



RAPPORT DE STAGE DE FIN D'ÉTUDES

STRUCTURATION DE L'ESPACE ET SÉLECTION ADAPTATIVE DES CARACTÉRISTIQUES D'OBJETS/BLOBS DANS UNE SCÈNE VIDÉO

Étudiant : Oussama EL HAMZAOU

Encadrant : Michel COLLOBERT

Résumé

Ce document présente un rapport de mon stage de fin d'études effectué dans le laboratoire de recherche TECH/IRIS/VIA du centre France Télécom R&D de LANNION (22), FRANCE.

Le sujet du stage est centralisé sur la détection de visages humains dans une scène vidéo. Les méthodes et algorithmes utilisées sont inspirés dans la majorité du fonctionnement du cerveau humain. La particularité de ce projet est qu'il vise un taux de réussite de détection pouvant atteindre 99% et une fréquence d'exécution temps réel (25 images par seconde).

Le programme applique une série de filtres sélectifs sur chaque image d'une vidéo afin de localiser les visages humains. La disposition en cascade des filtres utilisés permet la compensation des défauts de chaque filtre.

Dans ce stage, le travail effectué est divisé en deux parties. La première consiste à développer et améliorer quelques filtres. La deuxième partie (non présentée dans le présent document) vise la création d'un module intelligent capable d'optimiser la vitesse d'exécution et la qualité des résultats du programme.

Mots clés

Détection, neuroscience, visage, blob, vision

Abstract

This document presents a report on my training period at the research lab TECH / IRIS / VIA of France Telecom R&D center in LANNION (22), FRANCE.

The topic of the project is centralized on the detection of human faces in a video. The methods and algorithms used are based on the human brain working. The particularity of this project is that it seeks a detection rate of up to 99% and a frequency of real-time performance (25 frames per second). The program applies a series of selective filters on each frame in a video to locate human faces. The cascade filters system used allows each filter lacks compensation.

My work is divided into two parts. The first aims to develop and improve some filters. The second part (not included in this document) aims to create an intelligent module able to optimize execution speed and quality of the program results.

Keywords

Detection, Neuroscience, face, blob, vision

Remerciements

Je tiens à exprimer mes remerciements et ma gratitude à tous ceux qui ont voulu apporter l'assistance nécessaire au bon déroulement de ce stage.

En particulier l'expression de ma très grande reconnaissance à Monsieur Michel COLLOBERT pour la confiance qu'il m'a accordée en acceptant de m'encadrer, la pertinence de ses conseils, sa rigueur scientifique, son soutien sans faille, sa disponibilité pour répondre à mes questions et à mes attentes et pour l'attention qu'il a bien voulu apporter à ce rapport à divers stades de son élaboration.

Pour la même occasion, j'adresse mes remerciements à Monsieur Dominique PAVY, responsable de l'unité de recherche et développement VIA, et à tous les membres de l'équipe VIA, pour leur bonne humeur quotidienne, et pour leurs conseils.

Je remercie également Monsieur Luc JAULIN, mon cher professeur et mon tuteur à l'école ENSIETA, pour son suivi et ses conseils.

Contents

1	Introduction	5
2	Contexte général du projet	6
2.1	Groupe France Télécom	6
2.2	Présentation du projet	9
3	État de l'art	10
3.1	Présentation de l'application	10
3.2	Explication du principe de fonctionnement	11
3.3	Présentation de la tâche à accomplir	12
4	Point de départ : système de vision humain	13
4.1	L'anatomie de l'œil	13
4.2	La rétine	13
4.3	Les champs récepteurs	16
4.4	Les chemins de traitement de l'information visuel	19
5	Modules inspirés	21
5.1	Algorithme de détection du mouvement	21
5.2	Filtres d'extraction des caractéristiques du visage humain	25
5.3	Les chemins de traitement	27
5.4	Modules supplémentaires	29
6	Étapes suivantes	32
7	Conclusion	33

List of Figures

1	Interface de l'application de détection de visage	10
2	Application de différents filtres en cascade	11
3	Scéma de l'œil	13
4	La rétine	14
5	Schéma de la rétine	16
6	Champs récepteur d'une cellule bipolaire	17
7	Réponse d'une cellule bipolaire à centre ON	18
8	Réponses d'une cellules simple	19
9	Application de l'algorithme de détection du mouvement	21
10	Application d'un niveau de seuillage adéquat	22
11	Principe de détection du mouvement	22
12	Schématisation de l'algorithme de détection	24
13	Détection du mouvement avec l'algorithme amélioré	25
14	Réponses d'une cellule de type ON/OFF	26
15	Extraction des formes horizontales	27
16	Sensibilité de l'œil humain	28
17	Organisation des espaces de couleurs dans le projet	29
18	Détection horizontale des yeux	30
19	Détection verticale des yeux	30
20	Processus de détection de la zone des yeux	31

1 Introduction

Systèmes de reconnaissance, systèmes de sécurité, optimisation de capteurs audio ou vidéo...La détection de visages humains dans une scène vidéo constitue une étape importante dans une large palette de projets.

Le stage s'inscrit dans le cadre d'un projet visant le développement d'un programme performant, rapide et efficace pour la détection de visages. La stratégie de recherche s'appuie sur l'étude et l'analyse du fonctionnement du cerveau afin de s'en inspirer pour développer les algorithmes de détection.

Dans ce rapport je vous expose le déroulement du stage ainsi que le travail que j'ai effectué. Je commence par une bref introduction de l'entreprise et du laboratoire où se déroule le stage (le laboratoire TECH/IRIS/VIA), suivie par une description de l'état actuel du projet. Ensuite je résume les principes de fonctionnement du cerveau humain dont je me suis inspiré pour effectuer mon travail. Dans la partie suivante je détaille le travail réalisé ainsi que les résultats obtenus. Je fini par une présentation des étapes suivantes dans le projet.

Dans certaines parties du rapport, surtout au niveau de l'explication du fonctionnement du cerveau, j'ai essayé de ne pas effectuer beaucoup de changements aux informations trouvées dans les sources de l'étude bibliographique. Les spécialistes des neurosciences qui ont écrit ces documents expliquent clairement et précisément, et je ne voulais pas compliquer la compréhension de certains mécanismes en modifiant leurs explications.

2 Contexte général du projet

2.1 Groupe France Télécom

- *Présentation du groupe*

À partir de 2006, Orange devient la marque commerciale unique du groupe France Télécom, un des principaux opérateurs de télécommunications dans le monde.

Le groupe sert plus de 172 millions de clients sur les cinq continents au 31 mars 2008, dont les deux tiers sous la marque Orange. Le groupe a réalisé un chiffre d'affaires consolidé de 52,9 milliards d'euros en 2007 (13 milliards d'euros au 31 mars 2008). Au 31 mars 2008, le groupe comptait 111,9 millions de clients du mobile et 12 millions de clients haut débit (ADSL).

Lancé en juin 2005, le programme NExT (Nouvelle Expérience des Télécommunications) permet au groupe de poursuivre sa transformation d'opérateur intégré afin de faire de France Télécom l'opérateur de référence des nouveaux services de télécommunications en Europe. En 2006, Orange est ainsi devenue la marque unique du groupe pour l'Internet, la télévision et le mobile dans la majorité des pays où le groupe est présent. Dans le même temps, Orange Business Services est devenue la marque des services offerts aux entreprises dans le monde. France Télécom est le troisième opérateur mobile et le premier fournisseur d'accès Internet ADSL en Europe et parmi les leaders mondiaux des services de télécommunications aux entreprises multinationales.

- *Activités du groupe*

Les services mobiles

Les services de communication pilotent les activités mobiles du Groupe partout dans le monde. Avec Orange, l'univers du mobile s'ouvre à de nouveaux horizons, avec des services enrichis, simplifiés et convergents qui entrent, eux aussi, dans l'ère du Haut Débit.

Les services à domicile

Les services de communication résidentiels pilotent la stratégie du Groupe sur le fixe et l'Internet dans le monde. Cela couvre tous les produits et services de la maison, du téléphone fixe à l'Internet, de la Livebox à la télévision numérique, en passant par les accès et services Haut Débit.

Les contenus accessibles partout

Pour offrir le meilleur du numérique à ses clients, le Groupe propose un ensemble de contenus, toujours plus riches et innovants accessibles à n'importe quel moment.

La convergence

Les services de communication sont désormais accessibles quel que soit le réseau (fixe, mobile, Internet) ou le terminal utilisé (téléphone familial ou personnel, ordinateur à domicile ou au bureau, télévision). C'est la convergence, cela procure à l'utilisateur plus de simplicité.

Les services au bureau

Accroître la performance des entreprises, faciliter le travail de leurs collaborateurs, telles sont les missions d'Orange sur ce marché. L'offre mondiale Orange Business Services fait du Groupe un partenaire pour les entreprises qui recherchent une expertise et un accompagnement vers la convergence entre voix et données et dans leur usage des services IP (Internet Protocol).

- *Un groupe international*

La marque Orange est présente en :

- Europe (Belgique, Espagne, France, Moldavie, Pays-Bas, Pologne, Roumanie, Royaume-Uni, Slovaquie, Suisse)
- Afrique (Sénégal, Mali, Côte d'Ivoire, Cameroun, Guinée équatoriale, Botswana, Madagascar, Île Maurice)
- Moyen-Orient (Égypte, Jordanie)
- Asie et Caraïbes (République dominicaine, Vietnam, Nouvelle-Calédonie et Vanuatu)

Les filiales du groupe sont présentes dans 146 pays qui permettent à l'offre Orange Business Services d'être présente chez 3750 entreprises multinationales clientes et 8 millions d'utilisateurs du mobile. Grâce à cette offre, le groupe France Télécom est le n°1 mondial des fournisseurs d'accès réseaux IP-VPN, et le n°1 européen des solutions de services en communications pour les entreprises.

- *France Télécom R&D*

La R&D est la source principale d'innovation pour le groupe avec plus de 8500 brevets à son actif. Elle est composée de 3 800 chercheurs répartis dans 15 laboratoires sur trois continents (8 en France et 7 laboratoires répartis en Europe, Amérique et Asie). Son implantation permet de détecter l'ensemble des ruptures technologiques, et de favoriser les partenariats. La division Recherche & Développement a pour principales missions :

- de développer des produits et services pour le groupe, en respectant la qualité de service
- de dégager de nouvelles sources de croissance

- d’anticiper les révolutions technologiques et d’usage
 - d’imaginer dès maintenant les solutions du futur
- *L’unité d’accueil TECH/IRIS/VIA*
- Le Centre de Recherche & Développement (CRD) Technologies (TECH) a la responsabilité de :
- détecter et identifier les ruptures affectant le Groupe dans les domaines des technologies de la voix, du son, de l’image, des terminaux, des nouvelles interfaces et modes d’interaction, des nano et biotechnologies communicantes, ainsi que dans les évolutions des usages, de la perception client et des business models
 - veiller et assurer au Groupe, si nécessaire en partenariat avec des industriels ou des organismes académiques, un positionnement compétitif sur les technologies stratégiques et la normalisation associée
 - développer et fournir les briques technologiques des nouveaux services intégrés du Groupe
 - contribuer au développement des revenus de France Télécom par la protection intellectuelle et la valorisation de ses travaux et expertises

Le laboratoire IRIS "Image, Rich media, nouvelles Interactions et hyperlangageS" dirigé par Alexandre Nolle, a pour missions principales :

- d’éclairer l’avenir du Groupe pour le domaine de l’Image (audiovisuel, rich media, réalités virtuelle et augmentée) en étant au meilleur niveau mondial de Recherche sur les technologies de compression, d’indexation audiovisuelle et de la 3D en réseau (géovisualisation, avatars)
- de développer, en lien avec les autres CRDs, les solutions technologiques liées à l’image, au partage et au travail collaboratif, via l’étude et la mise en oeuvre de nouvelles interactions pour des services innovants
- d’œuvrer au rapprochement des mondes de l’Audiovisuel (ISO/MPEG) et du Web (W3C), et de contribuer au développement de travaux de normalisation internationale tels que MPEG-21 pour mieux répondre aux besoins de l’opérateur intégré

L’Unité de Recherche et Développement (URD) VIA (Vision par ordinateur, Interface collaborative, réalité Augmentée), dirigée par Dominique Pavy, apporte une triple compétence en vision par ordinateur, en travail collaboratif et en analyse d’usages et ergonomie. Depuis 2005, elle pilote un programme de recherche interdisciplinaire au sein de FT R&D sur la téléprésence et

l'immersion. Ce programme permet de coupler avec une analyse prospective des usages, les technologies de réalité augmentée et mixte, le développement d'interfaces collaboratives, et les dispositifs de communications interpersonnelles.

2.2 Présentation du projet

Dans le cadre du laboratoire VIA, le projet vise la détection d'objets dans une scène vidéo. On cherche surtout à localiser et suivre le visage humain à travers des séries de filtres extracteurs de différentes caractéristiques.

L'originalité du projet réside dans la méthode utilisée pour développer les algorithmes d'extraction. En effet, le point de départ est le cerveau humain. On essaie d'exploiter les résultats des études menées sur le fonctionnement des différents modules du cerveau afin de s'en inspirer pour créer l'application.

3 État de l'art

3.1 Présentation de l'application

La figure FIG.1 présente la fenêtre principale de l'application avec deux vues sur les images captées par les caméras (gauche et droite).

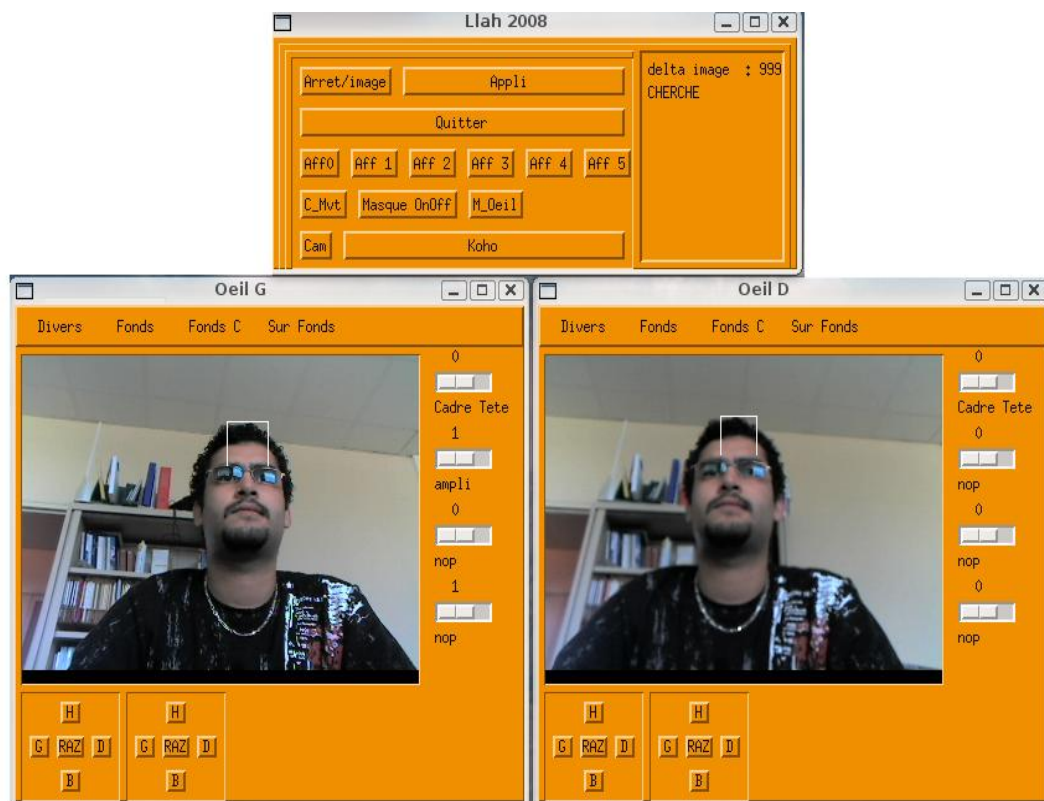


Figure 1: Interface de l'application de détection de visage

On peut voir que la fonction de détection est déjà développée. Mais la précision de la détection et le suivi du visage restent à améliorer. La fenêtre principale offre plusieurs options à l'utilisateur, elle permet notamment l'ouverture de plusieurs vues sur la scène afin de pouvoir comparer la qualité de capture entre la caméra gauche et celle de droite par exemple. L'option d'arrêt de l'image permet de fixer le flux vidéo sur une image précise, ce qui offre la possibilité de tester différents filtres plus efficacement.

3.2 Explication du principe de fonctionnement

L'application met en cascade plusieurs filtres afin de pouvoir détecter efficacement le visage humain. Ainsi, les imperfections de chacun des filtres sont compensées par les qualités des autres filtres. La figure FIG.2 présente un exemple de détection effectué à partir de la mise en cascade de quelques filtres. La source de traitement est un flux vidéo présentant un corps en mouvement. La détection du visage se fait sur plusieurs étapes.

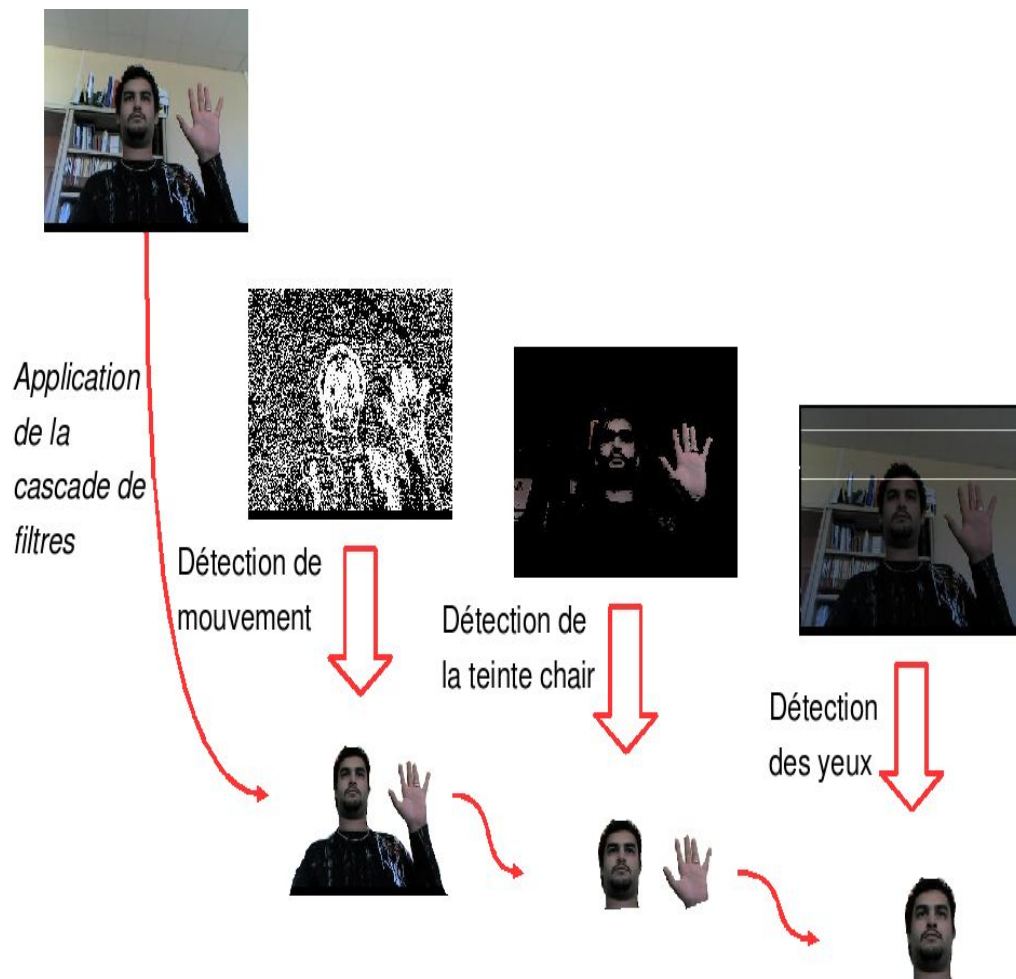


Figure 2: Application de différents filtres en cascade

L'application des filtres en cascade élimine petit à petit les éléments qui ne font pas partie du visage humain, pour obtenir à la fin du processus une détection précise du visage.

3.3 Présentation de la tâche à accomplir

Le travail à faire durant le stage est divisé en deux grands axes :

1. Améliorer et créer des filtres
2. Développer un module intelligent d'application des filtres

La première partie du travail consiste à créer et améliorer différents filtres de détection (mouvement, teinte chair, formes géométriques...). Durant la deuxième partie, on développe un module intelligent qui peut choisir le type de filtres à appliquer afin de gagner en précision et en temps d'exécution. En effet, lorsqu'un sujet est situé par exemple devant un mur qui a les mêmes informations de teinte que le visage humain, l'application du filtre détecteur de teinte chair est inutile, donc il vaut mieux passer cette étape pour ainsi gagner en temps d'exécution et éviter les problèmes de précision.

4 Point de départ : système de vision humain

Dans cette partie, je vais essayer de présenter d'une manière précise le système de vision biologique dès la rétine jusqu'aux centres de traitement dans le cerveau.

4.1 L'anatomie de l'œil

L'œil est l'organe de la vision . Il est de faible volume (6.5 cm^3), pèse 7 grammes, et a la forme d'une sphère d'environ 24 mm de diamètre, complétée vers l'avant par une demi-sphère de 8 mm de rayon, la cornée.

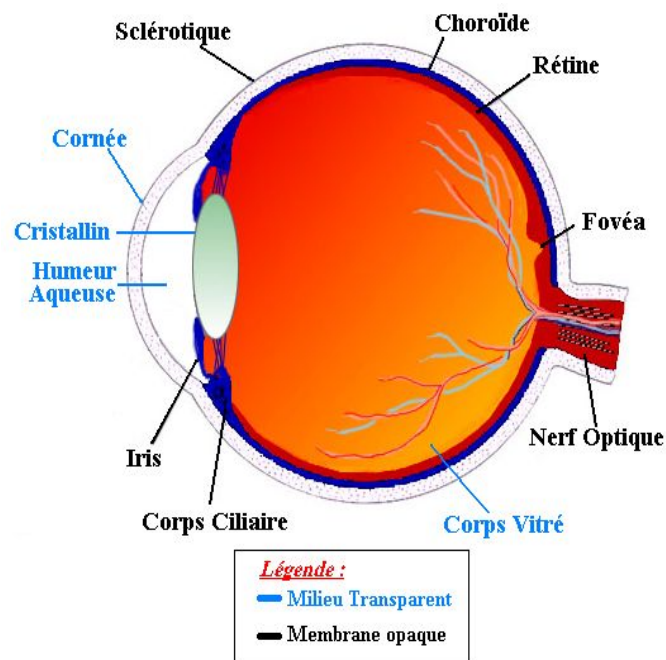


Figure 3: Scéma de l'œil

4.2 La rétine

Tapissant le fond de l'oeil, la rétine visuelle ou nerveuse est le lieu de traduction du message lumineux venant de l'extérieur en signaux nerveux envoyés au cerveau. Il s'agit d'un tissu neuronal très fin, qui fait partie du Système Nerveux Central, de 0,1 à 0,5 mm d'épaisseur, organisée en dix couches de cellules. Celles-ci comprennent l'épithélium pigmentaire, la couche des photorécepteurs, la membrane limitant, la couche granuleuse et la couche plexiforme externes et internes, la couche

des ganglionnaires, la couche des fibres optiques et finalement la membrane limitant interne. On va se concentrer ici uniquement sur trois couches : la couche des cellules photosensibles comprenant la couche des photorécepteurs et la couche granuleuse externe, la couche granuleuse interne et la couche des ganglionnaires. Ces trois couches représentent les cellules nerveuses de la rétine, reliées entre elles et ayant chacune une fonction précise.

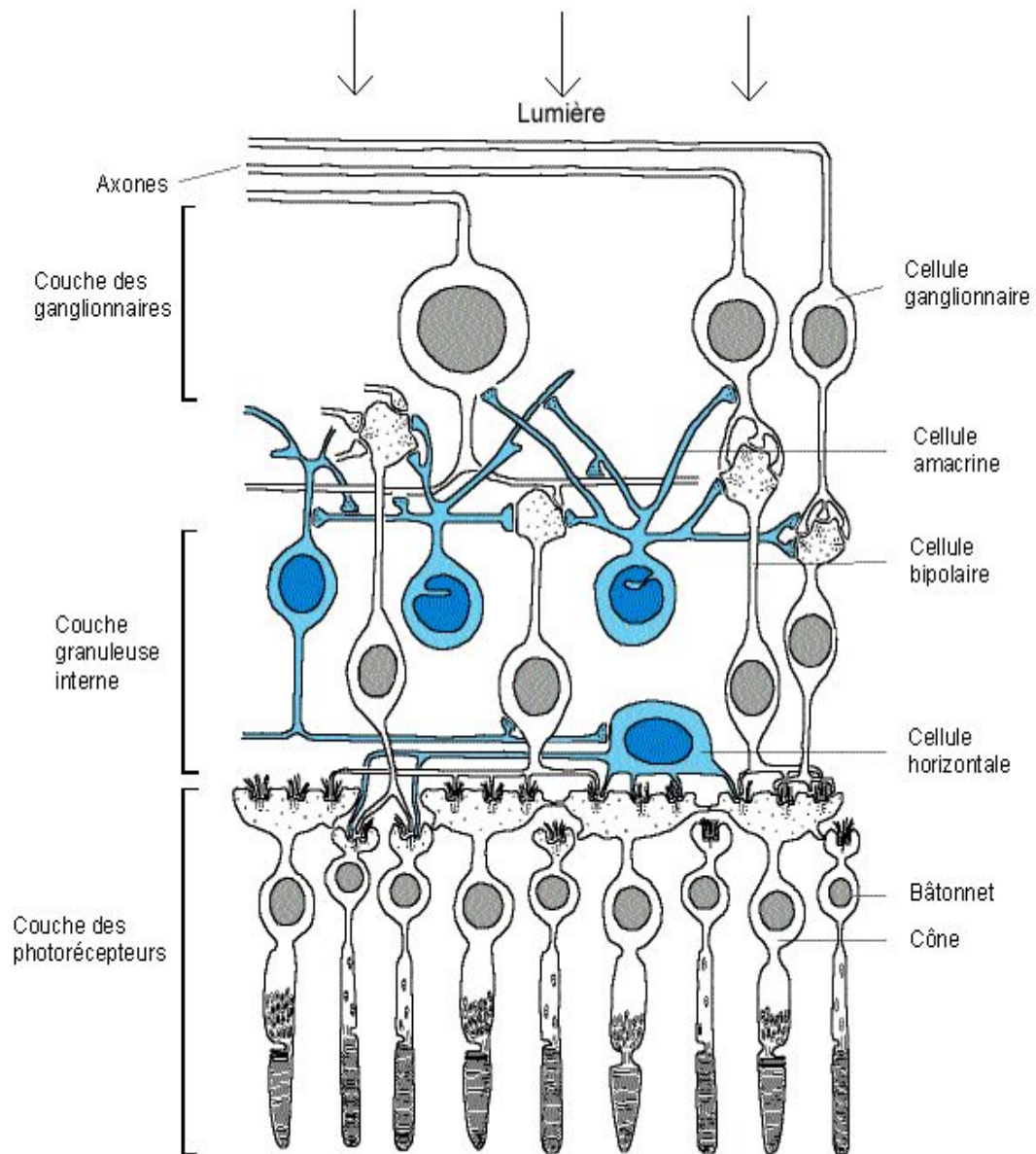


Figure 4: La rétine

La plus profonde par rapport à l'arrivée de la lumière est, paradoxalement, la couche des cellules photosensibles, aussi appelées photorécepteurs ou cellules sensorielles. La rétine est en effet "inversée", car la lumière doit traverser la rétine avant de pouvoir atteindre les photorécepteurs, sensibles à la lumière. Cette couche comporte environ 130 millions de cellules photosensibles différentes, portant des noms reflétant leur forme :

- Les bâtonnets, qui constituent environ 95% de ces cellules, soit au nombre de 120 millions, sont responsables de la vision nocturne, et ne sont sensibles qu'à la différence entre obscurité et lumière. Par contre, ils ont la plus grande sensibilité, et sont par cela adaptés à de faibles quantités de lumière.
- D'autre part, les cônes, cellules sensorielles plus grandes, forment les 5% restants des photorécepteurs. Les cônes sont responsables de la vision diurne (de jour), et font la différence entre les couleurs. Les cônes sont présents en majorité au niveau de la "tache jaune", dans la région centrale de la rétine. Au milieu de celle-ci se trouve la "fovéa", une légère dépression où sont présents uniquement des cônes, très serrés, et où les autres couches sont rejetées à la périphérie, pour laisser pénétrer la lumière plus facilement. La vision est à ce niveau plus précise, plus détaillée et plus sensible aux mouvements que sur le reste de la rétine. C'est d'ici que provient la plupart de l'information visuelle arrivant au cerveau.

La couche suivante, dénommée couche granuleuse interne, comporte une variété de neurones rétiniens, qui peuvent être classés en trois catégories :

- Les cellules bipolaires, tenant leur nom du fait qu'elles soient articulées entre les photorécepteurs et les cellules ganglionnaires de la 3^{ème} couche (les deux "pôles"), constituent la voie "directe" de transmission du message nerveux. Il existe plusieurs groupes de cellules bipolaires : les bipolaires de bâtonnets, reliant plusieurs bâtonnets à une cellule ganglionnaire, et les bipolaires de cônes, reliant un ou plusieurs cônes à une cellule ganglionnaire.

Les deux autres groupes de neurones de cette deuxième couche servent à moduler latéralement la transmission d'information, formant un "système de contrôle" qui permet au système rétinien de tenir compte des événements issus du voisinage, pour mieux s'adapter, par exemple, aux contrastes ou aux bords des objets, et aussi au mélange des couleurs.

- Les cellules horizontales, dont le nom décrit bien le placement, en contact avec les synapses (région de contact entre deux neurones, et lieu de transfert d'informations entre celles-ci) entre les récepteurs et les cellules bipolaires.

- Les cellules amacrines, en contact avec les synapses entre cellules bipolaires et ganglionnaires.

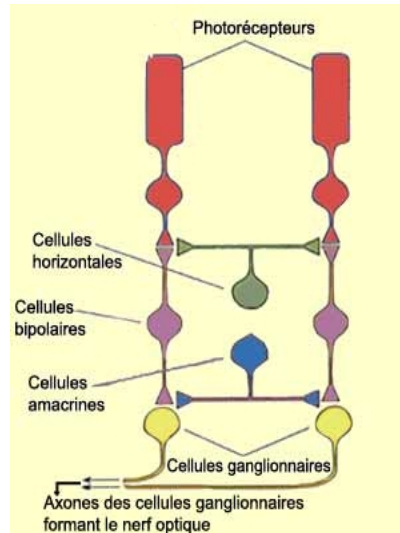


Figure 5: Schéma de la rétine

La dernière couche nerveuse de la rétine, "l'étage de sortie" vers le cerveau, est composée d'environ 1 million de cellules ganglionnaires, soit à peu près 130 fois moins nombreux que les cellules photosensibles. Ces neurones sont reliés d'une part aux cellules bipolaires, et ont d'autre part des fibres nerveuses, ou axones, qui se rejoignent au niveau de la papille pour former le nerf optique, relié au cerveau. La papille est dépourvue de photorécepteurs, et forme donc une "tache aveugle". Ceci n'entraîne pourtant pas de problème visuel, grâce à une compensation de la part du cerveau.

Grâce à cette organisation en couches, l'information est donc véhiculée vers le cerveau en plusieurs étapes, ce qui permet d'effectuer un traitement des signaux au niveau rétinien.

4.3 Les champs récepteurs

Les neurones des différentes couches de la rétine "couvrent" chacun une région de notre champ visuel. Cette région de l'espace où la présence d'un stimulus approprié modifie l'activité nerveuse d'un neurone est appelée le champ récepteur de ce neurone. Dans le système de vision humain, plus on passe d'une couche de la rétine à l'autre, et à plus forte raison si l'on se rend jusqu'aux neurones du cortex visuel, plus les champs récepteurs se complexifient. On s'intéresse ici

au comportement des cellules bipolaires et de certaines cellules du cortex visuel primaire en fonction de leur champ récepteur. Une cellule bipolaire reçoit des connexions synaptiques directes d'un certain nombre de photorécepteurs situés plus ou moins vis-à-vis elle : de un au centre de la fovéa, jusqu'à des milliers dans la rétine périphérique.

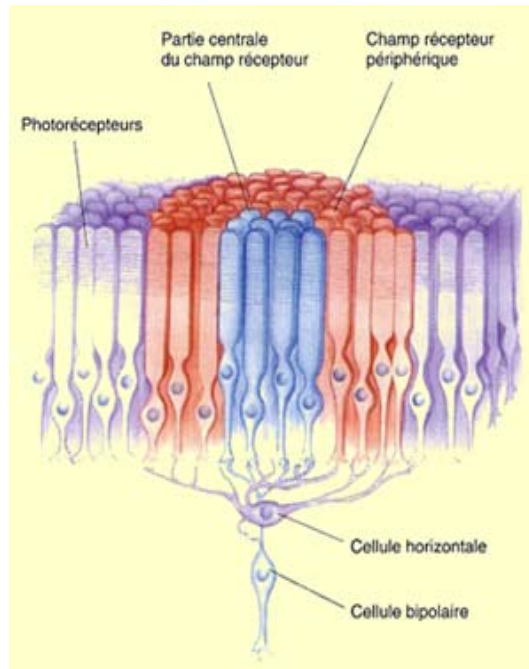


Figure 6: Champs récepteur d'une cellule bipolaire

L'éclairement du centre d'un champ récepteur d'une cellule bipolaire produit un effet (hyperpolarisation ou dépolarisation) opposé à l'éclairement de sa périphérie. Pour certaines cellules dites à centre ON par exemple, un jet de lumière va causer une dépolarisation lorsqu'il frappe le centre et une hyperpolarisation lorsqu'il frappe la périphérie.

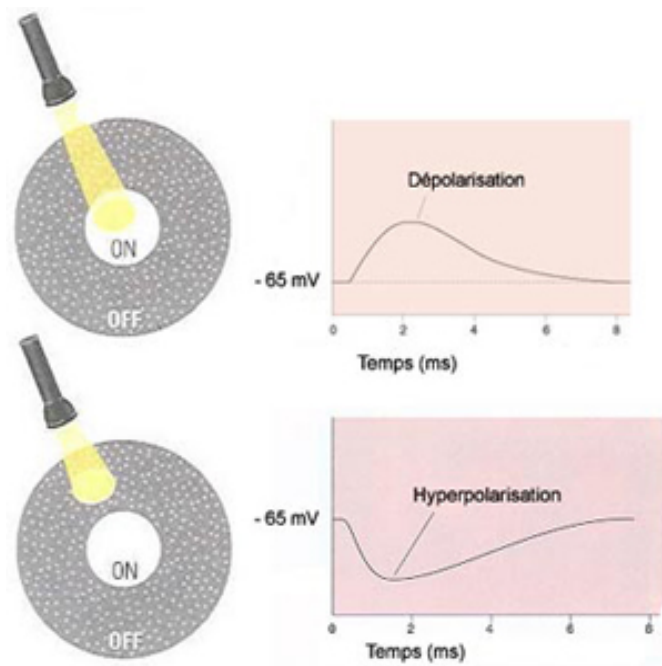


Figure 7: Réponse d'une cellule bipolaire à centre ON

Les changements au niveau du potentiel d'action des cellules (hyperpolarisation ou dépolarisation) sont très grands. Ils permettent une réponse précise et rapide du système nerveux.

Les neurones du cortex visuel primaire ont quant à eux des champs récepteurs non pas circulaires mais plutôt allongés. Ils répondent particulièrement bien à des traits de lumière ayant une orientation spécifique. Ces champs récepteurs répondant à une orientation donnée sont ceux des **cellules simples**.

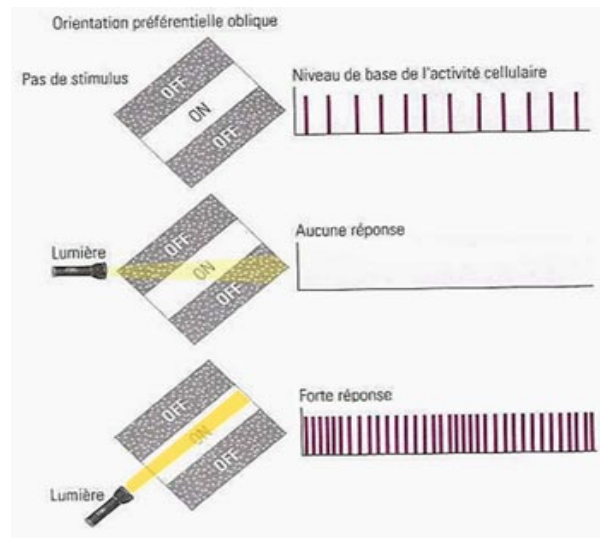


Figure 8: Réponses d'une cellule simple

Ces champs récepteurs de forme rectangulaire ont souvent une bande centrale répondant positivement à la lumière (ON) flanquée de deux bandes répondant à l'obscurité (OFF). Le dessin de la figure FIG.8 montre que lorsque le trait de lumière n'est pas exactement sur la bande ON, le stimulus n'est tout simplement pas efficace pour cette cellule.

4.4 Les chemins de traitement de l'information visuel

On définit d'abord quelques mots clés de cette partie. La vision en lumière du jour est qualifiée de vision photopique, tandis que le monde gris de la vision en faible conditions de lumière correspond à la vision scotopique. Les récepteurs de la rétine sont interconnectés de manière à former des " champs récepteurs " dont l'étendue est plus grande en vision scotopique. L'acuité de la vision des fins détails est alors sacrifiée au profit d'une sensibilité accrue. Le traitement de l'information visuel est divisé dès la rétine en deux parties distinctes : traitement de la luminance et traitement de la chrominance.

Comme indiqué précédemment, les cellules sensibles de la rétine sont de deux types :

- Les cônes : Ils agissent en vision photopique, c'est à dire lorsque la luminosité est entre moyenne et forte, et ils sont regroupés en 3 groupes mêlés dont les sensibilités spectrales complémentaires permettent l'interprétation trichrome des couleurs (Rouge, Vert, Bleu). Les cônes sont essentiellement regroupés

dans la région située dans l'axe optique de l'œil, la fovéa ; la distance moyenne les séparant y est d'environ $2,2 \mu m$.

- Les bâtonnets : Ils permettent la vision scotopique, c'est à dire lors des faibles éclaircements. La sensibilité scotopique de l'œil est maximale pour une longueur d'onde de 500 nm alors qu'en vision photopique, le maximum se situe aux alentours de 550 nm.

Il est important de noter que l'influence des cellules en bâtonnets n'est importante que pour la vision scotopique.

5 Modules inspirés

5.1 Algorithme de détection du mouvement

Afin de détecter un visage humain dans une scène vidéo, on suppose tout d'abord que le sujet est en mouvement (même de petites distances). La détection de ce mouvement nous permet de localiser les zones de travail, où on va appliquer d'autres algorithmes afin d'affiner le résultat. La méthode déjà utilisée dans l'ancienne version du projet se base sur une simple différence entre l'image courante et l'image précédente dans la séquence d'images constituant la vidéo. On prend en compte ensuite les pixels qui présentent une valeur de différence supérieure à un certain seuil. La figure FIG.9 donne un aperçu des résultats obtenus par cette méthode.

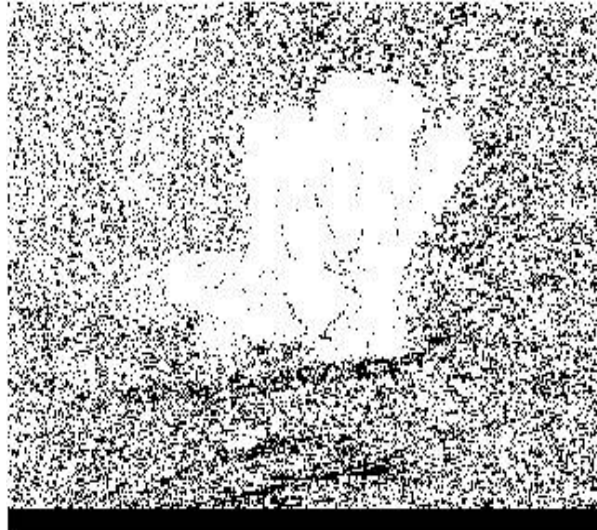


Figure 9: Application de l'algorithme de détection du mouvement

L'algorithme a été appliqué sur une main en mouvement devant un arrière plan noir. Le résultat peut être amélioré en appliquant un niveau de seuillage adéquat.

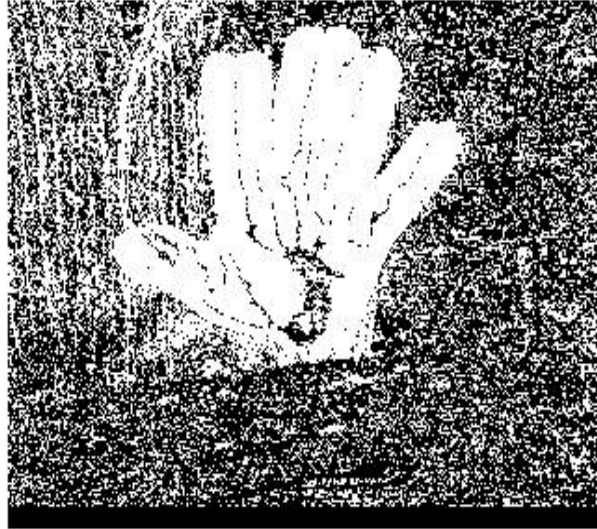


Figure 10: Application d'un niveau de seuillage adéquat

On voit clairement un dédoublement de la main. Ce problème est dû à la manière dont on présente le résultat. En effet, lorsqu'un point lumineux passe d'un pixel dans une image à un autre dans l'image suivante, on prend en compte les deux pixels. La figure FIG.11 schématise cette analyse.

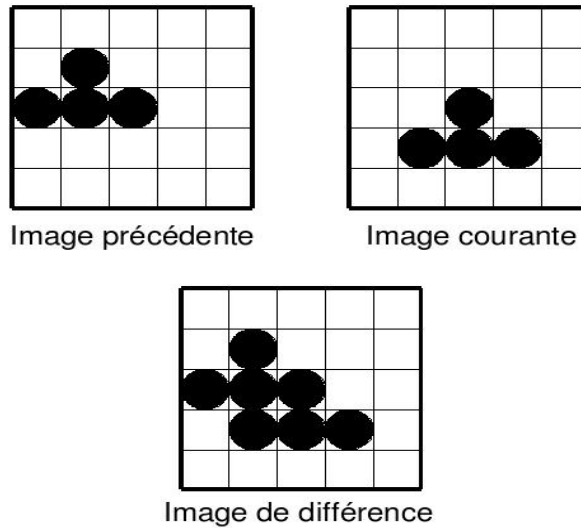


Figure 11: Principe de détection du mouvement

On voit sur l'illustration ci-dessus un dédoublement de la forme en mouvement, car l'algorithme qui analyse l'image courante prend en compte aussi les différences de luminosité observées au niveau des pixels de l'ancienne position de la forme en question. Ce problème gêne la détection, car les modules/algorithmes utilisés après la détection de mouvement s'appuient sur les contours des objets. Une première étape du travail consistait donc à améliorer le module de détection de mouvement afin d'obtenir une image plus claire. Pour réussir cette étape, on s'est inspiré du fonctionnement des champs récepteurs des cellules simples. En effet, ces cellules répondent par un changement de leur potentiel d'action à chaque fois qu'elles détectent une lumière dans la direction correspondant à leur champ récepteur. Une cellule sensible à la direction 45° par exemple ne répondra que s'il y a de la lumière dans cette direction dans la scène vue. On a donc créé plusieurs cellules d'analyse permettant de tester les différentes directions autour d'un pixel afin d'extraire l'information précise du mouvement. Pour expliquer le principe de l'algorithme, considérons un pixel (x,y) sur une image. On commence par analyser les différences de luminosité entre ce pixel et les pixels qui l'entourent dans l'image précédente, à savoir $(x-1,y-1)$, $(x-1,y)$, ..., $(x+1,y+1)$. À la détection d'une différence inférieure à un seuil fixé expérimentalement, on conclut pour la direction du mouvement, et on marque le pixel entourant comme référence de cette direction. En même temps, on signale que ce pixel avait été traité afin d'éviter un double traitement dans l'image. Afin de bien illustrer cette méthode, on va utiliser l'exemple de la figure FIG.12.

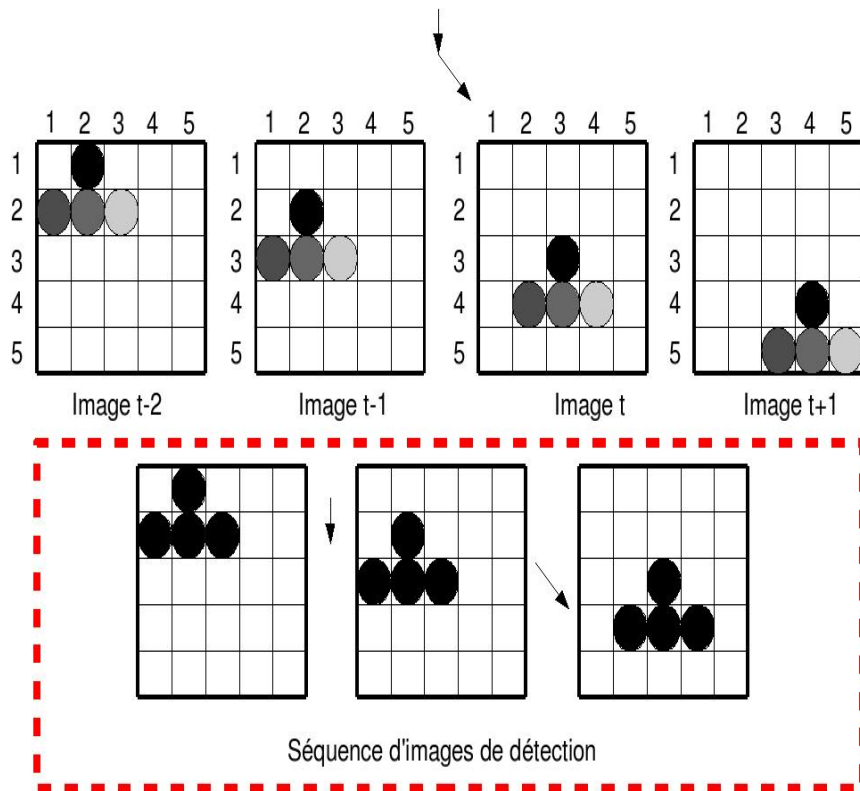


Figure 12: Schématisation de l'algorithme de détection

Prenons le pixel $(2,2)$ par exemple dans l'image $t-1$. En étudiant les différences de luminosité entre ce pixel et les pixels qui l'entourent dans l'image $t-2$, à savoir $(1,1)$, $(1,2)$, ..., $(3,3)$, on trouve que le pixel $(1,2)$ présente une différence inférieure au seuil défini, on signale un mouvement à cet endroit, et on marque les pixels testés comme traités. Ainsi dans la séquence des images de détection, on observe la suite du mouvement. L'algorithme qu'on a implémenté a largement amélioré la vitesse de détection (on travaille quasiment sur la moitié du nombre de pixels seulement). En plus on a obtenu une meilleure qualité de détection. L'illustration 13 présente un résultat obtenu avec le même test que l'illustration 10. On voit que les contours sont beaucoup plus nets, ce qui permet une suite de traitement plus efficace.



Figure 13: Détection du mouvement avec l'algorithme amélioré

5.2 Filtres d'extraction des caractéristiques du visage humain

Après la détection du mouvement, l'étape suivante consiste à appliquer certains filtres sur la zone détectée afin d'en extraire un visage humain (s'il existe bien sûr). L'application de ces filtres se fait en cascade, dans un processus similaire à celui décrit dans [8]. Le développement des filtres est basé sur une caractéristique des cellules ON/OFF évoquées précédemment. Ces cellules répondent en modifiant leur fréquence de potentiel d'action. On s'intéresse à la manière dont elles agissent et surtout la précision de leurs réponses. La figure FIG.14 montre des exemples de réponses.

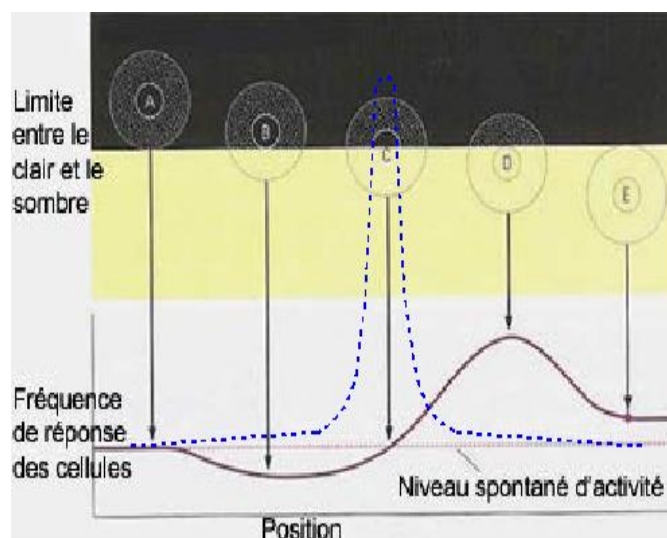


Figure 14: Réponses d'une cellule de type ON/OFF

La courbe bleu schématise le changement de fréquence de réponse de la cellule en conditions de luminosité optimale (spot lumineux dirigé vers le centre seulement dans le cas d'une cellule ON). Cette réponse est largement distinguée par rapport aux autres réponses de la cellule, qui sont considérées comme du bruit. On s'est inspiré du mode de fonctionnement de ces cellules pour définir le comportement des filtres développés. Ces filtres doivent donc avoir une sélectivité maximale pour des motifs et des caractéristiques bien précis. On a développé par exemple un filtre pour extraire les formes horizontales dans la scène. Ce filtre permet d'affiner la décision de détection au niveau de certaines zones présentant des caractéristiques de direction particulières. La figure FIG.15 présente un exemple d'application de ce filtre.



Figure 15: Extraction des formes horizontales

On voit que le filtre met en évidence les zones qui contiennent des motifs horizontaux, cela permet par exemple la vérification des résultats obtenus à partir d'autres filtres (processus en cascade). La sélectivité du filtre est contrôlée par un seuil qu'on peut déterminer expérimentalement. Une grande sensibilité risque d'inclure dans les résultats d'extraction des motifs non convenable, alors qu'une faible sensibilité peut causer la perte de données de détection.

5.3 Les chemins de traitement

On a vu que la rétine contient deux types de photorécepteurs : les cônes et les bâtonnets. Les cônes fonctionnent en lumière diurne et assurent la vision colorée alors que les bâtonnets fonctionnent en éclairage faible et ne permettent qu'une vision en nuances de gris. La principale idée du projet étant de s'inspirer du fonctionnement du cerveau, on a consacré une bonne partie du temps de développement à mettre en œuvre une séparation des chemins de traitement de la luminance et des couleurs. La méthode utilisée consiste à exploiter les possibilités offertes par les espaces de couleurs pour s'adapter au niveau du traitement. **Espaces de couleurs** :

Lors du traitement d'images ou de vidéos (suite d'images) numériquement, deux moyens sont offerts, RVB et YUV. RVB traite les données de manière intuitive, en effet l'image est stockée et manipulée comme un ensemble de valeurs pour chaque couleur fondamentale (rouge, vert, bleu) et pour chaque pixel. L'espace colorimétrique YUV est différent. Il est établi sur un constat concernant la manière

dont l'œil traite l'information. En effet, l'œil humain perçoit mieux les changements de luminosité que les changements de couleur. La figure FIG.16 montre ce constat.

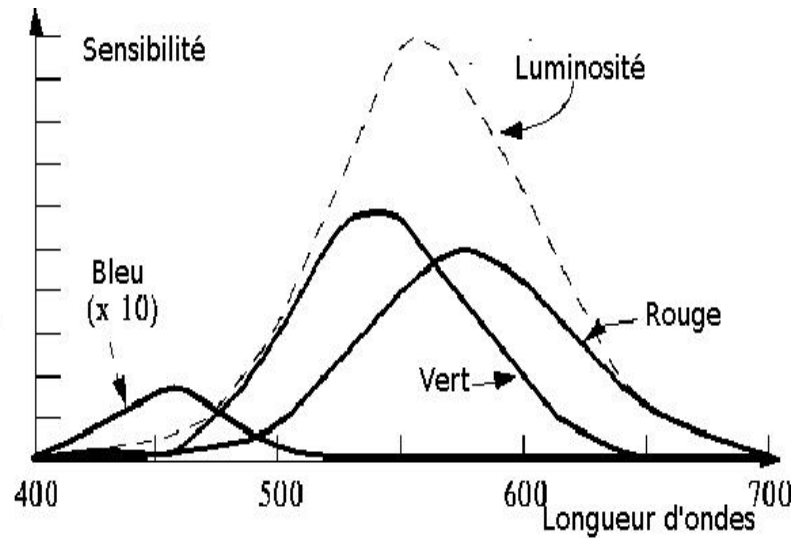


Figure 16: Sensibilité de l'œil humain

Le format YUV donne plus d'importance à la luminance et lui consacre une composante (la composante Y). Il permet ainsi une séparation entre la luminance et la chrominance (les composantes U et V). Pour le projet, on a choisi d'imiter le fonctionnement de l'œil humain. Les opérations d'acquisition et de traitement des données sont donc réalisées dans l'espace des couleurs YUV. Durant les traitements, on bascule vers le format RVB, plus intuitif, pour effectuer l'affichage des résultats. Cette organisation du projet est présentée sur le schéma de la figure FIG.17.

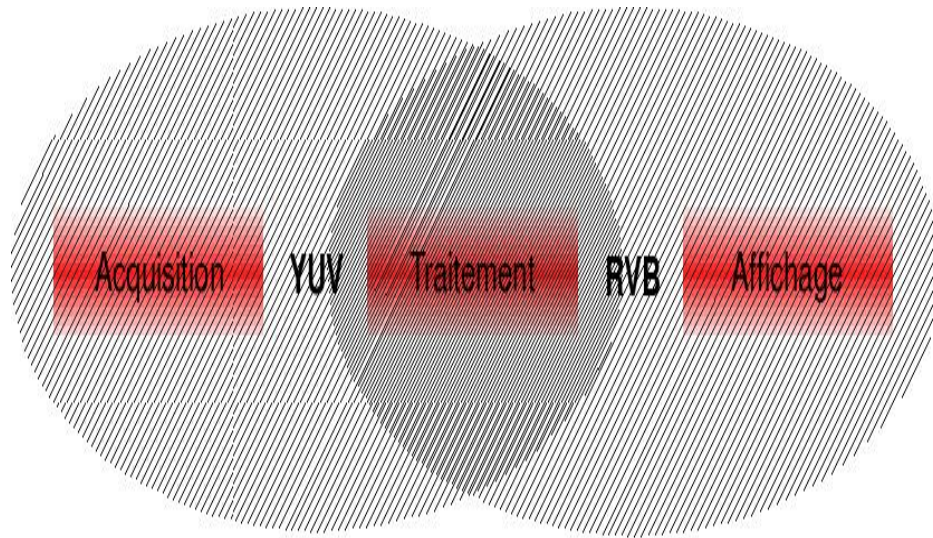


Figure 17: Organisation des espaces de couleurs dans le projet

5.4 Modules supplémentaires

Le module présenté dans cette partie s'intéresse à la détection de la région des yeux dans le visage. On cherche à réduire le temps d'exécution du programme en limitant les traitements à certaines zones bien déterminées. En utilisant une méthode similaire au processus présenté dans [4] l'article de Kung Peng (cf. bibliographie), on cherche à localiser de manière schématique les yeux dans un visage. L'idée derrière ce module consiste à offrir un moyen de vérification efficace qui permet de s'assurer que le résultat de la détection est un visage humain. En utilisant par exemple le détecteur de mouvement et le filtre de la teinte chair, on risque de confondre une main avec un visage, le module présenté dans ce chapitre permet de sélectionner une zone contenant, a priori des yeux ; ce qui offre la possibilité de vérifier leur existence. Afin d'effectuer cette tâche, on s'appuie sur certaines observations expérimentales. Dans un visage humain, les zones des yeux et du nez présentent beaucoup de changements d'intensité lumineuse. On va donc appliquer un gradient sur une zone supposée contenir un visage et on somme ensuite, horizontalement et verticalement, les valeurs obtenues. En cherchant les valeurs maximales on peut trouver la zone des yeux. Les figures FIG.18 et FIG.19 présentent les résultats d'exécution du module développé.

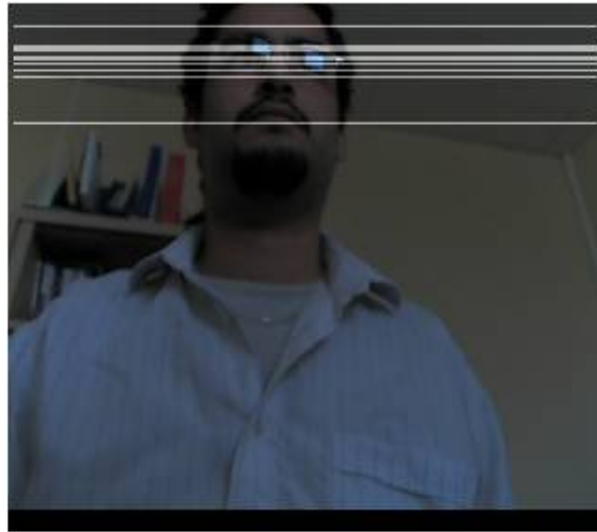


Figure 18: Détection horizontale des yeux

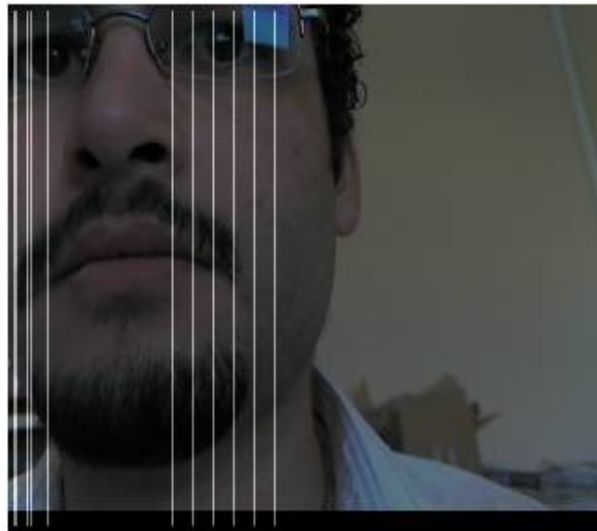


Figure 19: Détection verticale des yeux

On observe clairement une concentration des valeurs maximales dans la zone des yeux. Pour faire les expériences, on a essayé de fournir les conditions nécessaires pour avoir les résultats sans influence de l'environnement extérieur. Les zones de traitement ont été définies de tel sorte à exclure l'environnement entourant le visage du calcul du gradient. Le processus final repose sur la compilation des

résultats horizontalement et verticalement afin d'obtenir les intersections indiquant la position des yeux. La figure FIG.20 donne une idée du déroulement du processus.

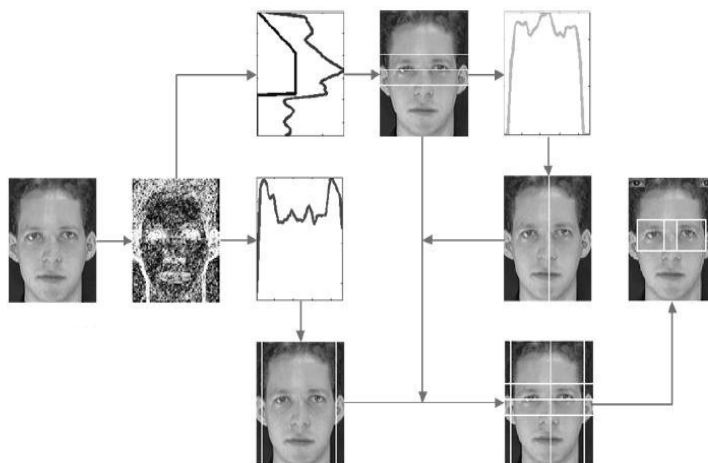


Figure 20: Processus de détection de la zone des yeux

6 Étapes suivantes

Durant cette première partie du stage, nous nous sommes d'abord familiarisé avec le projet existant, en analysant les codes sources des différents fichiers du programme. Ensuite nous avons étudié le fonctionnement du cerveau humain, et précisément le système de vision, ce qui a permis la transition vers l'étape suivante où nous avons développé et amélioré certains filtres et modules, utilisés en cascade pour effectuer la détection de visages humains dans une scène vidéo.

L'étape suivante dans le travail du stage vise la création d'un module intelligent capable de choisir le type de filtres à appliquer afin d'obtenir les meilleurs résultats et réduire le temps d'exécution. En effet, il peut y avoir des situations où certaines modules ou filtres sont absolument inutiles. Dans une scène présentant un être humain devant une autoroute par exemple, il est inutile d'appliquer le module de détection de mouvement (si l'autoroute est pleine de véhicules en mouvement bien sur). Donc il faut essayer d'analyser certaines caractéristiques de l'environnement de la scène avant de commencer la détection. nous n'avons pas encore décidé des méthodes qu'on pourrait utiliser dans la réalisation de ce module, mais nous allons probablement s'appuyer sur des théories et des algorithmes d'apprentissage artificiel. On doit néanmoins faire attention à bien choisir les algorithmes, car ces méthodes sont en général gourmandes en ressources systèmes et consomment beaucoup de temps de calcul.

7 Conclusion

Le sujet conjugue les principes de l'informatique et du développement avec les neurosciences, ainsi une grande partie du planning du stage était consacrée à l'étude bibliographique, ce qui m'a permis d'acquérir des connaissances suffisantes pour pouvoir développer dans ce cadre de bio-inspiration. Ce domaine en plein développement ouvre des perspectives intéressantes, surtout quand il est appliqué à la robotique, discipline qui m'attire de plus en plus, Ce stage m'a donc permis d'approfondir largement mes connaissances de programmation, d'algorithmique, de raisonnement et de méthodes de recherche.

Références

- [1] BERTHOZ, A. , SCIENCES, O. J., (eds.) LA DECISION, IMPRIMERIE FLOCH A MAYENNE, 2003
- [2] BERTHOZ, A. , SCIENCES, O. J., (eds.) LE SENS DU MOUVEMENT, IMPRIMERIE FLOCH A MAYENNE, 1999
- [3] HYUNDO KIM, ERIK MURPHY-CHUTORIAN, J. T. "SEMI-AUTONOMOUS LEARNING OF ONJECTS," WORKSHOP : VISION FOR HUMAN COMPUTER INTERACTION (), 2006, pp. 6
- [4] Ian Fasel, Bret forttenberry, J. M. "A generative framework for real-time object detection and classification," Computer Vision and Image Understanding (), 2005, pp. 30
- [5] IMBERT, M. , SCIENCES, O. J., (eds.) TRAITE DU CERVEAU, IMPRIMERIE FLOCH A MAYENNE, 2006
- [6] Kun Peng, Liming Chen, S. R. G. K. "A Robust and Efficient Algorithm for Eye Detection on Gray Intensity Face ," (), 2005, pp. 8
- [7] MARTIN A.FISCHLER, O. F. , of congress Cataloging-in-Publication Data, L., (eds.) Intelligence. The Eye, the Brain, and the Computer, Addison-Weskey Publishing Company ; Inc, 1987
- [8] Paul Viola, M. J. "Robust Real-time Object Detection," SECOND INTERNATIONAL WORKSHOP ON STATISTICAL AND COMPUTATIONAL THEORIES OF VISION-MODELING, LEARNING, COMPUTING, AND SAMPLING (), 2001, pp. 25
- [9] Paul Viola, M. J. "Fast and robust Classification using Asymmetric AdaBoost and a Detector Cascade," Neural Information Processing Systems (), 2001, pp. 8
- [10] PERRINET, L. "Comment déchiffrer le code impulsif de la Vision? Etude du flux parallèle, asynchrone et éparé dans le traitement visuel ultra-rapide", 2003.
- [11] LE CERVEAU Á TOUS LE NIVEAUX (<http://lecerveau.mcgill.ca>)
- [12] La perception des couleurs par l'œil (<http://www.bioinformatics.org>)
- [13] Organisation fonctionnelle du cortex visuel primaire (<http://acces.inrp.fr/>)
- [14] Encyclopédie électronique libre (<http://fr.wikipedia.org>)
- [15] Perception visuelle humaine (<http://www.rennes.supelec.fr/ren/perso/jweiss>)