitement que nous sommes amenés à réaliser le plus souvent est le produit convolution d'une image et d'un filtre.

5.1.1 Convolution - Théorie

En mathématiques, le produit de convolution est un opérateur bilinéaire et un produit commutatif, généralement noté $$, qui, à deux fonctions f et g sur un même domaine infini, fait correspondre une autre fonction $f * g$ sur ce domaine, qui en tout point de celui-ci est égale à l'intégrale sur l'entiereté du domaine (ou la somme si celui-ci est discret) d'une des deux fonctions autour de ce point, pondérée par l'autre fonction autour de l'origine — les deux fonctions étant parcourues en sens contraire l'une de l'autre (nécessaire pour garantir la commutativité).¹*

En traitement d'image, le produit de convolution est effectué dans le domaine discret où f et g représentent respectivement une image et un filtre et dont le résultat est une nouvelle image. On appelle filtre, un noyau de convolution, le plus souvent de petite taille utilisé pour appliquer un traitement afin de modifier une image (fig 5.1). Étant donné le domaine discret, la convolution d'une image par un filtre définit pour chaque pixel sa nouvelle valeur en sommant la valeur des pixels voisins (avec un voisinage de la taille du filtre) au pixel d'origine pondéré par la valeur du filtre, qui est lui centré sur le pixel d'origine, à la même position que le pixel voisin (fig 5.2).

1. Source : Produit de convolution - Wikipedia

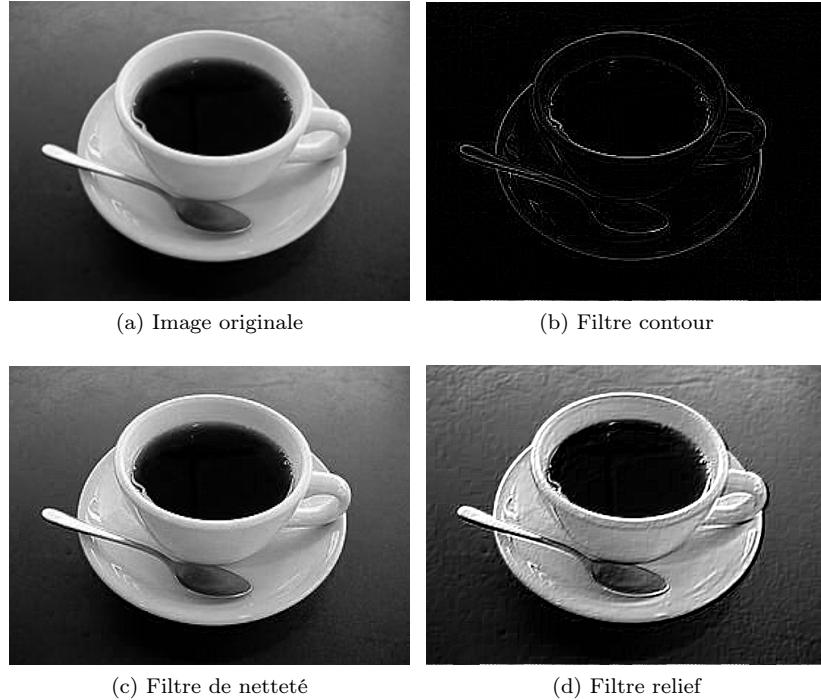


FIGURE 5.1 – Différentes filtres de convolution appliquée à une image.

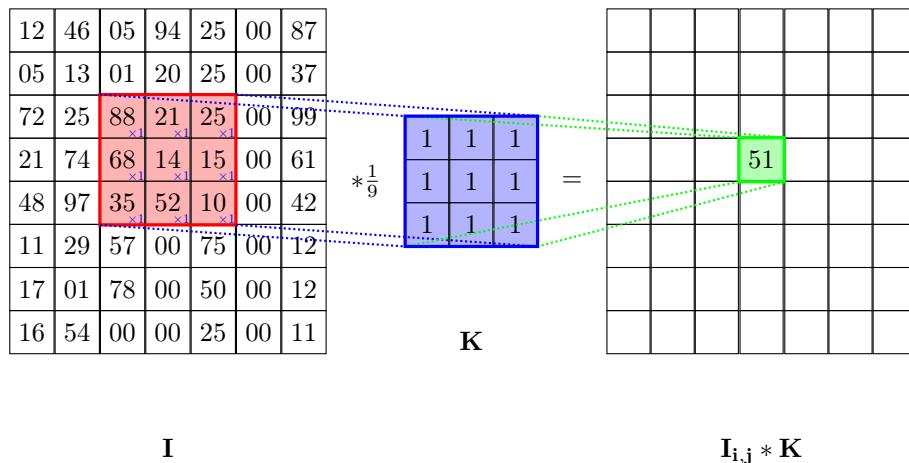


FIGURE 5.2 – Convolution d'une matrice (image) (I) par un filtre (K)

Comme on peut s'en rendre compte, l'image résultat est une nouvelle image, totalement indépendante de l'image d'origine dont chaque pixel est complètement indépendant de ces voisins. C'est cette propriété qui va nous intéresser car cela signifie qu'il est possible de calculer chaque pixel de l'image résultat dans n'importe quel ordre et qu'avec des moyens de calculs suffisants il est possible de calculer tous les pixels de l'image résultat en une seule fois. On dit d'une telle fonction qu'elle est parallélisable. Ce fort potentiel de parallélisation fait de ce traitement une cible de choix pour l'optimisation et donc le calcul sur carte graphique (section 2 - Calcul haute performance).

5.1.2 Convolution - Implémentation sur GPU

Il existe de nos jours plusieurs langage de programmation graphique nous permettant de mettre à profit la puissance de notre GPU. Nous avons choisi d'implémenter notre algorithme de convolution en OpenCL OpenGL ESet CUDA

OpenCL OpenCL ou *Open Computing Language* est un language de programmation basé le C créé par **Khronos Group** en 2009. Un programme OpenCL est composé de deux partie : La partie **code hôte**

qui est l'application principale et qui va se charger d'orchestrer les différentes tâches. Cette partie est écrite en C avec des appels à l'API OpenCL pour gérer tout ce qui allocation de mémoire et gestion des périphériques OpenCL. La partie **noyau** ou **code périphérique** qui représente un programme qui va être exécuté sur un périphérique. Cette partie est écrite en OpenCL-C et représente les calculs à effectuer. Il faut aussi bien différencier hôte, qui dans notre cas sera le processeur et qui sert à allouer la mémoire et transférer les données au périphérique, et périphérique(s) qui représentera ici une ou plusieurs carte graphique et qui sera utilisé pour appliquer des traitements aux données qui lui sont transmises.

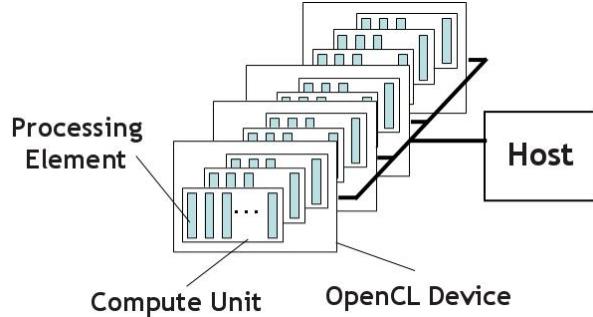


FIGURE 5.3 – Schéma OpenCL - Hôte et périphériques²

OpenCL est compatible avec la plupart des systèmes et des architectures aujourd'hui présents sur le marché sans aucune modification de code nécessaire. Cet avantage est aussi l'un de ses plus gros inconvénients car il ne permet pas d'exploiter au mieux chaque architecture comme peut le faire CUDA avec NVIDIA, et les performances de ce dernier ne sont donc pas équivalentes sur chaque architecture.

OpenGL ES OpenGL ES est une sous partie d'OpenGL, qui est une interface de programmation multi plateforme et multi langage, créée pour le rendu 2D/3D. Même si OpenGL ES n'a pas été créé pour la programmation sur carte graphique à proprement parler elle offre des outils intéressants tel que les shaders, des petits programmes informatiques exécutés sur carte graphique. Tout comme pour OpenCL, une application hôte va être chargée d'allouer la mémoire (buffer), et de transmettre des données à ces shaders qui seront ensuite exécutés par le *pipeline* de rendu graphique.

5.2 Matériel

5.3 Nectar - Architecture micro services

2. Source : <https://www.anandtech.com/show/7334/a-look-at-alteras-opencl-sdk-for-fpgas/>

Annexes

Annexe 1

Intro

1 Partie 1

Bla

1.1 Sous-partie 1

Bla

1.2 Sous-partie 2

Bla

1.3 Sous-partie 3

Bla

2 Partie 2

Bla

2.1 Sous-partie 1

Bla

2.2 Sous-partie 2

Bla

2.3 Sous-partie 3

Bla

Annexe 2

Intro

Prérequis

Bla
— item1;
— item2;
— item3;
— item4.
Bla

1 Partie 1

Bla

1.1 Sous-parie 1

Bla

1.2 Sous-parie 2

Bla

2 Partie 2

ATTENTION !
Texte d'avertissement

Bla

3 Partie 3

Bla

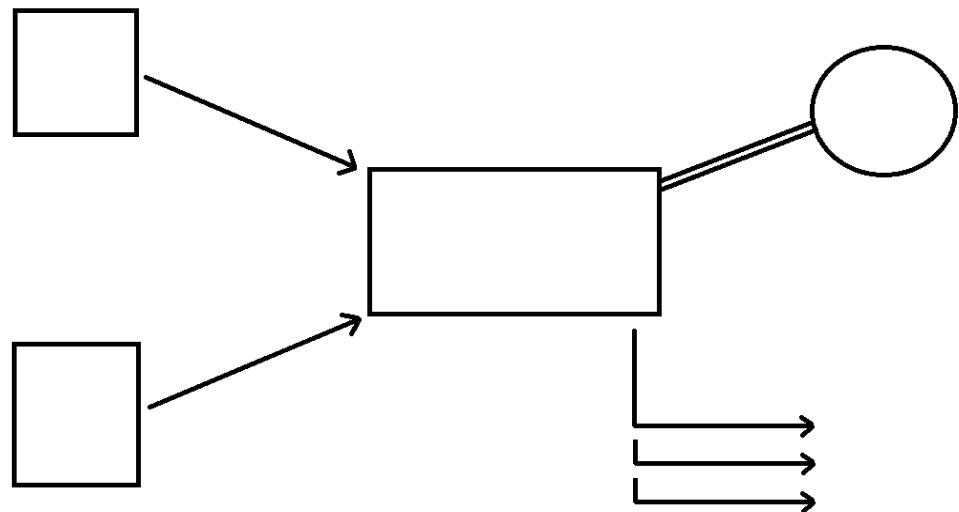


FIGURE 5.4 – Presentation schema

Paragraphe 1

Bla

Paragraphe 2

Bla

Paragraphe 3

Bla

Bibliographie

- [1] Sam Aaron. Sonic pi - the live coding music synth for everyone, 2003.
- [2] John Canny. A computational approach to edge detection. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 1986.
- [3] Richard O Duda and Peter E Hart. Use of the hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1) :11–15, 1972.
- [4] Brett Jones, Rajinder Sodhi, Michael Murdock, Ravish Mehra, Hrvoje Benko, Andrew Wilson, Eyal Ofek, Blair MacIntyre, Nikunj Raghuvanshi, and Lior Shapira. Roomalive : Magical experiences enabled by scalable, adaptive projector-camera units. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST ’14, pages 637–644, New York, NY, USA, 2014. ACM.
- [5] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. Augmented reality : A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*, volume 2351, pages 282–293. International Society for Optics and Photonics, 1995.
- [6] NVIDIA Corporation. NVIDIA CUDA C programming guide, 2010. Version 3.2.
- [7] Reactable. Reactable experience - music knowledge technology, 2003.
- [8] Unity Technologies. Unity. <https://unity3d.com/>. Last accessed 23 March 2018.