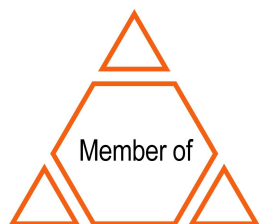


# Rapport de stage chez VICOMTech:

## High Throughput Screening in vivo sur embryons de poissons zèbres



INI-GraphicsNet

COLLIN Nicolas  
TN09-Février 2007  
GB03 - A06

# REMERCIEMENTS

En premier lieu, je souhaiterais remercier Julián Flórez et Jorge Posada, respectivement Directeur général y directeur adjoint de VICOMTech (Centre de Technologies d'interaction visuelle et communications) pour m'avoir accueilli dans leur entreprise.

Plus spécialement, je voudrais remercier Céline Paloc, ma maîtresse de stage à VICOMTech, qui m'a supervisé, m'a aidé et m'a permis de réaliser ce stage dans les meilleures conditions, et Fernando Boto pour sa grande disponibilité et pour avoir partagé son expérience et ses connaissances avec moi.

Enfin, je souhaiterais remercier tout le personnel de VICOMTech pour m'avoir accueilli dans leur espace de travail, pour m'avoir permis de m'intégrer à la vie de l'entreprise et de partager une ambiance de travail très agréable.

CONFIDENTIAL

# SOMMAIRE

<b>1</b>	<b>PRESENTATION.....</b>	<b>4</b>
1.1	LE GROUPE INI-GraphicsNET.....	4
1.2	VICOMTECH.....	4
<b>2</b>	<b>LE PROJET ZEBRASCREEEN.....</b>	<b>6</b>
2.1	PRESENTATION.....	6
2.2	OBJECTIFS.....	7
<b>3</b>	<b>TACHES REALISEES.....</b>	<b>8</b>
3.1	INTRODUCTION.....	8
3.2	PREMIERE PARTIE : L'EVALUATION.....	8
3.3	SECONDE PARTIE : EXTRACTION DE LA FREQUENCE CARDIAQUE.....	11
<b>4</b>	<b>CONCLUSION.....</b>	<b>14</b>

CONFIDENTIEL

# 1 Présentation

## 1.1 Le groupe *Ini-GraphicsNet*

Le groupe Ini-graphicsNet (International Network of Institutions for Advanced Education, Training and R&D in Computer Graphics Technology, Systems and Applications) constitue un des plus grands centres mondiaux en matière de nouveaux moyens de communication y formes de communication.

Les recherches de INI-GraphicsNet se concentrent principalement dans des travaux d'application orientés recherche mais aussi dans le développement de technologies. Le réseau travaille beaucoup en collaboration avec les universités locales. Grâce à sa structure décentralisée, INI-GraphicsNet a accès à un large champ de compétences et de résultat de recherches de niveau international.

Le groupe est composé de nombreux centres répartis dans le monde entier :

Darmstadt (Allemagne), Rostock (Allemagne), Guimarães (Portugal), Providence y Omaha (USA), San Sebastián (Espagne), Séoul (Corée du Sud), Singapour y Trento (Italie). Sur ces sites, plus de 370 employés et presque 550 chercheurs réalisent un volume de recherche annuel de 40 millions d'euros. Enfin, le groupe est aussi constitué de Spin-off qui développent des prototypes résultants des recherches afin de créer des produits potentiellement introduisible sur le marché.

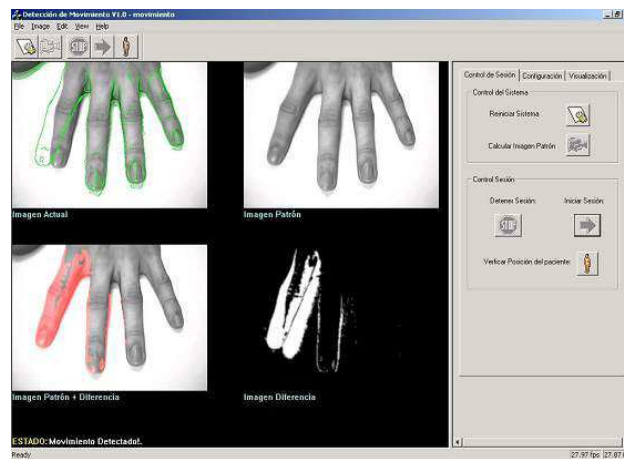
## 1.2 *VICOMTech*

L'Association Centre de technologies d'interaction Visuelle et de communication – VICOMTech – est un centre de recherche appliquée qui travaille dans le domaine des technologies graphiques et multimédias. C'est une association à but non lucratif fondée en 2001 par la fondation INI-GraphicsNet de Fraunhofer-IGD et le groupe de radio et télévision basque EiTb (Euskal Telebista - Televisión Vasca S.A, y Eusko Irratia - Radiodifusión Vasca, S.A.).

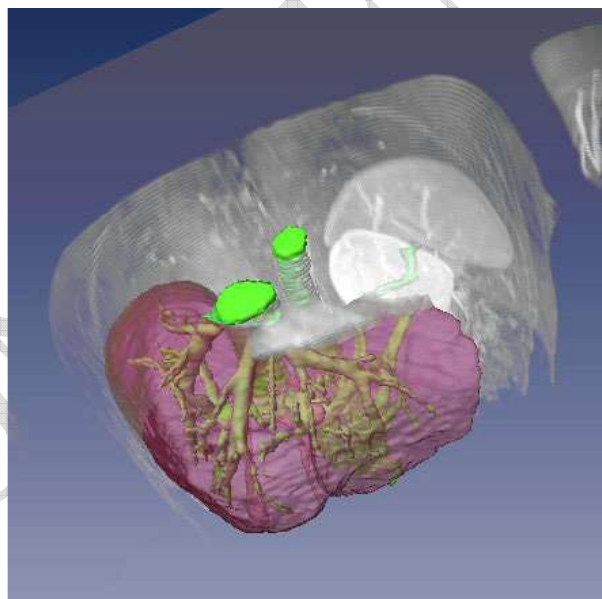
Les domaines d'applications se divisent en 5 unités de recherche :

- Télévision digitale et services interactifs
- Applications médicales et technologies de l'information et des communications (TICS) pour la santé.
- Reconstruction virtuelle du patrimoine archéologique et historique, Système d'information géographique (SIG) et tourisme.
- Activités ludoéducatives, éducation, génie linguistique.
- Multimédia et applications industrielles.

Dans le domaine des technologies biomédicales, il se développe des systèmes médicaux basés sur les technologies de dernière génération. Les grandes lignes de recherche incluent une analyse d'images médicales 2D/3D (traitement d'image digital, segmentation, ...), une gestion de l'information clinique (transmission, standards, interfaces), et des technologies de visualisation (réalité virtuelle et augmentée), vision par ordinateur et capture de mouvement.



**Figure 1 : Capture de mouvements.**



**Figure 2: Modèle 3D du foie et ses principaux vaisseaux sanguins.**

Ces technologies s'utilisent pour développer des applications comme les systèmes de télémédecine, des outils d'aide au diagnostique et à la chirurgie et des systèmes de suivi de patients.

C'est dans ce domaine d'application de VICOMTech que j'ai réalisé mon stage.

## 2 Le projet Zebrascreen

### 2.1 Présentation

Aujourd'hui, les techniques existantes dans le domaine de l'imagerie permettent une acquisition d'images de haute résolution des organismes biologiques. Cependant, l'analyse de ces images reste majoritairement l'inspection visuelle et manuelle afin d'identifier des caractéristiques fonctionnelles ou morphologiques (phénotypes). Le manque d'utilisation de l'informatique dans l'identification de ces caractéristiques constitue un frein à l'avancé de la recherche biologique et biomédicale, plus particulièrement dans les domaines de la génomique et des découvertes de nouveaux médicaments.

Le processus d'élaboration d'un nouveau médicament se déroule en général en 10 à 15 années entre sa découverte et sa mise sur le marché. La phase de test pré-clinique est une des plus longue et des plus coûteuse. L'accélération de cette phase est donc un enjeu très important pour l'industrie pharmaceutique.

Le projet zebrascreen cherche à développer un système qui automatise la réalisation d'essais ou de « screening » in vivo sur des embryons de poissons zèbres. L'innovation réside dans le caractère in vivo de la recherche. Ce type de système automatisé existe déjà sur des échantillons in vitro, sur des cellules généralement. Le système requiert des méthodes avancées d'analyse d'image pour l'identification et la classification automatique des caractéristiques aussi bien morphologiques que fonctionnelles des organes et/ou des tissus.



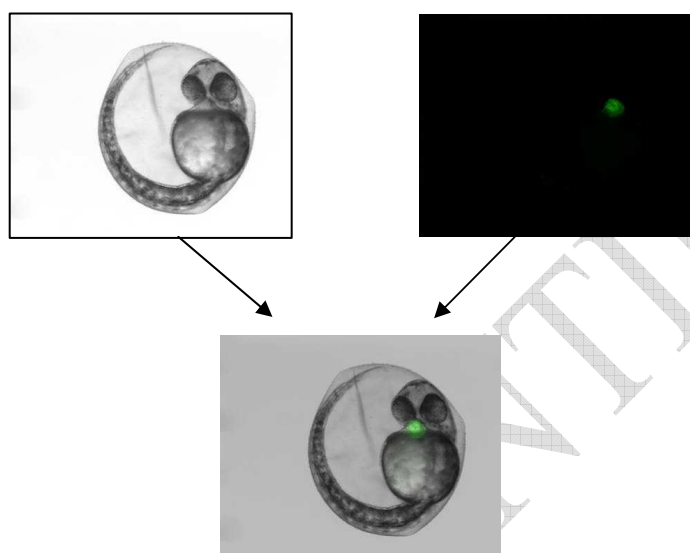
Figure 3: embryon de poisson zèbre après 19h de développement.

Le choix du poisson zèbre comme modèle animal in vivo repose sur plusieurs avantages que présente celui-ci :

- C'est un vertébré et par conséquent, il présente une grande homologie génétique avec le reste des vertébrés, y compris l'Homme (jusqu'à 85 % d'homologie).
- Les embryons sont transparents et les organes internes peuvent s'observer facilement.
- Grâce à la grande connaissance de son génome, le poisson zèbre est un organisme qui présente une haute traçabilité génétique.
- La capacité à générer des mutants à grande échelle permet d'avoir un potentiel d'obtention de modèles de maladies (ou d'anomalies) plus grand, plus rapide et plus varié qu'avec d'autres modèles animaux.
- Le cœur d'embryon de poisson zèbre se développe en 48h.

## 2.2 Objectifs

L'objectif essentiel de ce projet est l'automatisation de tous le processus de traitement des échantillons. Les avantages que présentent ce système sont : l'obtention de hauts rendements en terme de nombre d'échantillons analysés (High Throughput screening), identification quantitative et qualitative de phénotypes, ainsi qu'une information phénotypique objective (indépendante de l'observateur puisque faite par machine).



**Figure 4 : Image résultant de la superposition de l'image acquise sous lumière normale (champ clair) et celle acquise sous fluorescence.**

Le projet cherche donc des méthodes avancées d'analyse et de traitement d'image afin d'identifier et de classer les phénotypes. Ainsi, des statistiques peuvent être fournies à partir d'une base d'images de plusieurs milliers d'images sur les effets d'une substance.

Le projet se réalise dans le cadre d'une collaboration internationale d'entreprises, de centres de recherche et d'universités. Le rôle de VICOMTech dans ce projet est de réaliser toute la partie analyse d'image, c'est-à-dire, l'extraction des données (des caractéristiques phénotypiques, morphologiques et fonctionnelles) mais aussi la classification des données (association des caractéristiques extraites des images à des phénotypes prédéfinis). Toutes les images utilisées sont des images issues de microscopes à fluorescence. Il s'observe uniquement le cœur de l'embryon marqué à la GFP (Green Fluorescence Protein). De plus, tout ces processus doivent être automatisés ce qui ajoute une contrainte supplémentaire.

## **3 Tâches réalisées**

### **3.1 Introduction**

Le stage se décompose en deux thèmes de travail. Un premier est l'évaluation de logiciels existants de High Throuput Screening en se basant sur différents critères (possibilités de traitements d'images, de mesures sur les images, d'automatisation, de segmentation, etc.). Un second est l'implémentation et l'intégration d'algorithmes de traitement d'image dans les différents logiciels afin de réaliser une étude comparative des différentes solutions possibles au niveau de l'efficacité, de la rapidité des processus etc.

Le travail s'est réalisé à partir du mois de septembre jusqu'au mois de janvier. Les deux parties ont eu à peu près la même durée, soit deux mois et demi.

### **3.2 Première partie : l'évaluation**

La première partie du stage consiste d'une part à évaluer les logiciels existants selon certains critères et d'autre part à rédiger un document officiel correspondant à cette évaluation afin d'orienter les choix de fonctionnement du système final. La partie de traitement d'images a un rôle primordial dans le projet puisqu'il va permettre l'extraction des données phénotypiques qui nous intéressent. Cette étape s'inscrit dans une suite de procédés: Acquisition–traitement-classification (base de donnée). Pour ce faire, plusieurs solutions sont possibles :

- Utiliser un unique logiciel commercial
- Associer plusieurs logiciels
- Créer un Application

C'est pourquoi une évaluation des technologies existantes en ce domaine et notamment des logiciels commerciaux a été réalisée. Dans le cadre d'une association de différentes applications commerciales, la compatibilité entre les différents éléments du système doit être prise en compte et gérée. Cela va être un point clé de l'évaluation des programmes, observer les communications possibles entre chacun d'entre eux (au niveau des formats de fichiers par exemple).

L'évaluation s'est faite sur trois logiciels différents : AxioVision (Carl Zeiss), MetaXpress™ (Molecular devices) y Acapella™ (Evotec), avec l'objectif de trouver les avantages et inconvénients de chacun d'eux, les comparer et choisir l'outil le mieux adapté pour réussir le High Throughput Screening (HTS) ou High Content Screening (HCS) in vivo. L'aspect in vivo est un élément clé pour comprendre la difficulté de réaliser l'analyse d'image de façon automatisée. Les logiciels existants sont développés pour travailler sur des images isolées, des cellules par exemple ou des éléments relativement statiques. Or, dans le cas de l'analyse in vivo de cœurs d'embryon, il ne s'agit plus seulement d'obtenir des informations de type « statiques » (morphologie) mais aussi dynamiques (fréquence cardiaque). C'est pourquoi, des vidéos vont être utilisées et non plus des images « isolées ». L'enjeu est d'essayer d'adapter et/ou étendre les outils disponibles pour les images aux vidéos.



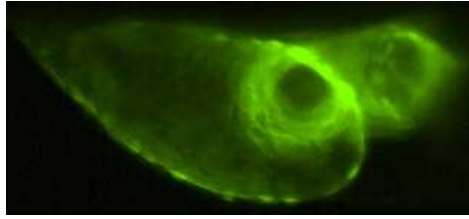


Figure 5 : cœur d'embryon de cœur de poisson zèbre (fluorescence).

Bien que tous les logiciels soient développés pour réaliser du High Throughput Screening, chacun a ses propres caractéristiques et ses orientations. Par exemple, AxioVision est plutôt orienté vers l'acquisition d'images en série, Acapella vers son caractère modulaire, sa capacité à ajouter et intégrer des modules externes.

La logique adoptée a été d'évaluer chaque logiciel individuellement et ensuite de les comparer et d'observer les interactions possibles entre eux.

Les principaux paramètres utilisés pour l'évaluation individuelle sont :

- Les ressources systèmes requises
- Les formats d'entrée et de sortie
- Les capacités des modules de traitement d'image
- Possibilité d'extraire des données et des statistiques.
- Interface utilisateur.
- Possibilité de base de données

Un des points clé de l'analyse d'image repose dans un premier temps sur la reconnaissance du cœur dans l'image, et ensuite, sur la distinction des deux chambres cardiaques, le ventricule et l'oreillette. Ces processus de reconnaissance d'objet dans l'image fait appel à ce qu'on appelle la « segmentation » d'image. Cette segmentation dépend beaucoup de la qualité de l'image fournie en entrée. La difficulté de travailler avec des échantillons in vivo va apparaître ici avec le caractère aléatoire des positions des embryons dans leurs puits.

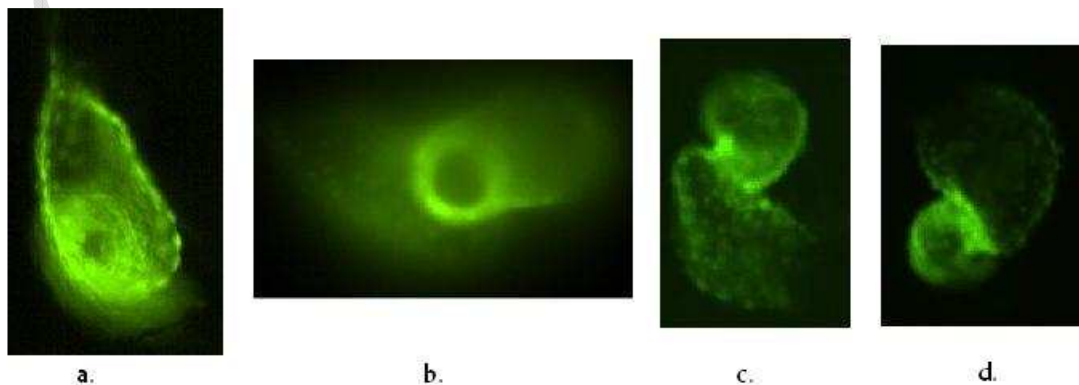
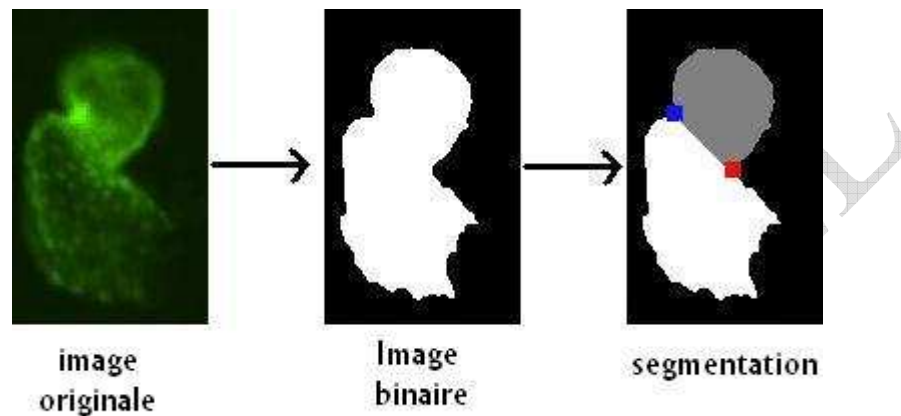


Figure 6 : Différents cœurs d'embryons de poissons zèbres avec diverses orientations

On peut voir, à l'œil nu, que certains cas vont poser moins de problème que d'autres pour distinguer les deux chambres cardiaques. En plus de l'orientation du cœur par rapport à l'objectif du microscope, la diffusion du cœur (en terme de netteté des contours) va jouer un rôle important dans la capacité à segmenter et à faire des mesures de qualité par la suite (volume, aire des chambres, etc.). L'image 6.b par exemple ne présente pas des contours nets et la détection du cœur dans l'image va être compromise ne sera pas bonne alors que la 6.c donnera de meilleurs résultats. La netteté d'une image influence sa binarisation qui joue directement sur la segmentation.



**Figure 7 : Etapes pour arriver jusqu'à la segmentation des cavités cardiaques.**

Toute l'évaluation consiste à comprendre le fonctionnement de chaque module disponible, les tester avec les vidéos et voir les résultats qu'ils fournissent avec celles-ci. Le problème des logiciels existants est qu'il est difficile d'obtenir une bonne segmentation sans changer les paramètres des processus de traitements de l'image. Il ne faut pas oublier que l'objectif est d'automatiser le système dans son ensemble, et que par conséquent, il ne peut y avoir qu'une seule et unique manière de traiter les images, indépendante de la nature de la vidéo d'entrée (orientation du cœur, qualité vidéo, etc.). C'est pourquoi, afin d'adapter et d'améliorer les capacités des logiciels commerciaux et pallier les éventuelles lacunes dans un domaine du traitement de l'image (pour la segmentation par exemple), la génération de nouveaux modules utilisables à l'intérieur même du logiciel peut représenter un puissant outil. Cependant, les modalités d'usage de ces modules diffèrent d'un logiciel à un autre (en terme de langage de programmation, intégration au logiciel, etc.) ce qui est pris en compte dans leur évaluation.

Un autre aspect prépondérant de l'évaluation des logiciels a été d'observer les possibilités d'automatisation du traitement d'images. Il s'agit de reproduire pour chaque vidéo une même séquence d'instructions (comme un script) afin de la traiter entièrement ; c'est-à-dire, segmenter, réaliser les mesures, et intégrer dans une base de données. Ainsi, pour l'ensemble des embryons analysés dans une expérimentation, le traitement se répète automatiquement pour tous les embryons sans que l'utilisateur n'ait à intervenir ce qui assure une certaine rapidité d'exécution ainsi qu'une stabilité de l'analyse (identique pour chaque échantillon).

Tout le travail concernant l'évaluation s'est donc porté, tout d'abord, sur la compréhension de l'enjeu de chaque paramètre d'évaluation, le test des capacités de chaque logiciel pour tous les critères et enfin la réalisation d'un rapport officiel en espagnol à ce propos ( $\approx 40$  pages), et apportant quelques orientations pour le choix du futur système. Tout ce travail a duré plus de deux mois.

### 3.3 Seconde partie : extraction de la fréquence cardiaque

Après l'évaluation des logiciels, des essais ont été menés pour réaliser l'étape d'extraction du rythme cardiaque indépendamment des logiciels commerciaux. Pour cela, un programme de base, avec son environnement, a été créé afin de pouvoir greffer les modules dessus. Ces modules correspondent aux différentes méthodes utilisées pour calculer la fréquence cardiaque. Mais avant de développer ces modules, une étude a été réalisée afin de déterminer les méthodes possibles et optimales pour extraire le signal du battement cardiaque.

L'étude comporte deux aspects: un sur le signal en lui-même et un autre sur le traitement du signal. En premier lieu, il s'agit de savoir à partir de quel signal on va calculer la fréquence cardiaque : c'est-à-dire, si on va observer de l'aire du coeur, le périmètre, l'intensité moyenne, le coefficient de corrélation avec un cercle, etc. En second lieu, il faut déterminer avec quels outils on va extraire la fréquence du signal et avec quels paramètres. Pour cela, 24 méthodes différentes ont été testées sur 4 vidéos de durées différentes présentant des orientations distinctes. Cette expérimentation a été réalisée avec les outils d'un des logiciels existants (AxioVision) afin de créer par la suite ses propres algorithmes en se basant sur les résultats de l'étude.

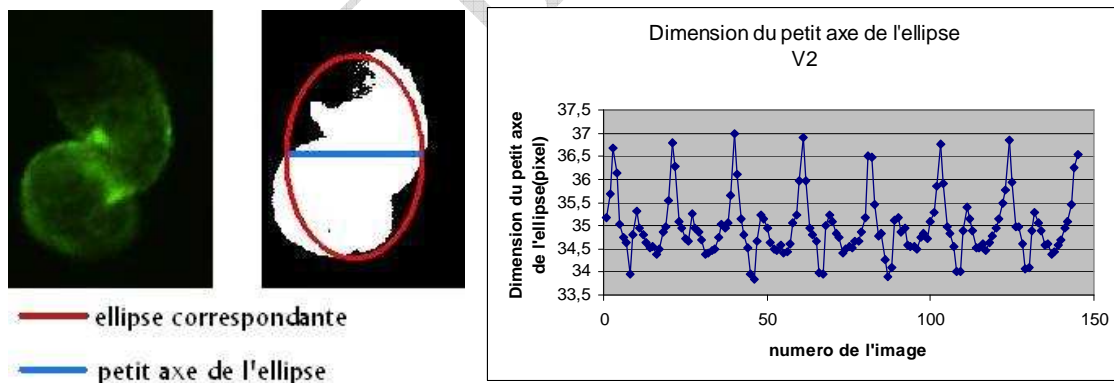


Figure 8 : A gauche, schéma de la mesure du petit axe de l'ellipse. A droite, signal temporel de cette mesure.

Toutes les mesures réalisées dans le cadre de cette étude se sont faites à partir d'images dites « masque » (image binaire permettant de restreindre les mesures à une zone de l'image originale et non à l'image entière. Cette zone est appelée « région d'intérêt »). Ceci implique que ce masque doit être créé par une étape binarisation de l'image d'origine et c'est à partir de ce masque que les mesures seront faites. Par exemple, sur l'illustration 8, le signal acquis est la longueur du petit axe de l'ellipse ayant le même moment géométrique d'inertie que le cœur. A partir du signal obtenu, on applique une transformée de fourrier.

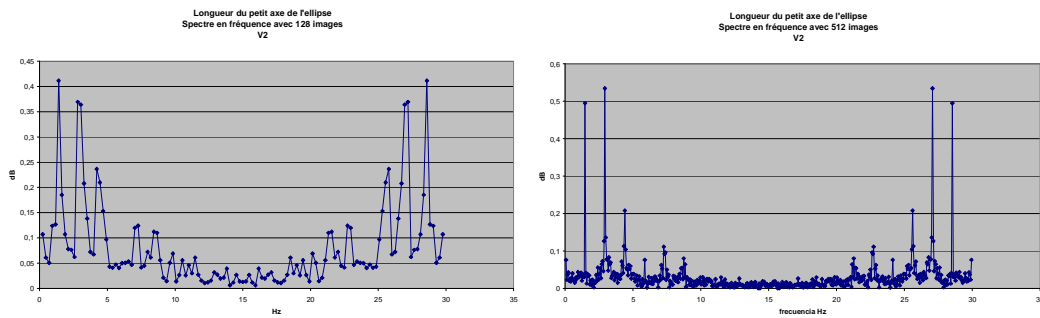


Figure 9 : Spectres en fréquences correspondant au signal de la Figure 8 : A gauche, schéma de la mesure du petit axe de l'ellipse. A droite, signal temporel de cette mesure.

C'est à ce niveau qu'on entre dans l'étude du traitement du signal, à savoir, le nombre de points nécessaire pour obtenir d'une part la bonne fréquence, et avec une bonne précision. Pour trouver la fréquence à partir du signal, on utilise une transformée de Fourier. Le tout est de savoir combien de point utilisé pour la calculer (sur le plan de la rapidité et des résultats). C'est pourquoi plusieurs solutions ont été testées : 128, 256, 512 et 1024 images. Les résultats varient selon les vidéos en entrée et selon les méthodes utilisées pour obtenir le signal.

Au terme de cette étude, quelques méthodes ont donné de meilleurs résultats que d'autres et en se basant sur les principes de celles-ci, des algorithmes ont été implémentés et intégrés dans le programme de base de VICOMTech. Le développement du programme s'est fait en C++ (Microsoft Visual .NET).

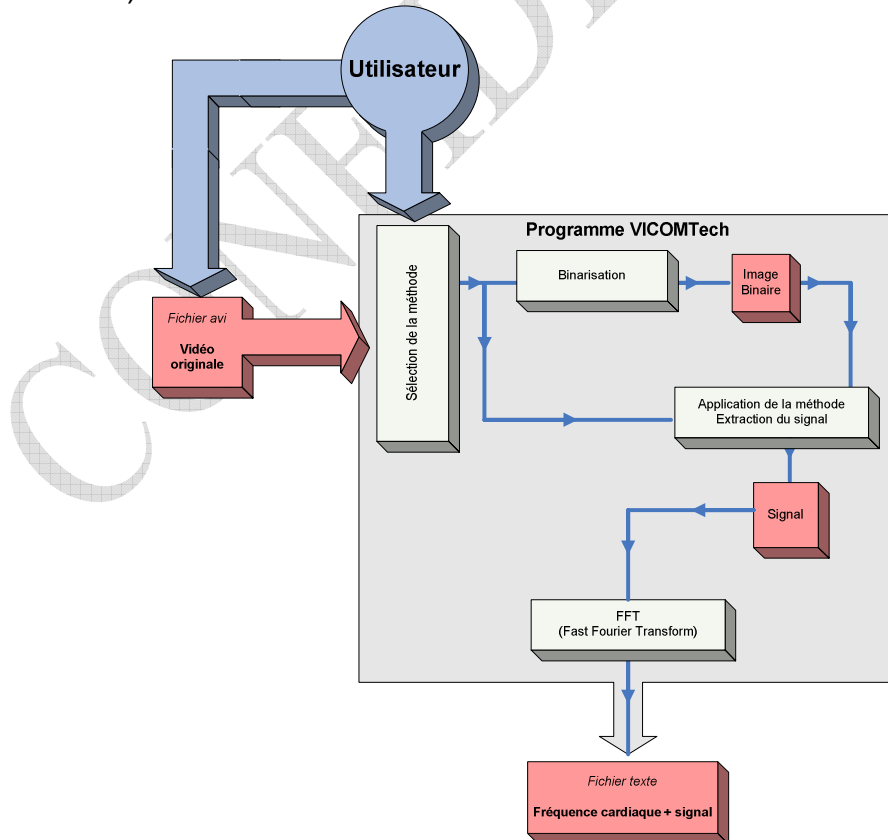


Figure 10 : Schéma d'organisation du programme de VICOMTech fournissant la fréquence cardiaque à partir d'une vidéo.

Le fonctionnement du programme est décrit par le diagramme de l'illustration 10. Il se compose de différents éléments :

- Une interface graphique.
- Un outil de sélection des fichiers.
- Un outil de sélection des méthodes de calcul.
- Les instructions de calcul de la fréquence cardiaque.

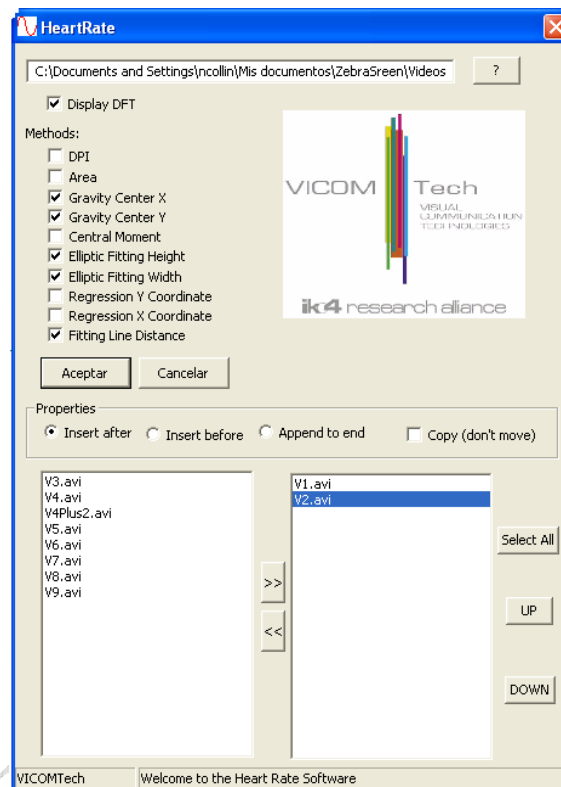


Figure 11 : Interface graphique du programme de VICOMTech.

Pour implémenter les algorithmes d'extraction de signal, des bibliothèques de fonctions ont été utilisées, et en particulier la bibliothèque OpenCV qui est spécialement développée pour le traitement d'image. Il a donc pu être possible, en combinant des fonctions, de créer des méthodes d'extraction du signal du pouls, certaines s'apparentant aux méthodes déjà utilisées dans les applications commerciales mais beaucoup plus adaptées au problème et à l'hétérogénéité des vidéos d'entrée ; d'autres méthodes ont été entièrement créées. Chaque procédé développé a été testé, a évolué et été amélioré en fonction des comparaisons avec les résultats obtenus de l'étude.

L'objectif est de ne conserver qu'une seule et unique méthode dans le programme, celle qui fonctionnera le mieux, avec un maximum de vidéos différentes. Les tendances ont été adoptées après les premiers résultats mais les expérimentations suivent leur cours.

## 4 Conclusion

Ces 6 mois de stage à VICOMTech m'ont permis de découvrir de nombreux aspects du monde du travail, de la recherche scientifique et du développement de nouvelles technologies. J'ai eu l'opportunité de travailler sur plusieurs aspects d'un même projet, me permettant d'avoir une vue d'ensemble du projet et pas uniquement l'aspect technique sur lequel j'ai travaillé.

Dans la première partie du stage, j'ai dû m'adapter tout d'abord au mélange des langues. En tant que français, j'ai évalué des programmes en anglais, alors que la langue de travail est l'espagnol et que les documents se rédigeaient également en espagnol. En réalisant l'évaluation, chaque paramètre contrôlé a entraîné une recherche pour en comprendre son fonctionnement afin être capable de juger chaque logiciel. De nombreux tests et expérimentations ont été menés dans ce cadre là. Cette première partie peut donc se résumer en trois étapes clés qui ont été, la recherche, l'évaluation et la rédaction d'un document officiel (décrivant l'évaluation).

La seconde partie mettait en jeu, à la fois des thèmes de recherche et de développement ; la recherche avec la réalisation de l'étude sur l'extraction du signal, et le développement avec l'implémentation des processus permettant l'obtention du signal du battement cardiaque et leur intégration dans un programme créé par VICOMTech. Ces deux aspects permettent d'acquérir une méthodologie de travail très précieuse pour tout type de travaux futurs. L'utilisation des connaissances en traitement de signal (avec les FFT) a été combinée avec la programmation afin de réaliser ce programme fonctionnel et efficace.

A partir des travaux accomplis, les étapes futures sont la réalisation de mesures différentes (non seulement des mesures de fréquence cardiaque mais aussi de morphologie, etc.) sur chacune des chambres du cœur (oreillette et ventricule) ainsi que la mise en place de la classification phénotypique après analyse.

A travers toutes ces étapes, j'ai pu acquérir à la fois des connaissances dans les divers domaines techniques, mais aussi des méthodes de travail. De plus les échanges avec les autres ingénieurs et chercheurs ont été d'une grande aide et une source d'expérience pour mener à bien mon stage dans les meilleures conditions possibles.