

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Saad Dahlab - Blida 1 (USDB)



Faculté des Sciences
Département d'Informatique

MEMOIRE DE FIN D'ETUDES

Pour l'obtention du Diplôme de Master en Informatique

Option : systèmes d'informatiques et réseaux

THEME

Développement d'une Application de Détection Somnolence au Volant du Conducteur Pour Eviter les Accidents de la Route

ORGANISME D'ACCUEIL :

Centre de Développement des Technologie Avancées (CDTA)



Réalisé par :

Mr. KADDOUR Khaled et Mr. LAZEREGUE Abdelwadoud.

Promotrice : Mme. YKHLEF Hadjer (USDB)

Encadreur : Mr. HIMEUR Yassine (CDTA)

Co-Encadreur : Mr. AMOURI Billel (CDTA)

Présidente : Mme. TOUBALINE Nesrine (USDB)

Examinatrice : Mme. DJEDDAR Afrah (USDB)

2018/2019

Remerciements

En préambule à ce mémoire, nous remercions ALLAH qui nous a aidé et donné la patience et le courage durant cette longue année d'étude.

Nous souhaitons adresser nos remerciements les plus sincères aux personnes qui nous ont apporté leur aide et qui ont contribué à l'élaboration de ce mémoire ainsi qu'à la réussite de cette formidable année universitaire.

Ces remerciements vont tout d'abord à notre promotrice Mme YEKHLEF HADJER pour sa disponibilité tout en long de la réalisation de ce Mémoire, Ainsi pour son inspiration, aide et son suivi.

Nos remerciements vont aussi au corps professoral et administratif de département d'informatique de l'université SAAD DAHLEB de BLIDA pour la richesse et la qualité de ses enseignements et qui déploie des grands efforts pour assurer à leurs étudiants une formation de qualité.

Nous remercions très chaleureusement aussi, Mr HIMMEUR YASSINE, Notre encadreur, pour sa confiance et ses encouragements.

Nous remercions aussi notre encadreur Mr AMOURI BILLEL, pour le temps qu'il nous a consacrés et leurs précieux conseils.

Nos remerciements vont également à tout le personnel de Centre de Développement des Technologie Avancées "CDTA" pour son hospitalité et un particulièrement à celui qui nous a guidés tout au long du stage.

Nos remerciements iront également vers tous ceux qui ont accepté avec bienveillance de participer au jury de ce mémoire.

On n'oublie pas nos parents pour leur contribution, leur soutien et leur patience.

Enfin, nous adressons nos plus sincères remerciements à tous nos proches et amis, qui nous ont toujours encouragées au cours de la réalisation de ce mémoire.

Merci à tous et à toutes.

Dédicaces

Je dédie ce modeste travail

À mes parents qui depuis mon plus jeune âge ont toujours fait leur maximum, en consacrant temps et argent, pour m'éveiller et m'encourager dans mes passions. C'est grâce à vous et pour vous que j'ai fait mon mémoire. Aucun mot sur cette page ne saurait exprimer ce que je vous dois, ni combien je vous aime. Qu'Allah vous bénisse, vous assiste, vous vienne en aide ;

A mon binôme ABDELWADOUD ;

A mes chers frères, sœurs et mes neveux ;

A ma promotrice Mme YAKHLEF HADJER ;

A mon encadreur Mr HIMEUR YASSINE ;

A mon co-encadreur Mr AMOURI BILLEL ;

A tous mes collègues, plus particulièrement : ABDELHAKIM, ZAKARIA, AYMEN, et toute la section Master2 SIR en témoignage de mon amitié sincère ;

A tous mes amis, plus particulièrement : MOHAMED YASSINE, AHMED, AYOUB en témoignage de mon amitié sincère;

A tous ceux qui m'ont soutenu, qu'ils trouvent ici l'expression de mon amour et ma profonde.

Gratitude

KHALED

Dédicaces

Je dédie ce modeste travail

A mes parents, pour m'avoir encouragé, Sans eux, je n'en serais pas là.

J'exprime toute ma reconnaissance et gratitude à l'administration et à l'ensemble du corps enseignant de l'Université de (SAAD DAHLEB) pour leurs efforts à nous garantir la continuité et l'aboutissement de ce programme de Master.

Je tiens à remercier aussi et chaleureusement mes encadreurs Mme (YEKHLEF HADJER), Mr (HIMMEUR YASSINE) et Mr (AMOURI BILLEL) de m'avoir permis de mener ce travail, pour leur engagement et leur soutien ainsi que pour la pertinence de leurs remarques et de leur feed-back.

Mes vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont porté à notre recherche en acceptant d'examiner notre travail et de l'enrichir par leurs propositions.

Enfin, Je tiens également à remercier toutes les personnes qui ont participé de près ou de loin à la réalisation de ce travail, à mes amis, ma famille.

Merci.

ABDELWADOUD

ملخص

يموت العديد من الأشخاص كل يوم جراء حوادث المرور الناتجة عن نعاس و تعب السائق. و لتفادي هذا النوع من الحوادث، يجب الكشف المبكر عن النعاس الذي يعتبر مشكلا معقدا نظرا للتغير غير المتوقع للحالة الفيزيولوجية و السلوكية للسائق. في السنوات الأخيرة ، و بسبب التطور التكنولوجي الذي شهدته رؤية الكمبيوتر، أصبح من الممكن مراقبة حالة السائق بدون التدخل في مهمته عن طريق تحليل تعابير الوجه و حالة العينين.

يتم مراقبة حالة السائق اعتمادا على حالة العينين عن طريق ثلاث خطوات أساسية تتمثل في كشف الوجه ثم تحديد موقع العينين و أخيرا التعرف على حالة العينين باستعمال تقنيات التقاط الصور و معالجتها.

تطرقنا في مشروعنا الى الطرق الأكثر تداولاً في مجال الكشف عن الوجه و التي تتمثل في HOG ، Viola & Jones و CNN . و بعد اختبارها على قاعدة البيانات FDDDB و YAWDD خلصنا الى أن طريقة Viola et Jones تعطي نتائج حسنة و بتدعيمها بطريقة التتبع Suivi de Corrélation أصبح الكشف أحسن بكثير.

أما فيما يخص تحديد موقع العين ، فقد قمنا بتدريب نماذجنا الخاصة اعتمادا على طريقتي HOG و Viola et Jones على قاعدة البيانات التي قمنا بإنشائها فكانت نماذج Viola et Jones أحسن من نماذج HOG .

و في الأخير قمنا بحساب نسبة الأعين المغلقة في وحدة زمنية معينة و ذلك بعد تحديد حالة العينين للكشف عن نعاس السائق بقصد تنبيهه.

الكلمات المفتاحية : النعاس أثناء السياقة ، الكشف عن النعاس ، نظام غير تدخلي ، مساعدة السائق ، تتبع الوجه ، حالة العينين.

Résumé

De nombreuses personnes meurent chaque jour suites aux accidents de la route causés par la somnolence et la fatigue du conducteur. Pour éviter ce type d'accident, il convient de détecter rapidement la somnolence, qui dû au changement imprévu de l'état physiologique et comportemental du conducteur. En raison du développement technologique de la vision par ordinateur, il est possible de surveiller l'état du conducteur sans interférer avec sa tâche en analysant les expressions faciales et l'état des yeux.

L'état du conducteur est contrôlée en fonction de l'état des yeux à travers trois étapes de base: la détection du visage, le positionnement des yeux et enfin la reconnaissance de l'état des yeux à l'aide de techniques de traitement des images.

Dans notre projet, nous avons discuté les plus courantes méthodes de détection des visages : HOG, Viola et Jones et CNN. Après des tests sur les bases de données FDDB et YAWDD, nous avons constaté que la méthode de Viola et Jones donnait de bons résultats et qu'en prenant en charge la méthode de Suivi de corrélation, la détection était bien meilleure.

En ce qui concerne l'emplacement des yeux, nous avons formé nos propres modèles basés sur les méthodes HOG et Viola et Jones sur la base de données que nous avons créé. Les modèles de Viola et Jones sont meilleurs que les modèles HOG.

Enfin, nous avons calculé le pourcentage des yeux fermés dans une unité de temps donnée après avoir déterminé l'état des yeux afin de détecter la somnolence du conducteur dans l'intention de le prévenir.

Mots clés: somnolence au volant, détection de la somnolence, système non intrusif, aide à la conduite, suivie du visage, troubles de la vue.

Abstract

Many people die every day from road accidents caused by drowsiness and driver fatigue. To avoid this type of accident, drowsiness must be detected quickly, a complex problem due to the unforeseen change in the physiological and behavioral state of the driver. In recent years, due to the technological development of computer vision, it is possible to monitor the state of the driver without interfering with his task by analyzing facial expressions and eye states.

The state of the driver is controlled according to the state of the eyes through three basic steps: the detection of the face, the positioning of the eyes and finally the recognition of the state of the eyes with the help of techniques of capture and image processing.

In our project, we discussed the most common methods of face detection, namely HOG, Viola and Jones and CNN. After testing on the FDDB and YAWDD databases, we found that the Viola and Jones methods performed well and that by supporting the Correlation Tracking method, the detection was much better.

Regarding the location of the eyes, we formed our own models based on the HOG and Viola and Jones methods on the database we created. The Viola and Jones models are better than the HOG models.

Finally, we calculated the percentage of eyes closed in a given unit of time after determining the state of the eyes to detect sleepiness of the driver with the intention of preventing it.

Key words: driving drowsiness, drowsiness detection, non-intrusive system, driving assistance, facial follow-up, visual disturbances.

Table des matières

INTRODUCTION GENERALE	1
CHAPITRE 1 : NOTIONS DE BASE SUR LE TRAITEMENT D'IMAGE ET DE VIDEO	4
1.1. INTRODUCTION	5
1.2. DEFINITION DE L'IMAGE NUMERIQUE	5
1.3. CARACTERISTIQUES D'UNE IMAGE NUMERIQUE	6
1.3.1. <i>Pixel</i>	6
1.3.2. <i>Résolution d'une image</i>	6
1.3.3. <i>Dimension d'une image</i>	6
1.3.4. <i>Espace de couleur</i>	6
1.4. VIDEO	8
1.5. ACQUISITION DES DONNEES	8
1.6. CONCLUSION	9
CHAPITRE 2 : APPROCHES DE DETECTION DE SOMNOLENCE.....	10
2.1. INTRODUCTION	11
2.2. RELATION ENTRE LA SOMNOLENCE ET LA CONDUITE	13
2.3. APPROCHES DE DETECTION DE SOMNOLENCE.....	13
2.3.1. <i>Approches axées sur la performance du conducteur</i>	13
2.3.2. <i>Approches axées sur les signaux physiologiques</i>	14
2.3.3. <i>Approches axées sur le traitement d'image</i>	15
2.4. CONCLUSION	18
CHAPITRE 3 : DETECTION ET SUIVI DU VISAGE.....	20
3.1. INTRODUCTION	21
3.1.1. <i>Détection d'objet</i>	21
3.1.2. <i>Region of interest</i>	22
3.2. ALGORITHMES DE DETECTION DU VISAGE	22
3.2.1. <i>Viola & Jones</i>	23
3.2.1.1. <i>Caractéristiques</i>	23
3.2.1.2. <i>Image Intégral</i>	24
3.2.1.3. <i>Cascade attentionnelle</i>	25
3.2.2. <i>Histogram of oriented gradients</i>	26
3.2.2.1. <i>Calcul de vecteur des caractéristiques Histogrammes</i>	26
3.2.2.2. <i>Classification</i>	27
3.2.3. <i>Réseaux de neurones convolutionnels</i>	27
3.2.4. <i>Evaluation</i>	28
3.2.4.1. <i>Intersection Over Union</i>	30
3.2.4.2. <i>Matrice de confusion</i>	31
3.2.5. <i>Résultats et discussion</i>	32
3.3. SUIVI DU VISAGE	34
3.3.1. <i>Cam-shift</i>	34
3.3.2. <i>Suivi de Corrélation</i>	35
3.3.3. <i>Résultats et discussion</i>	36
3.4. CONCLUSION	37
CHAPITRE 4 : DETECTION DES YEUX.....	38
4.1. INTRODUCTION	39
4.2. COLLECTION DE DONNEES	39
4.3. ECHANTILLONS POSITIFS	39

4.4.	ECHANTILLONS NEGATIFS	40
4.5.	ENTRAINEMENT	40
4.6.	RESULTATS ET DISCUSSION	41
4.6.1.	<i>Viola & Jones</i>	41
4.6.2.	<i>Histogram of oriented gradients</i>	45
4.7.	CONCLUSION	48
CHAPITRE 5 : DETECTION DE LA SOMNOLENCE.....		49
5.1.	INTRODUCTION	50
5.2.	DETECTION ET SUIVI DU VISAGE	51
5.3.	DETECTION DES YEUX	51
5.4.	ETAT DES YEUX	51
	52	
5.5.	POURCENTAGE DES YEUX FERMES	53
5.6.	EVALUATION	53
5.7.	RESULTATS ET DISCUSSION	54
5.8.	CONCLUSION	55
CONCLUSION GENERALE.....		56
BIBLIOGRAPHIE.....		59

Liste des figures

Figure 0.1: Raspberry PI 3 modèle B+	2
Figure 1.1 : Les étapes principales de traitement d'image.	5
Figure 1.2 : Une image sous 3 résolutions différentes [3].	6
Figure 1.3 : Palette de niveaux de gris (grayscale).	7
Figure 1.4 : Palette de RGB [3].	7
Figure 1.5 : palette de HSV.....	8
Figure 2.1 : le nombre des tués et des blessés entre 1970 et 2015 en Algérie où A est le nombre de tués et B le nombre de blessé [8].	11
Figure 2.2 : Les différentes méthodes de détection de somnolence.	13
Figure 2.3: illustration des capteurs utilisés dans les méthodes axées sur la performance du conducteur [15].	15
Figure 2.4 : Architecteur utilisé dans l'approche de Bhargava Reddy et al [29].	17
Figure 3.1: Un exemple de détection de la position d'objets et leurs catégories [3].....	21
Figure 3.2: exemple sur les ROI de visage et les deux yeux [32].	22
Figure 3.3 : Exemple de sous image de 5 x 5 dans le réel et sur un ordinateur [3].....	23
Figure 3.4 : Illustration des caractéristiques de Haar :.....	24
(A, b) deux rectangles, (c) trois rectangles, (d) quatre rectangles [33].....	24
Figure 3.5 : Illustration de l'image intégrale où la surface $S = C + A - (B+D)$	25
Figure 3.6 : Illustration du classificateur en cascade.	25
Figure 3.7 : les filtres qui utilisent dans le calcul de gradient verticale et horizontale.	26
Figure 3.8 : Des Photos prises de l'ensemble des données YawDD [32].....	29
Figure 3.9: Des Photos prises de l'ensemble des données Fddb [39].	29
Figure 3.10: Illustration de l'intersection et l'union.	30
Figure 3.11 : Les résultats de nos expériences sur les trois algorithmes ; La première ligne représente les résultats de Viola Jones, le deuxième HOG, le dernier CNN où Le cadre bleu représente le class réel et l'autre cadre représente les prédictions des algorithmes.....	33
Figure 3.12: illustration de suivi d'un visage.	35
Figure 4.1 : Des exemples de résultats dont le premier cas où les cadres bleus représentent les class réel et les cadres verts représente les class prédit par le model numéro 5.	42
Figure 4.2: Des exemples de résultats dans le deuxième cas où les cadres bleus représentent les class réel et les cadres verts représente les class prédit par le model numéro 5.	43
Figure 4.3 : Des exemples des résultats dans le troisième cas où les cadres bleus représentent les class réel et les cadres verts représente les class prédire par le model numéro 5.	45
Figure 4.4 : les résultats de test des modèles HOG dans la première condition.	46
Figure 4.5: les résultats de test des modèles HOG dans la deuxième condition.	46
Figure 4.6 : les résultats de test des modèles HOG dans la deuxième condition	47
Figure 4.7 : Des exemples des résultats dans le troisième cas où les cadres bleus représentent les class réel et les cadres verts représente les class prédire par le model qui a $c=1000$, $\sigma=0.05$	47
Figure 5.1 : L'organigramme de système proposé.....	50
Figure 5.2 : les résultats de test des modèles de détection des yeux sur le 25% de l'ensemble de donné de chapitre 3.....	52
Figure 5.3 : Les résultats de test du modèle avec $\sigma=0.04$ et $C=1000$	52

Liste des tableaux

Tableau 2.1: Les accidents routiers et les accidents routiers causés par la somnolence du conducteur entre 2011 et 2015 [9].	12
Tableau 2.2 : comparaison entre les méthodes de détection de somnolence.	18
Tableau 2.3 : comparaison entre les méthodes de détection de somnolence basées sur le traitement d'image.	19
Tableau 3.1 : Matrice de confusion d'un modèle de classification binaire.	31
Tableau 3.2: Les résultats de test sur l'ensemble des données FDDB.	32
Tableau 3.3: Les résultats de test sur l'ensemble des données YawDD.	32
Tableau 3.4 : les résultats de test sur l'ensemble des données YawDD	37
Tableau 4.1: les résultats de test dans la première condition.	41
Tableau 4.2: les résultats de test dans la deuxième condition	43
Tableau 4.3 : les résultats de test dans la troisième condition.	44
Tableau 5.1 : Les résultats de test de notre méthode sur l'ensemble des données NTHU.	54
Tableau 5.2 : les résultats de test de la méthode de Adrian sur l'ensemble de données NTHU.	55

Liste des abréviations

- ABS:** Anti-lock Braking System.
- AdaBoost:** Adaptive boosting.
- API:** Application Programming Interface.
- ASIRT:** Association Record Safety international Road Travel.
- CEW:** Closed Eyes in the Wild.
- CFT:** Correlation Filter Tracking.
- CNN:** Convolutional Neural network.
- CNPSR:** Centre National de Prévention et de Sécurité Routières.
- ECG:** Electrocardiography.
- EEG:** Electroencephalographic.
- EOG:** Electro-oculographie.
- Fast-R-CNN:** Fast Region-based Convolutional Network.
- Fddb:** Face Detection Data Set and Benchmark.
- FN:** False Negative.
- FP:** False Positive.
- FPS:** Frame Per Second.
- HOG:** Histogram of oriented gradients.
- HSV:** Hue, Saturation, Value.
- IOU:** Intersection Over Union.
- KNN:** K-Nearest Neighbors.
- KSS:** Karolinska Sleepiness Scale.
- LSTM:** Long short-term memory.
- MLP:** Multi-Layer Perceptron.
- MSO:** Multi Salient Objet.
- MTCNN:** Multi-Task Cascaded Convolutional Neural Networks.
- NHTSA:** National Highway Traffic Safety Administration.
- NTHU:** National Tsuing Hua University.
- PDO:** Property Damage Only.

PNN: Prototype Nearest Neighbor.

PPP: Point Par Pouce.

R-CNN: Region-based Convolutional Network.

RGB: Rouge, Green, Blue.

RNN: Recurrent Neural Network.

ROI: Region Of Interest.

SDLP: Standard Deviation of Lane Position.

SSD: Single Shot Multi Box Detector.

SVM: Support Vector Machine.

TN: True Negative.

TP: True Positive.

YawDD: Yawning Detection Dataset.

CCD: Charge Coupled Device.

Introduction générale

Les systèmes de détection de la somnolence de conducteur disponibles actuellement appartiennent généralement à deux catégories: les systèmes très coûteux dédiées pour les voitures de haut de gamme; et des systèmes abordables et peu robustes. La somnolence du conducteur est l'une des principales causes d'accidents de la route dans lesquels de nombreuses personnes décèdent ou se blessent chaque année. L'objectif de notre travail est de concevoir un système de détection de la somnolence du conducteur afin de l'aider à éviter les accidents causés par la collision. Cette détection se fait en fonction d'états des yeux et en utilisant des algorithmes appropriés implémenter au sein d'un Raspberry pi 3 modèle B+.

Raspberry pi 3 modèle B+ est un ordinateur pas cher qui peut se prêter à de nombreuses tâches légères et moyennes.il a les caractéristiques suivantes :

- 1.4GHz 64-bit processeur quatre cœurs.
- 1GB SRAM.
- Réseau local sans fil bandede 2,4 GHz et 5 GHz.
- Vitesse Ethernet 300Mbps.
- Bluetooth 4.2 / BLE.



Figure 0.1: Raspberry PI 3 modèle B+

Afin d'atteindre ces objectifs, une revue intensive de la littérature a été conduite et s'est orientée principalement vers les thématiques d'intérêt dans le cadre de cette mémoire à savoir les algorithmes de détection d'états des yeux, l'algorithme de détection Viola-Jones, et HOG. Initialement, les performances des algorithmes de détection de la somnolence sont analysées et évaluées. Ensuite une validation expérimentale sur le système a été présentée.

Ce mémoire est organisé en cinq chapitres comme suit :

- Le premier chapitre explicitera les notions de base sur le traitement d'images et vidéos.
- Le deuxième chapitre mettra en évidence, les différentes approches de détection de somnolences existantes.
- Dans le troisième chapitre nous présenterons les différentes approches de détection et suivie de visage les plus courantes.
- Le quatrième chapitre traitera le processus de formations des modèles de détection des yeux.
- Le dernier chapitre, il sera consacré à l'évaluation de la méthode proposé sur la base de données proposé par laboratoire de NTHU (National Tsuing Hua University).

Chapitre 1 : Notions de base sur le traitement d'image et de vidéo

1.1. Introduction

Aujourd'hui, le traitement d'images est un domaine très vaste qui a connu un développement important depuis quelques dizaines d'années. Ce domaine a contribué au développement de plusieurs domaines tels que la médecine, l'astronomie, l'informatique...etc. Le traitement des images est une méthode permettant d'effectuer certaines opérations sur une image afin d'obtenir une image améliorée ou d'en extraire des informations utiles. On peut le définir aussi comme un type de traitement du signal dans lequel l'entrée est une image et la sortie peut être une image où des caractéristiques associée à cette image.

Le traitement d'image est basé sur trois étapes essentielles : l'importation de l'image via des outils d'acquisition d'image comme la caméra, l'analyse et la manipulation de l'image, et finalement, l'extrait des informations.

Dans ce chapitre, nous abordons les notions de base nécessaires à la compréhension des techniques de traitement d'images et Vidéos.

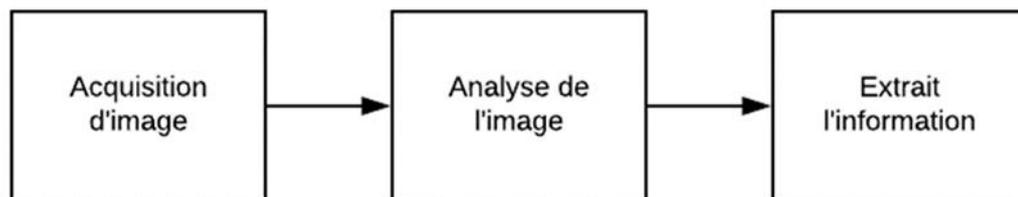


Figure 1.1 : Les étapes principales de traitement d'image.

1.2. Définition de l'image numérique

Une image numérique peut être considérée comme une représentation discrète de données possédant à la fois des informations spatiales (mise en page) et d'intensité (couleur), on peut aussi considérer comme un signal lumineux multidimensionnel [1].L'image peut aussi être représentée mathématiquement selon de fonction bidimensionnelle $I(x,y)$, De $\mathfrak{R} \times \mathfrak{R}$ dans \mathfrak{R} où le couplet d'entrée est considéré comme une position spatiale et le singleton de sortie comme l'intensité.

1.3. Caractéristiques d'une image numérique

1.3.1. Pixel

Le pixel (Picture Element) est le plus petit élément contrôlable d'une image. Il est souvent abrégé p ou px qui possède une valeur qui peut être un scalaire qui représente un niveau de gris, ou un vecteur représentant une couleur, ou toute autre chose [2].

1.3.2. Résolution d'une image

La résolution d'une image est définie par le nombre de pixels dans une longueur donnée (en pouce) dans cette image. Elle est exprimée en PPP (Point par pouce) où le pouce est 25.4 mm, ce nombre a une relation directe avec la qualité d'une image, ça signifie qu'une image avec haute qualité a un grand nombre de pixel dans un pouce.

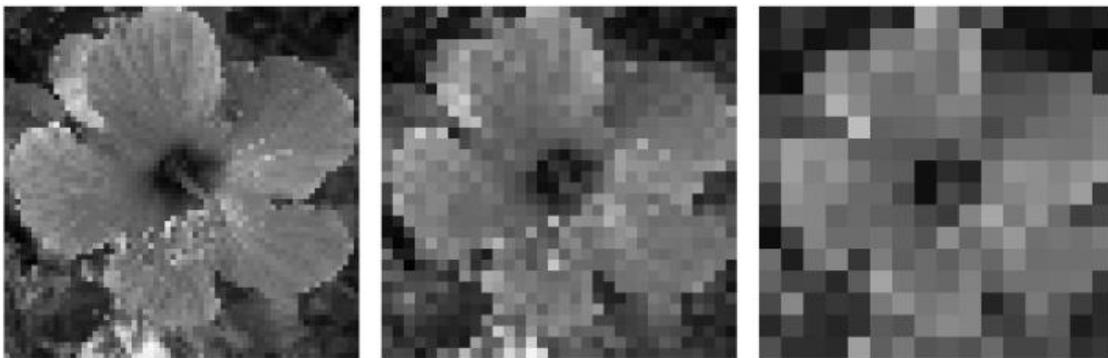


Figure 1.2 : Une image sous 3 résolutions différentes [3].

1.3.3. Dimension d'une image

Il s'agit de la taille de l'image, qui représente le nombre total de pixels représentant l'image. Ce nombre de pixels représente une matrice bidimensionnelle (C x R) où le C le nombre de colonnes et R est le nombre de lignes (ex. 640 x 480, 800 x 600, 1024 x 768, etc.) [1].

1.3.4. Espace de couleur

Chaque image numérique contient un ou plusieurs canaux de couleur, qui définit une intensité ou une certaine couleur dans une position $I(x, y)$ de l'image. Et la conversion d'une image de la représentation abstrait à la représentation réelle requiert la présence d'une palette

de couleurs qui associe chaque valeur de la représentation abstraite à une couleur de cette palette.

Les palettes de couleurs les plus connues sont les niveaux de gris (grayscale), RGB et HSV [1]:

- **Images à niveaux de gris (grayscale)** : C'est un système de codage informatique des couleurs avec un seul Canal où chaque valeur de cette canal varie de 0 (noir) à 255(blanc).

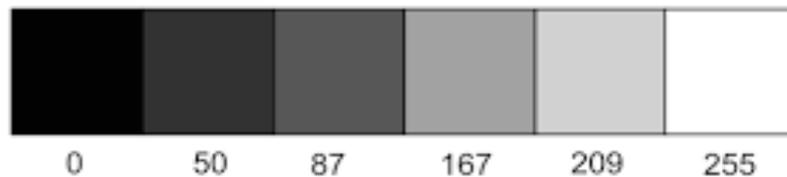


Figure 1.3 : Palette de niveaux de gris (grayscale).

- **RGB (rouge, vert, bleu)**: C'est un système de codage informatique des couleurs avec 3 canaux où le premier représente le rouge, le deuxième représente le vert et le dernier représente le bleu. Chaque valeur de ces canaux est variée de 0 à 255.

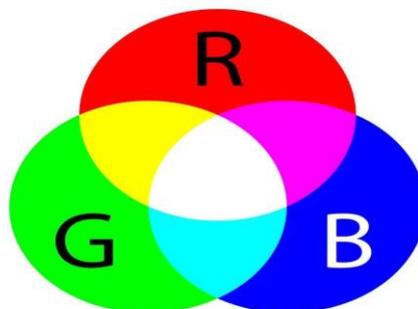


Figure 1.4 : Palette de RGB [3].

- **HSV (Teinte, saturation, luminosité)** : aussi, c'est un système de codage informatique des couleurs avec 3 canaux où le premier représente la teinte, le deuxième représente la saturation et le dernier représente la luminosité.

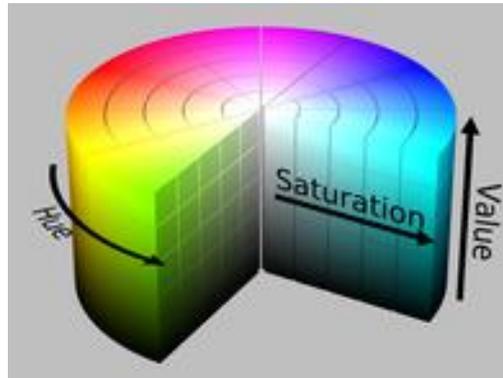


Figure 1.5 : palette de HSV.

1.4. Vidéo

La vidéo est une source multimédia qui combine une séquence d'images à une certaine cadence FPS (L'œil humain est capable de distinguer environ 20 images par seconde), et généralement, elle a des composants audio correspondant aux images affichées à l'écran. On trouve deux types des vidéos :

- **Les vidéos entrelacées** : ce type de vidéo est divisé en deux champs, un pour les lignes impaires de l'image et l'autre pour les lignes paires [4].
- **Les vidéos numériques** : ce genre de vidéo contient un seul champ qui inclut les lignes impaires et paires de l'image [4].

1.5. Acquisition des données

L'acquisition d'image constitue un des maillons essentiels de toute chaîne de conception et de production d'image. Pour pouvoir manipuler une image sur un système informatique, il est nécessaire de lui faire subir une transformation qui la rendra lisible et manipulable par ce système. La transformation de l'image réelle à une image numérique capable d'être interprétée par le système se fait grâce à une procédure de numérisation. Ces systèmes de saisie peuvent être classés en deux types principaux : les scanners et les caméras numériques [5] :

- **Scanner** : c'est un périphérique informatique qui permet de numériser des documents ou autres choses comme les empreintes digitales par exemple.
- **Caméra numérique** : Une caméra numérique est un appareil de numérisation qui utilise des capteurs électronique (CCD) pour transformer l'information lumineuse en un signal électrique.

1.6. Conclusion

Dans ce chapitre, nous avons abordé quelques notions de bases et certaines définitions élémentaires concernant le traitement des images et des vidéos après l'acquisition pour les utiliser par la suite dans notre système, où nous avons défini l'image numérique et le vidéo puis nous avons cité leurs caractéristiques (pixel, résolution, taille et espace de couleurs), et enfin les outils d'acquisition.

Chapitre 2 : Approches de détection de somnolence

2.1. Introduction

Selon les dernières recherches menées par l'Association Record Safety international Road Travel (ASIRT) [6], on trouve que les accidents de la circulation sont l'une des principales causes de mort dans le monde dans lesquels d'environ 1.2 millions personnes décède et d'environ 518 milliards dollars sont perdu chaque année. En moyenne de 3287 personnes et 1.4 milliards dollars chaque jour. La situation en Algérie n'est pas différente de celle d'autres pays. Selon les statistiques faites par le Centre National de Prévention et de Sécurité Routières algérienne (CNPSR) [7] en 2014, on trouve que 4812 personnes morts à cause des accidents de la circulation. Et avec une autre étude faite par Mariam. B et Laurent. C [8], ils trouvent que le nombre de tués et blessés dans les accidents de la circulation en Algérie est en croissance entre 1970 et 2015.

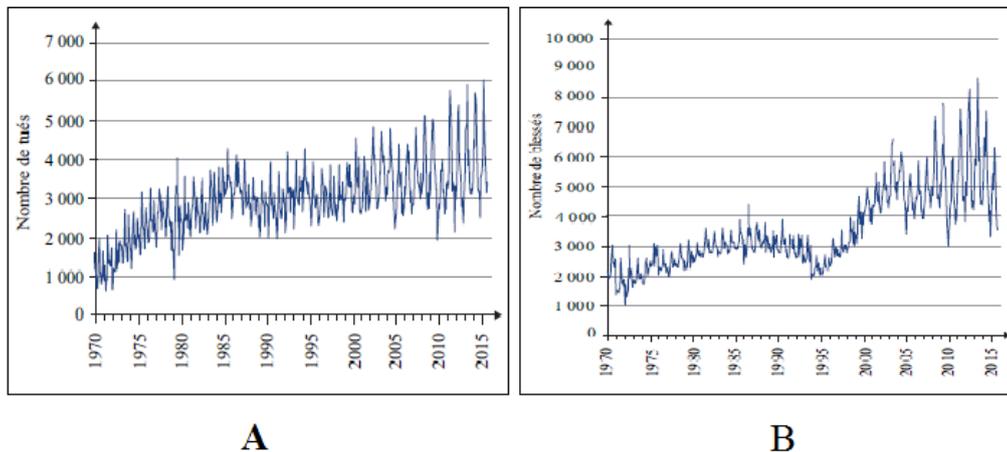


Figure 2.1 : le nombre des tués et des blessés entre 1970 et 2015 en Algérie où A est le nombre de tués et B le nombre de blessé [8].

L'une des principales causes des accidents est la somnolence du conducteur. À partir des études faites par National Highway Traffic Safety Administration (NHTSA) [9] en U.S.A, on trouve que 3662 personnes sont mortes entre 2011 et 2015 à cause de la somnolence du conducteur. Le tableau 2.1 représente les statistiques de cette étude. Selon les valeurs de tableaux, on observe que les pertes matérielles et humaines causées par la somnolence sont très élevées, et pour priver les accidents et réduire le nombre des morts et des blessés, plusieurs recherches ont été lancées et plusieurs technologies ont été développées (par exemple : les airbags, la ceinture de sécurité, ABS, etc...). Et dans les dernières années, une nouvelle technique a été proposée et utilisée dans les voitures populaires (comme :

Mercedes-Benz, Volvo, Audi, BMW, etc...). Cette technique est appelée « système de détection somnolence au volant du conducteur ».

Les méthodes qui sont utilisées dans cette technique sont divisées en deux groupes principaux : les méthodes axées sur les performances du conducteur et les méthodes axées sur l'état du conducteur. En outre, les méthodes axées sur l'état du conducteur sont divisé en deux groupes aussi: les méthodes utilisant des signaux physiologiques et les méthodes utilisant les traitements des images.

		Accidents	Accidents causé par somnolence	
		Nombre	Nombre	pourcentage
2011	Tués	29867	721	2.1%
	Blessé	1530000	29000	1.9%
	PDO	3778000	36000	1.0%
	Total	5338000	66000	1.2%
2012	Tués	31006	744	2.4
	Blessé	1634000	34000	2.1%
	PDO	3950000	47000	1.2%
	Total	5615000	81000	1.4%
2013	Tués	30202	717	1.4%
	Blessé	1591000	32000	2.0%
	PDO	4066000	39000	1.0%
	Total	5687000	72000	1.3%
2014	Tués	30056	747	2.5 %
	Blessé	1648000	33000	2.0 %
	PDO	4387000	53000	1.2 %
	Total	6064000	87000	1.4 %
2015	Tués	32166	736	2.3 %
	Blessé	1715000	33000	1.9 %
	PDO	4548000	56000	1.2 %
	Total	6296000	90000	1.4 %
2011-2015	Tués	153297	3662	2.4 %
	Blessé	8118000	160000	2.0 %
	PDO	20728000	232000	1.1 %
	Total	29000000	396000	1.4 %

Tableau 2.1: Les accidents routiers et les accidents routiers causés par la somnolence du conducteur entre 2011 et 2015 [9].

2.2. Relation entre la somnolence et la conduite

La somnolence est définie comme une diminution du niveau de conscience dépeint par la somnolence et la difficulté à rester alerte, mais la personne se réveille avec une simple excitation provoquée par les stimuli. Cela peut être dû à une absence de repos, à des médicaments, à un abus de substances ou à un problème cérébral. La somnolence est principalement le résultat de la fatigue qui peut être à la fois mentale et physique. D'autre part, n'importe quelle activité poursuit suffisamment longtemps, Causera à une difficulté à maintenir des performances qualifiées, y compris la conduite automobile. Et pour arrêter ou réduire le nombre des accidents, l'état de somnolence du conducteur devrait être surveillé en permanence.

2.3. Approches de détection de somnolence

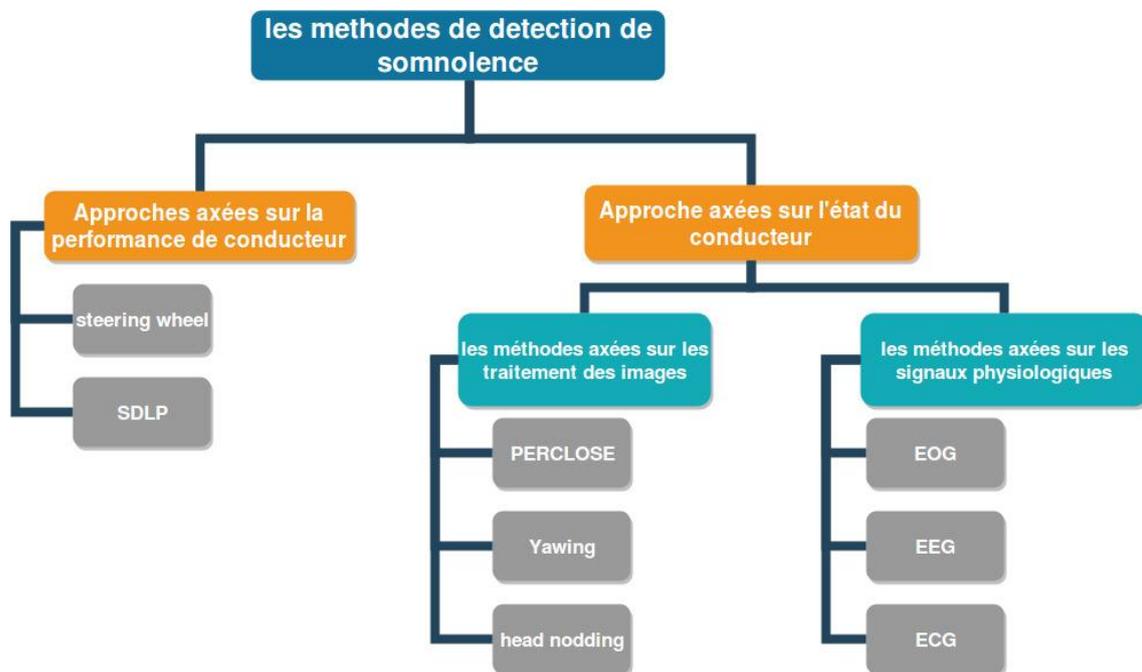


Figure 2.2 : Les différentes méthodes de détection de somnolence.

2.3.1. Approches axées sur la performance du conducteur

Généralement, le mode de conduite est influencé par plusieurs facteurs, les tâches de conduire (le changement de vitesses et l'accélération, courbure et largeur de la voie etc.) et les caractéristiques de conducteur (expérience de conducteur). En basant sur ces facteurs, on peut détecter l'état du conducteur. Dans la plupart des cas, on utilise des captures pour

surveiller les facteurs précédents par le placement de ces captures dans les différents composants de voiture comme le volant et la pédale d'accélération. Dans ces méthodes, plusieurs recherches étaient proposées, mais la détection de mouvement de volant et la détection de l'écart type de la position sont les mesures les plus couramment utilisées.

Krajewski et al [10] ont créé un système basé sur l'extraction des caractéristiques du mouvement de volant, qui a une relation avec le temps, la fréquence et l'environnement. Puis ils ont utilisé les algorithmes d'apprentissage SVM (Support Vector Machine) [11] et KNN (K-Nearest Neighbors) [12] pour classifier ces caractéristiques. Ce système atteint une précision de 86.1%. Il y a une autre mesure qui a été proposé, cette mesure est appelé « Standard Deviation of Lane Position » (SDLP) où nous pouvons évaluer les niveaux de somnolence du conducteur. Généralement, on utilise une caméra pour surveiller la voie. M.Inger et al [13] ont fait une comparaison entre SDLP et KSS [14] pour évaluer le SDLP et ils trouvent qu'il y a une relation directe entre KSS et SDLP, quand le SDLP augmente le KSS augmente aussi.

Ces approches affectées par plusieurs paramètres comme le type de voiture, l'état de la route et l'expérience du conducteur. Pour traiter ces mesures nous avons besoin beaucoup de temps et de données pour réaliser un système avec bonne précision.

2.3.2. Approches axées sur les signaux physiologiques

Quand une personne se sent somnolente ou fatiguée, de nombreux changements apparaissent sur son corps. Ces changements incluent les changements de la vitesse du rythme cardiaque, l'état du cerveau et l'état des muscles. Basant sur ces changements, plusieurs recherches et approches sont proposées. Dans ces approches généralement utilisent des capteurs portés par le conducteur comme on a indiqué dans la figure 2.3, Ces capteurs permettent à capter plusieurs types des signaux EEG (l'état du cerveau), EOG (mouvement des yeux) et ECG (rythme cardiaque). Dans Quelques recherches, ils ont utilisé les signaux EOG pour surveiller l'état du conducteur pendant la conduisent. Et Ceci est fait en classifier les mouvements des yeux à l'aide des signaux vertical et horizontal de l'EOG obtenu à partir des capteurs. La classification se fait selon l'un des algorithmes d'apprentissage SVM [15], PNN (Prototype Nearest Neighbor) et MLP (Multi-Layer Perceptron) [16].

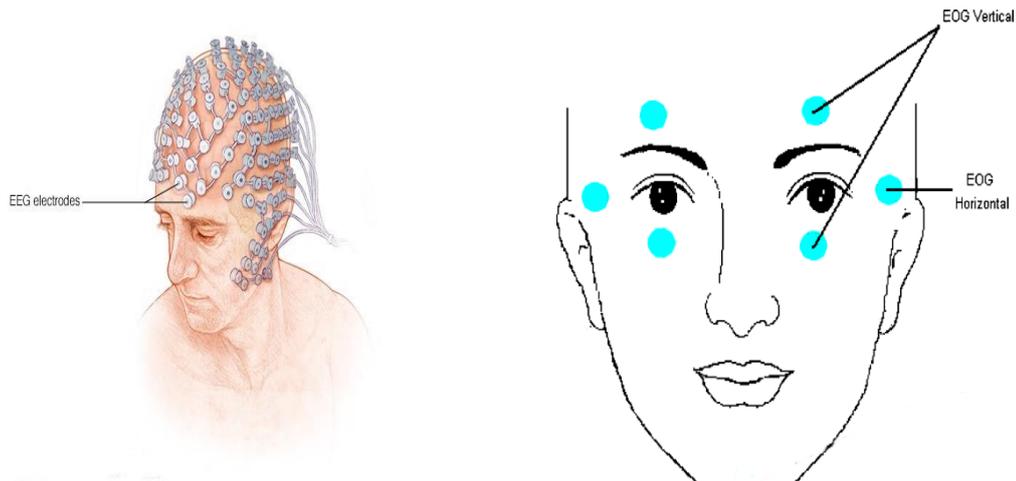


Figure 2.3: illustration des capteurs utilisés dans les méthodes axées sur la performance du conducteur [15].

Les signaux ECG sont aussi utilisés pour détecter l'état du conducteur. Ceci se fait à l'aide de variabilité du rythme cardiaque, dans lequel l'augmentation du rythme cardiaque est une indication de diminution du niveau de conscience, ce qui peut se produire au cours de la somnolence du conducteur [17]. EEG C'est également l'un des meilleurs signaux qui aident à étudier l'état du conducteur. Le EEG a différentes bandes de fréquences où la bande delta entre 0.5 et 4 Hz correspond à l'activité du dormir, la bande thêta entre 4 et 8 Hz correspond à la somnolence, la bande alpha entre 8 et 13 Hz correspond à relaxation et la bande beta entre 13 et 25 Hz correspond à la vigilance [18].

2.3.3. Approches axées sur le traitement d'image

Les approches précédentes sont classifiées comme des approches intrusives car elles sont influencées l'attention de conducteur. Maintenant, en raison du développement technologie, les recherches sont orientées vers les méthodes basées sur les traitements des images où on utilise une caméra pour extraire les symptômes de somnolence du conducteur comme les mouvements des yeux, les bâillements fréquents, les mouvements de tête.

La plupart des études publiées sur l'utilisation des traitements des images se concentrent sur les mouvements des yeux et plus précisément sur la mesure PERCLOSE [19] car plusieurs études prouvent que cette mesure est une mesure fiable pour prédire la somnolence.

Garcia et al [20] présentent un système consisté à 3 étapes principales, la première étape est le prétraitement où ils ont détecté le visage humain et les yeux avec l'algorithme

de viola et Jones, la deuxième étape est la localisation des pupilles et l'estimation de l'état des yeux avec la projection intégrale, et la dernière étape est prédire l'état du conducteur avec la mesure PERCLOSE.

Un autre système a été proposé par W. Zhang et al [21] où ils ont utilisé aussi la mesure PERCLOSE plus d'autres cinq mesures qui sont : la durée maximale de la fermeture, la fréquence de clignement, le niveau moyen d'ouverture des yeux, la vitesse de fermeture des yeux et la vitesse d'ouverture des yeux, et pour classifier ces mesures W. Zhang et al ont utilisé le Discriminant linéaire de Fisher.

Adrian [22] a proposé une autre approche où après la détection du visage et ses repères, il calcule l'EAR (Eyes aspect ratio) pour la détection de l'état des yeux (ouvert ou fermé) enfin, l'alarme est lancée si les yeux sont fermés dans 48 trames Consécutives.

Reza Ghoddoosian et al [23] propose un autre system où ils ont utilisé seulement trois caractéristiques : la fréquence de clignement, le niveau moyen d'ouverture des yeux et la vitesse d'ouverture des yeux et pour classifier ces mesures, ils ont utilisé les réseaux de neurones récurrents (l'architecteur LSTM).

Quoique des études prouvent que les mouvements des yeux sont des symptômes fiables pour prédire la somnolence, les autres symptômes ne sont pas moins importants que les mouvements des yeux où Dariusz Sawicki [24] propose un système qui peut distinguer entre les bâillements fréquents et l'ouverture de la bouche quand on parle, chante ou sourit. Ce système est composé de deux étapes principales : la première est la distinction des lèvres, ceci se fait avec l'un des algorithmes de détection des repères faciaux, et la deuxième étape est l'estimation de l'état de ces lèvres après le calcul de rapport de la hauteur de la bouche à leur largeur.

Sur le même symptôme et avec d'autres techniques M. Omidyeganeh et al [25] présentent un autre système pour prédire l'état du conducteur où ils ont utilisé l'algorithme de viola et Jones pour détecter le visage et la bouche, puis ils ont utilisé la théorie de back-projection pour mesurer à la fois le taux et l'ampleur des changements dans la bouche pour prédire son état et l'état du conducteur.

Les mouvements de tête sont aussi des symptômes qui peuvent être utilisés pour surveiller l'état du conducteur où les chercheurs Lam Thanh Hien et Do Nang Toan [26] proposent une approche où après avoir détecté le visage humain, certaines caractéristiques du visage sont extraites y compris deux coins éloignés des yeux, deux points de la bouche et

le bout du nez; ensuite, ils sont utilisés pour calculer les paramètres prédéterminés telles que la distance point à point, la distance point à bord et la zone triangulaire contrainte par deux points des coins de la bouche et de la pointe du nez. Enfin, la somnolence est évaluée en fonction des seuils prédéfinis pour ces paramètres.

Tous les systèmes mentionnés précédemment utilisent un seul symptôme pour détecter la somnolence. Mais, il y a d'autres approches proposées qui fusionnent plusieurs caractéristiques. M. Omidyeganeh et al [27] fusionnent deux caractéristiques : bâillement fréquent et PERCLOSE pour réduire le faux positif et augmenter la précision du système. Esra Vural et al [28] inventent un autre système qui fusionne plusieurs caractéristiques : le bâillement fréquent, les mouvements de tête, et les clignements des yeux ... etc. la plupart des approches développées extraient les caractéristiques à l'aide des méthodes décrites dans les approches précédentes.

Plus récemment, certaines approches qui utilisent l'apprentissage en profondeur, par exemple l'utilisation de CNN et RNN sont apparues. L'apprentissage en profondeur basé sur les CNNs et RNNs constitue une avancée décisive, en particulier pour les tâches de vision par ordinateur tel que la classification d'images, la détection d'objets la reconnaissance des émotions, la segmentation de scènes, etc. en fonction de ce concept, plusieurs approches sont proposées pour détecter la somnolence.

Bhargava Reddy et al [29] proposent un système basé sur les réseaux de neurone convolutifs .ce système est composé de deux étapes où dans la première étape, ils utilisent MTCNN (Multi-Task Cascaded Convolutional Neural Networks) pour détecter le visage, les yeux, le nez et lèvres gauche et droite, ...etc. Ce résultat représente l'entrée de la CNN de deuxième étape qui va extraire l'état des caractéristiques déjà extraite dans l'étape précédente et prédire l'état du conducteur.

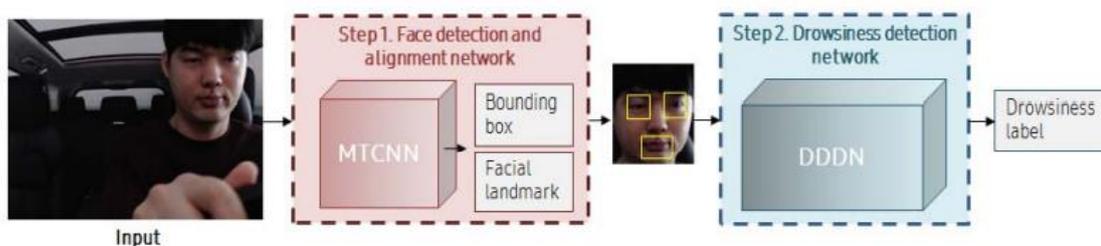


Figure 2.4 : Architecteur utilisé dans l'approche de Bhargava Reddy et al [29].

Un autre système inventé par Nisha Gandhi et al [30] utilise deux types d'apprentissage en profondeur ; ils utilisent CNN (Réseau neuronal convolutifs) pour détecter l'état des yeux (fermés ou ouverts) et RNN (réseau de neurones récurrent) pour détecter l'état du conducteur.

2.4. Conclusion

Ce chapitre présente les différents systèmes de détection automatique de somnolence du conducteur, où ces systèmes sont divisés en trois types principaux :

- **les méthodes basées sur les performances du conducteur** : ces méthodes sont des méthodes non intrusives, mais, ces approches sont affectées par plusieurs paramètres comme le type de voiture, l'état de route et l'expérience de conducteur.
- **Les méthodes basées sur les signaux physiologiques** : ces méthodes donnent des bons résultats mais elles sont des méthodes intrusives qui dérangent le conducteur tout en portant des capteurs.
- **Les méthodes basées sur les traitements des images** : ces méthodes sont des méthodes difficiles mais elles donnent des bons résultats et aussi elles sont des méthodes non intrusives.

Le tableau 2.2 contient une simple comparaison entre les méthodes de détection de somnolence au volant du conducteur.

Mesure	avantages	Inconvénients
La performance du conducteur	-Non intrusive	- Dépend de conditions extérieures
Le traitement d'image	-Non intrusive	- affectés par les conditions d'éclairage en arrière-plan
Les signaux physiologiques	-Détection anticipée	- intrusive -Faible qualité de signal dans les solutions non intrusives

Tableau 2.2 : comparaison entre les méthodes de détection de somnolence.

Le tableau 2.3 contient une simple comparaison entre les méthodes de détection de somnolence au volant du conducteur basées sur les traitements des images.

Mesure	Avantages	Inconvénients
Etat des yeux (PERCLOSE)	<ul style="list-style-type: none"> - contenir plusieurs informations décrivant l'état du conducteur. - Le grand danger réside dans la conduite avec les yeux fermés. 	<ul style="list-style-type: none"> - influencé par les lunettes et l'éclairage.
Bâillements fréquents (Yawning)	<ul style="list-style-type: none"> - détection du mouvement de la bouche facile. 	<ul style="list-style-type: none"> - ne prédire pas toujours la somnolence.
Mouvements de tête (head nodding)	<ul style="list-style-type: none"> - la combinaison avec d'autres symptômes donne des bons résultats. 	<ul style="list-style-type: none"> - ne prédire pas toujours la somnolence. - besoin de vision stéréoscopique (caméra 3d ou logiciel 3D).

Tableau 2.3 : comparaison entre les méthodes de détection de somnolence basées sur le traitement d'image.

Chapitre 3 : Détection et suivi du visage

3.1. Introduction

La détection des visages joue un rôle important dans des nombreuses applications telles que la vidéosurveillance, la reconnaissance faciale et la détection de la somnolence du conducteur etc... . C'est ce qui a fait la détection de visage est l'un des sujets les plus étudiés de la communauté de la vision par ordinateur. L'analyse des images et la détection des visages sont des capacités étonnantes pour l'être humain. Mais pour la machine, cette tâche reste un défi ouvert, malgré les efforts investis et les résultats obtenus à ce jour.

Ce chapitre est consacré à la détection du visage du conducteur dans des conditions réalistes. Nous commençons le chapitre avec une introduction, puis nous décrivons les algorithmes les plus connus dans le sujet de la détection des visages. Finalement, nous faisons une évaluation à ces algorithmes.

3.1.1. Détection d'objet

Comme nous avons dit. La tâche de détection des objets en général et des visages en particulier est une tâche très facile pour l'homme, mais, elle est très difficile pour l'ordinateur. Cette tâche est généralement inclue deux concepts : La détection de la position qui identifie par une point de début de l'objet, sa longueur et ça largeur ; et la classification où on identifie les types des objets qu'on a détectés sa position dans l'étape précédente.



Figure 3.1: Un exemple de détection de la position d'objets et leurs catégories [3].

3.1.2. Region of interest

Dans notre mémoire de fin d'études, nous concentrons sur le problème de détection de somnolence du conducteur basant sur l'état des yeux (Ouverts ou fermés). Et pour détecter cet état, nous faisons couper les régions du visage et yeux pour analyser séparément comme il est indiqué dans la figure 3.2. Ces régions sont appelées ROI (Region of interest) [31].

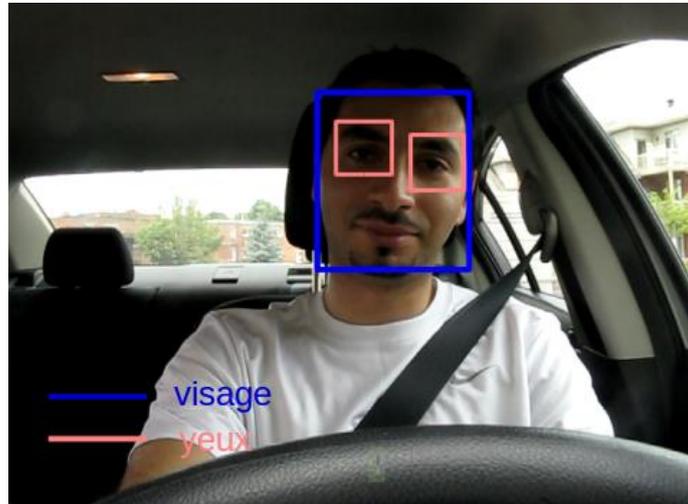


Figure 3.2: exemple sur les ROI de visage et les deux yeux [32].

3.2. Algorithmes de détection du visage

La détection des visages est l'identification de tous les visages d'une image ou d'une vidéo en fonction de ses caractéristiques telles que la forme, les yeux, le nez, la couleur de la peau, etc.... Ces propriétés sont souvent visibles pour l'être humain, mais, elles n'est pas visibles pour l'ordinateur car ce dernier voit l'image sous forme de tableau bidimensionnel, où chaque élément prend un entier compris entre 0 et 255 dans le cas d'une image en niveaux de gris (grayscale) ou d'un vecteur de 3 entiers où chacun représente une couleur (rouge, bleu et vert) dans le cas des images en couleur. Comme nous avons dit dans le chapitre 1.



Figure 3.3 : Exemple de sous image de 5 x 5 dans le réel et sur un ordinateur [3].

Et pour que l'ordinateur puisse simuler des êtres humains, plusieurs techniques ont été inventées pour rendre l'ordinateur voire certain caractéristiques importantes tel que les bordures, les coins et les lignes...etc. La binarisation, le seuillage et la normalisation sont quelques anciennes techniques utilisées dans la vision par ordinateur pour détecter le visage. Ces techniques sont influencées par certains changements dans la luminosité et la couleur de fond. Ce problème a conduit à l'invention des nouvelles techniques basées sur l'extraction des caractéristiques comme l'approche de Viola et Jones qui utilise les caractéristique de haar [33] et Rekha et Kurian qui utilise les caractéristique de HOG [34].

3.2.1. Viola & Jones

3.2.1.1. Caractéristiques

Les classificateurs de Viola et Jones utilisent les caractéristiques de type Haar [33] qui sont des caractéristiques locales calculées en soustrayant la somme d'une sous-région de la caractéristique de la somme de la région restante de la caractéristique. Virole et Jones utilisent trois types des caractéristiques : deux rectangles, trois rectangles et quatre rectangles. La valeur d'une caractéristique à deux rectangles est la différence entre la somme des pixels dans deux régions rectangulaires. Les régions ont la même taille et la même forme et sont adjacentes horizontalement ou verticalement. Une caractéristique à trois rectangles calcule la somme dans deux rectangles extérieurs soustraits de la somme dans un rectangle central. Enfin, une caractéristique à quatre rectangles calcule la différence entre les paires diagonales de rectangles [33].

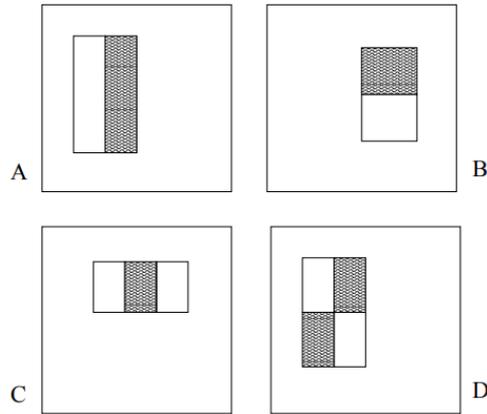


Figure 3.4 : Illustration des caractéristiques de Haar :
 (A, b) deux rectangles, (c) trois rectangles, (d) quatre rectangles [33].

3.2.1.2. Image Intégral

Il est facile de voir que même pour les petites images, il existe de nombreuses combinaisons des caractéristiques (plus de 180 000 pour une image 24 x 24). Et comme l'algorithme nécessite une itération sur toutes les caractéristiques, celles-ci doivent être calculées de manière très efficace. Pour ce faire, Viola et Jones introduisent l'image intégrale. Cette image est calculée avec la formule suivante :

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (3.1)$$

où $ii(x, y)$ est l'image intégrale et $i(x', y')$ l'image principal. Avec l'image intégrale, toute somme rectangulaire peut être calculée en quatre références de tableau (figure 3.5). Il est clair que la différence entre deux sommes rectangulaires peut être calculée en huit références. Comme les entités à deux rectangles définies ci-dessus impliquent des sommes rectangulaires adjacentes, elles peuvent être calculées en six références de matrice, huit dans le cas des trois entités rectangulaires et neuf pour des entités à quatre rectangles [33].

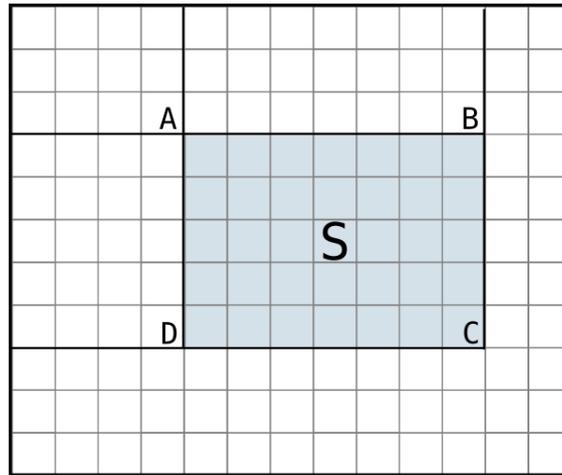


Figure 3.5 : Illustration de l'image intégrale où la surface $S = C + A - B - D$

3.2.1.3. Cascade attentionnelle

Pendant la phase d'entraînement, Viola et Jones ont utilisé l'algorithme d'AdaBoost (Adaptive boosting). Où L'idée générale de cet algorithme consiste pour chaque classificateur faible ultérieur à corriger les erreurs du classificateur précédent. Pour ce faire, ils attribuent une pondération à chaque exemple d'apprentissage, entraînent les classificateurs, choisissent le meilleur classificateur et mettent à jour les pondérations en fonction de l'erreur du classificateur. Les exemples mal étiquetés auront des poids plus importants et seront donc correctement classés par le prochain classificateur choisi. Tous les classificateurs obtenus à l'étape précédente sont des classificateurs faibles et pour créer un classificateur puissant, Viola et Jones ont utilisé la cascade attentionnelle. Cette technique permet la création d'une série des classificateurs qui sont organisés de complexe a plus complexe.

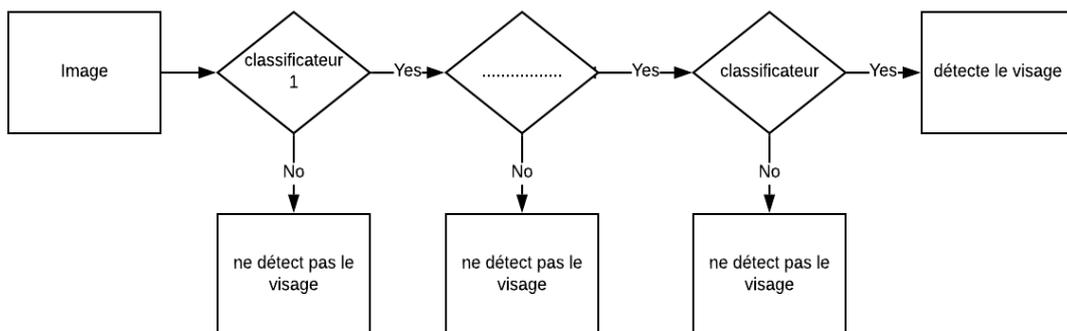


Figure 3.6 : Illustration du classificateur en cascade.

3.2.2. Histogram of oriented gradients

Histogram of oriented gradients (HOG) est une caractéristique utilisée en vision par ordinateur pour la détection d'objet. Cette technique calcule des histogrammes locaux de l'orientation du gradient sur une grille dense, c'est-à-dire sur des zones régulièrement réparties sur l'image. Cette caractéristique a été utilisée pour la première fois par Navneet Dalal et Bill Triggs en 2005 pour détecter les piétons. Mais elle peut être utilisée pour détecter d'autres objets comme les voitures, Les avion, les visages...etc. Cette méthode est très simple et on peut la résumer en deux étapes principales : le calcul du vecteur des caractéristiques Histogrammes et la classification.

3.2.2.1. Calcul de vecteur des caractéristiques Histogrammes

Comme nous avons dit. La méthode de HOG est destinée à la détection des piétons à l'origine. Pour cette raison Navneet Dalal et Bill Triggs [35] a utilisé une fenêtre de taille (64x128). Bien sûr, une fenêtre peut avoir n'importe quelle taille selon l'objet que l'on souhaite détecter. La première étape dans la réalisation de vecteur d'histogramme est le calcul du gradient vertical et horizontal. Ceci est facilement réalisé avec les filtres indiqués dans la figure 3.7.

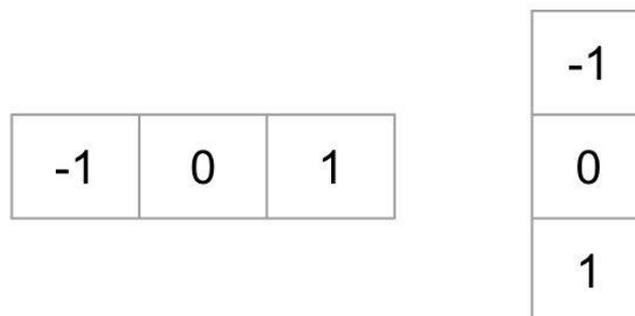


Figure 3.7 : les filtres qui utilisent dans le calcul de gradient verticale et horizontale.

Quant à la deuxième étape, elle consiste à calculer la longueur et la direction du vecteur en fonction des formules (3.2) et (3.3) et les résultats de la première étape.

$$g(x, y) = \sqrt{i_x^2 + i_y^2} \quad (3.2)$$

$$\Theta = \arctan\left(\frac{i_y}{i_x}\right) \quad (3.3)$$

3.2.2.2. Classification

Cette étape nécessite un modèle basé sur l'un des algorithmes d'apprentissage supervisé tels que l'arbre de décision, la forêt aléatoire, les machines à vecteurs de support, etc... . Comme ces techniques font partie des algorithmes d'apprentissage supervisé, elles nécessitent un grand nombre d'échantillons positifs (l'objet que l'on souhaite détecter) et d'autre échantillons négatifs (sans l'objet que l'on souhaite détecter). Navneet Dalal et Bill Triggs prouve que le SVM [11] est un bon choix dans le cas de détection des piétons.

3.2.3. Réseaux de neurones convolutionnels

Les réseaux de neurones convolutionnels (ConvNets ou CNN) constituent une catégorie de réseaux de neurone qui se sont révélés très efficaces dans des domaines tels que la reconnaissance et la classification des images. Les CNNs sont des variantes biologiquement inspirées du perceptron multicouche et ils ont été appliqués avec succès dans le domaine d'identification des objets et des visages. LeNet était l'un des tout premiers réseaux de neurones convolutifs proposés par Yann Lecun en 1988. Il était principalement utilisé pour des tâches de reconnaissance de caractères telles que la lecture des codes zipes, des chiffres, etc... .

En 2012, Alex Krizhevsky et Hinton ont gagné la compétition ImageNet en utilisant CNN, qui a ouvert une ère de reprise pour l'apprentissage en profondeur. Généralement, les ConvNets sont constitués de trois types de couches : les couches de convolution, les couches d'activation, les couches de pooling et les couches entièrement connectées. généralement, les

CNN obtient une image RGB en entrée et produit un ensemble de cadres pouvant contenir l'objet que l'on souhaite détecter (visages, voitures, chats...etc.). Tout ce que nous avons besoin est tout simplement sélectionner les cadres qui a une grande probabilité pour contenir l'objet.

Il y a plusieurs méthodes qui nous permettent de faire ceci parmi eux : R-CNN [36], Fast-R-CNN [37] et SSD [38].

3.2.4. Evaluation

La détection du visage par la vision par ordinateur, comme elle est démontrée précédemment, très complexe. Les différents algorithmes de détection du visage décrit précédemment possèdent des avantages et des inconvénients qui doivent être considérés lors de la conception du système complet de détection de somnolence. Pour ce faire, il est primordial de valider les techniques choisies sur des ensembles de données relativement volumineux. Même si de telles séquences d'images ne représentent pas exactement les conditions réelles d'utilisation, elles procurent néanmoins une idée fiable du comportement des différents modules dans un environnement contrôlé.

Dans le reste de cette sous-section nous évaluerons les 3 algorithmes décrits dans la sous-section précédente. L'implémentation de ces algorithmes est très difficile et l'obtention des résultats satisfaisants est plus difficile mais opencv et dlib ont rendu les choses très faciles où ces bibliothèques open source multiplateformes dédiées aux traitements des images et l'apprentissage automatique. Elles contiennent des modèles robustes et des APIs qui permettent d'implémenter ces algorithmes dans quelque ligne. Dans cette évaluation nous avons utilisé deux datasse FDDB et YawDD :

- **YawDD [32]** : est une ensemble des données qui contient deux sous-ensembles de données vidéo de conducteurs présentant diverses caractéristiques faciales, à utiliser pour tester des algorithmes et des modèles destinés principalement à la détection des bâillements, mais également à la reconnaissance et au suivi du visage et de la bouche. Les vidéos sont prises dans des conditions d'illumination réelles et variables. Notre évaluation se fait sur 10 vidéos de la première partie de dataset, Parce que l'annotation des visages dans toutes les vidéos prend très longtemps.



Figure 3.8 : Des Photos prises de l'ensemble des données YawDD [32].

- **FDDB [39]** : (Face Detection Data Set and Benchmark) cette dataset a été créé par Vidit Jain et Erik Learned-Miller dans l'université du Massachusetts en 2010 destinés à la détection et la reconnaissance des visages, cette dataset contient 2845 images et 5171 visages.



Figure 3.9: Des Photos prises de l'ensemble des données FDDB [39].

3.2.4.1. Intersection Over Union

Intersection Over Union (IOU) est une mesure basée sur l'indice de Jaccard qui évalue le chevauchement entre deux cadres. Cela nécessite une cadre réelle (A) et une cadre estimée par l'algorithme (B). En appliquant IOU, nous pouvons déterminer si une détection est valide ou non. IOU est calculé selon la formule:

$$IOU = \frac{A \cap B}{A \cup B} \quad (3.4)$$

À l'aide d'IOU, on peut calculer les TP, TN, FP et FN:

- TP (True Positive) : la détection est correcte cela signifie que $IOU > \text{seuil}$
- FP (False Positive) : la détection est incorrecte cela signifie que $IOU \leq \text{seuil}$
- FN (False Negative): Les cadres réelles qui n'ont pas été découvertes par l'algorithme
- TN (True Negative) ne prend pas en considération, car il y a nombreuse Régions qui ne sont pas indiqués dans l'ensemble de données et n'est pas détecté par le système.

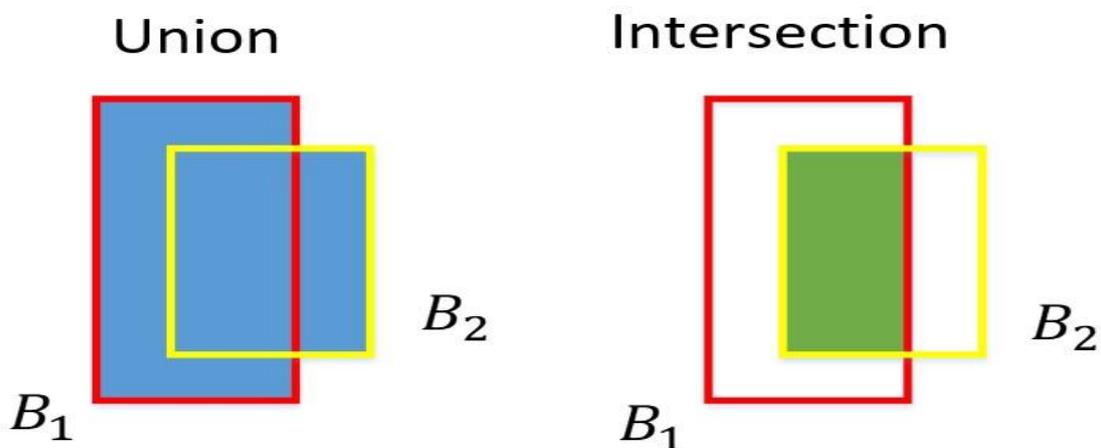


Figure 3.10: Illustration de l'intersection et l'union.

3.2.4.2. Matrice de confusion

Une matrice de confusion est un tableau souvent utilisé pour décrire les performances d'un modèle de classification sur un ensemble des données de test pour lesquelles les valeurs vraies sont connues. Dans notre cas est le visage.

		Classe estimée par l'algorithme	
		Visage	Absence de visage
Classe réelle	visage	TP	FN
	Absence de visage	FP	TN

Tableau 3.1 : Matrice de confusion d'un modèle de classification binaire.

Basant sur la matrice de confusion, on peut calculer deux mesures qui aident dans l'étude de comportement des algorithmes sur les ensembles des données. Ces mesures sont : la précision et le rappel :

- **Précision** : La précision est le rapport entre les observations positives (correctement prédites) et le total des observations positives et observations négative (incorrectement prédites), cette mesure est calculé avec la formule suivant :

$$Précision = \frac{TP}{TP + FP} \quad (3.5)$$

- **Rappel** : Le rappel est le rapport entre les observations positives (correctement prédites) et toutes les observations de la classe réelle dans l'ensemble de données. cette mesure est calculée avec la formule (3.6) :

$$rappel = \frac{TP}{TP + FN} \quad (3.6)$$

3.2.5. Résultats et discussion

Dans cette section, nous faisons une discussion sur les résultats obtenues à partir les tests qui nous avons effectué sur un laptop de marque DELL doté d'un processeur i3 avec 2 noyaux de 1.8 GHz, 8 GB de RAM et disque dur de 500 GB. Comme nous avons dit, le test se fait sur les ensembles des données FDDB et YawDD. Les résultats de ce test sont présents dans le tableau 3.2 et 3.3 où la précision décrit le comportement de l'algorithme sur les objets détectés et le rappel décrit le comportement de l'algorithme sur les objets mentionné dans l'ensemble des données. Le meilleur algorithme est celui qui possède la meilleure valeur de précision et rappel.

Il est un peu difficile d'analyser les deux mesures au même temps. Pour des raisons de simplicité, nous avons calculé le F1-score. Ce dernier combine les deux mesures précédentes en une seule mesure selon l'équation (3.7) :

$$F1 - Score = \frac{2 \times \text{précision} \times \text{rappel}}{\text{précision} + \text{rappel}} \quad (3.7)$$

	TP	FP	FN	Précision	rappel	F1-Score
Viola-Jones	1299	94	3375	0,93	0,27	0.42
HOG	2994	316	1680	0,90	0,64	0.74
CNN	3826	461	848	0,89	0,81	0.84

Tableau 3.2: Les résultats de test sur l'ensemble des données FDDB.

	TP	FP	FN	Précision	rappel	F1-Score	FPS
Viola-Jones	14723	4	9151	0,99	0,61	0.75	98
HOG	23314	82	560	0,99	0,97	0.97	41
CNN	23718	0	156	1,0	0,99	0.99	9

Tableau 3.3: Les résultats de test sur l'ensemble des données YawDD.

D'après les valeurs des tableaux, nous observons que l'approche basée sur le CNN présente le F1-score le plus élevé, suivie de l'approche HOG et de l'approche Viola Jones. Sur la base de cette comparaison uniquement, nous ne pouvons pas décider que CNN est la meilleure option pour notre problème parce que cette approche fonctionne à une vitesse estimée de 9 FPS sur nos machines et de 0,5 FPS sur le raspberry. Cela affecte négativement sur les performances finales du système et le rend incapable de fonctionner en temps réel.

D'autre part, la vitesse de l'approche HOG et viola-Jones sont estimés à 41 et 98 FPS sur nos machines et ils sont estimés à 5 et 22 FPS sur le raspberry. Ces valeurs aussi, ne permettent pas de dire que l'approche de HOG où viola-Jones est une bonne option à notre problème parce qu'il est un peu difficile de fonctionner en temps réel sur raspberry avec 5 FPS surtout que la détection des visages est la première étape du système, et inévitablement, l'ajout d'autres étapes qui réduisent la vitesse et quoique l'approche de viola-Jones et fonctionne normal sûr les deux (nôtres machine et raspberry), il ne faut pas oublier que l'approche viola-Jones a F1-score le plus bas.

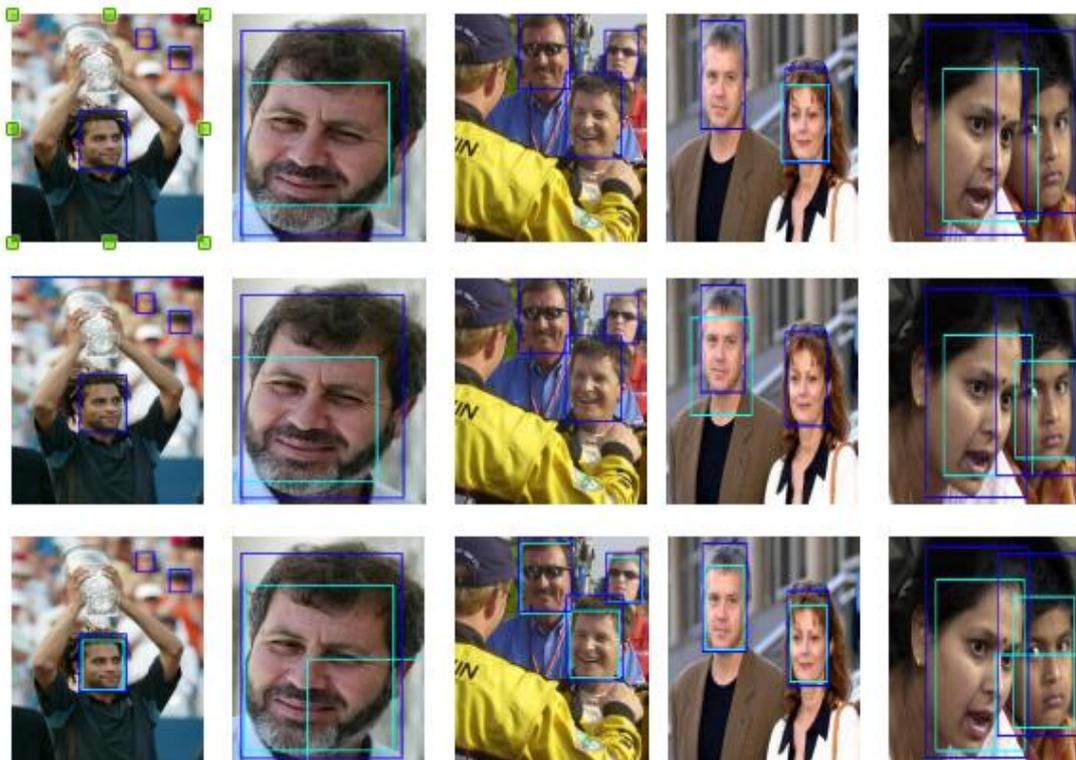


Figure 3.11 : Les résultats de nos expériences sur les trois algorithmes ; La première ligne représente les résultats de Viola Jones, le deuxième HOG, le dernier CNN où Le cadre bleu représente le class réel et l'autre cadre représente les prédictions des algorithmes.

3.3. Suivi du visage

Le suivi des objets en générale et le suivi de visage en particulier est l'une des tâches les plus difficiles de la vision par ordinateur et il est associé à un large éventail d'applications surtout la vidéosurveillance. Le suivi des objets dans sa forme la plus simple, est le processus de localisation des objets en mouvement dans le temps dans chacune trame d'une séquence vidéo. Cette méthode comporte deux étapes principales:

- La première étape consiste à représenter l'objet que l'on souhaite suivre par des descripteurs des caractéristiques de forme et/ou d'apparences qui permettent de le reconnaître dans les trames successives par exemple l'histogramme des couleurs.
- La deuxième étape est représentée dans la prédiction de la position de l'objet dans les images suivantes basant sur le modèle de représentation et de son état dans l'image précédente.

Plusieurs algorithmes ont été proposés dans ce sujet mais le filtre de Kálmán, camshif et les suivis de filtres de corrélation sont les algorithmes les plus célèbres et utilisés dans les dernières années. Dans cette section nous avons décrit l'algorithme camshif et le principe général des suivis de filtres de corrélation(CFT).

3.3.1. Cam-shift

L'algorithme camshift est une amélioration de l'algorithme mean-shift publié la première fois par GARY Bradski en 1998 [40]. Où cet algorithme permet à résoudre le problème de changement de taille de l'objet.

- **Mean-Shift**

L'idée de base du camshift est identique à celle du suivi de brut-force, à savoir que nous pouvons supposer que dans les trames suivantes de vidéo, la cible ne changera pas beaucoup de position ni d'apparence. Puisque nous mesurons notre candidat cible sur une région, nous pouvons supposer qu'une partie de l'objet cible restera sous la région de la position précédente.

Généralement, La procédure Mean-Shift est basée sur la distribution globale des caractéristiques d'intensité (des couleurs ou des niveaux de gris) de l'objet cible. Initialement, un modèle pour l'objet à suivre est sélectionné et l'histogramme de

couleur pour la région qui englobe l'objet cible est ainsi construit. Ensuite, la fonction de densité de probabilité est calculée de la façon suivante :

$$q_u = C \sum_{i=1}^n K(\|x_i\|)^2 \delta[b(x_i) - u] \quad (3.8)$$

où C est une constante de normalisation, k est un noyau qui favorise les centroïdes d'objets, u est l'ensemble d'intervalles de l'histogramme, n est le nombre de pixels de l'objet, b est la valeur de l'intervalle de l'histogramme au pixel x_i et enfin δ est la fonction de Kronecker. Par la suite, la distance de Bhattacharyya est utilisée afin de comparer les histogrammes associés aux objets. Cette opération est répétée jusqu'à ce que la valeur de similarité ne dépasse pas un seuil ou que le nombre limite d'itérations soit atteint (de quatre à six itérations en général). Une phase de réinitialisation du modèle pour les objets cibles permet de gérer les occultations entre eux [41].

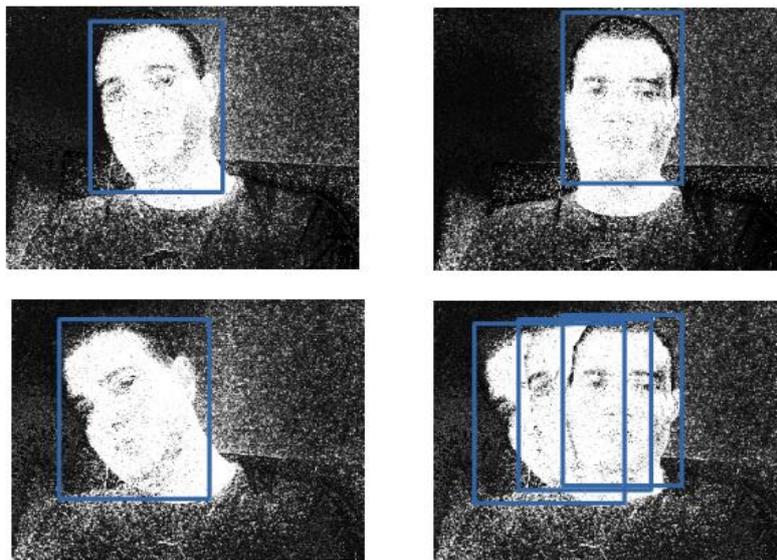


Figure 3.12: illustration de suivi d'un visage.

3.3.2. Suivi de Corrélation

Ces dernières années, les suivis de filtres de corrélation (CFT) ont reçu une grande attention dans le domaine du suivi d'objets visuels et ont obtenu des très bons résultats. Généralement, les suivis de filtres de corrélation représentent l'objet que l'on souhaite suivre

avec un filtre formé à partir d'instances décalées de la cible. Ce dernier est initialement sélectionné sur la base d'une petite fenêtre de suivi centrée sur l'objet dans la première trame. Dans la trame suivante, la cible est suivie en corrélant le filtre sur une fenêtre de recherche. La réponse du filtre est évaluée et le maximum donne la nouvelle position de la cible. Le filtre est formé en ligne et mis à jour successivement à chaque image afin que le suivi s'adapte aux changements de cible modérés. L'avantage majeur du suivi du filtre de corrélation est l'efficacité du calcul. La raison en est que le calcul peut être effectué efficacement dans le domaine de Fourier selon la formule suivante.

$$G = F \odot H \quad (3.9)$$

Où G est le résultat de corrélation de la trame entrée et le filtre, F la transformation de Fourier de la trame et H la transformation de Fourier de filtre. Il convient de souligner ici que la corrélation devient une multiplication élémentaire dans le domaine de Fourier [42].

Bien que ces filtres permettent d'effectuer efficacement des tâches de localisation, les besoins en formation requis les rendent inappropriés pour le suivi en temps réel. Mais l'approche de Minimum Output Sum of Squared Error (MOSSE) proposée par D.S.Blome et al [42] en 2010 a changé la situation. Où l'idée de base de cette approche est de trouver un filtre qui minimise la somme de carré d'erreur:

$$\min_H \sum_i |F_i \odot H - G_i|^2 \quad (3.10)$$

Le suivi MOOSE fonctionne bien sur les objets en mouvement, mais pas sur les objets redimensionnés. Ce problème a été résolu dans le 2014 par l'approche Danelljan et al [43] qui utilisent une pyramide d'échelle pour estimer la dimension d'objet après la détection de sa position. L'implémentation de cette approche se trouve dans la bibliothèque dlib.

3.3.3. Résultats et discussion

Le test a été effectué toujours sur la même machine, sauf que cette fois on utilise seulement le dataset qui contient les vidéos (YawDD). Les résultats de ce test sont présentés dans le tableau 3.4. En observant ces valeurs, on voit clairement la supériorité

de dlib-correlation sur camshift en termes de la précision et rappel et la supériorité de camshift sur dlib-correlation en terme vitesse. Cela augmente la difficulté du choix car la vitesse ne peut être préférée à la précision et rappel.

	TP	FP	FN	Précision	rappel	F1-Score	FPS (PC-raspberry)
camshift	13738	10136	10136	0.57	0.57	0.57	100-44
Dlib-correlation	22489	1385	1385	0.94	0.94	0.94	47-7

Tableau 3.4 : les résultats de test sur l'ensemble des données YawDD

3.4. Conclusion

La détection de visage est une étape très importante dans la détection de somnolence du conducteur car le visage contient tous les importants symptômes de somnolence (bâillement, mouvement de tête, yeux fermé). Et pour analyser ces symptômes, il est important que la méthode établie soit précise et rapide. La principale préoccupation de ce chapitre est de choisir une méthode capable de détecter un visage dans toutes les conditions possible. Parmi ces méthodes nous avons étudié deux types des méthodes la détection et suivi.

- dans la détection, nous avons trouvé que l'approche CNN est la plus précise et l'approche de viola-Jones est la plus rapide.
- dans la suivi, nous avons trouvé que l'approche Dlib-correlation est la plus précise et l'approche camshift est la plus rapide.

Chapitre 4 : Détection des yeux

4.1. Introduction

Ces dernières années, la détection des yeux est devenue un sujet de recherche important dans le domaine de la vision par ordinateur et de la reconnaissance des formes parce que les yeux contiennent des informations essentielles pour de nombreuses applications (ex : la reconnaissance de l'expression faciale, le diagnostic médical, détection de somnolence...etc.). Par conséquent, la détection des yeux est une étape cruciale et compliquée. Dans cette étape, Nous allons former nos propres modèles qui utilisent les bibliothèques opencv et dlib. Contrairement à l'étape précédente où nous avons utilisé des modèles préformés par ces deux bibliothèques.

4.2. Collection de données

La collecte de données est le processus de collecte et de mesure d'information sur des variables ciblées. Ce processus est le roi de l'apprentissage machine. On ne pourrait pas former un classificateur robuste sans un bon ensemble des données. Pour former notre propre classificateur, nous avons besoin plusieurs d'échantillons, ce qui signifie que nous avons besoin de nombreuses images montrant l'objet cible (échantillon positif) et encore plus d'images sans l'objet cible (échantillon négatif). Ces images doivent être variées de qualité, d'arrière-plan, nature de l'objet et conditions d'éclairage.

4.3. Echantillons positifs

Pour construire les échantillons positifs. Nous collectons environ de 2699 images qui montrant des humains ou des visages humains pour assurer l'existence des yeux qui représente l'échantillon positif dans notre cas d'étude. Puis, nous étiquetons tous les yeux dans chaque image avec un outil open source appelé labeling. Cet outil permet de construire un fichier XML qui contient le nom de l'image, taille, position de chaque yeux dans l'image et son étiquette (ouvertes, ou fermer). Après cette étape, Nous avons collecté environ 4040 yeux ouverts et 1532 fermés ces yeux sont divisés entre des hommes, des femmes, avec des lunettes et sans Lunettes...etc. dans la collection de ces échantillons, nous avons utilisé de nombreux ensembles des données:

- CEW (Closed Eyes In The Wild) [44] : est un ensemble des données Contient des visages avec 1192 yeux fermé et 1231 yeux ouvert.
- Helen dataset [45]: Ce l'ensemble de données a dirigé à la localisation des caractéristiques du visage. Ce l'ensemble contient 2000 images des personnes dans des positions et éclairage différent.
- BIOID [46]: Ce l'ensemble de données comprend 1521 images où chaque image contient d'un visage de 23 personnes différentes.

4.4. Echantillons Négatifs

Pour construire les échantillons négatifs. Nous avons collecté environ de 16846 images qui ne contiennent pas l'objet cible (yeux). La construction de ces échantillons. Ce n'était pas difficile où nous avons utilisé deux ensembles des données MSO (Multi Salient Objet) [47] et Coco train 2014 [48] en retirant tous les images qui peuvent contient les yeux.

Il convient de souligner ici que les ensembles d'images qui nous avons collecté et nous avons utilisés lors de nos expériences, sont normalisés géométriquement en une taille de 300×300 pixels et rendre tous les images en grayscale.

4.5. Entraînement

Dans cette partie, nous avons utilisé l'ensemble de données précédentes pour former nos modèles. Cet ensemble de données est divisé en deux parties: 75% de cet ensemble de données représente la partie de formation et le reste représente la partie de teste :

- **Viola-Jones:**

Pour former les modèles de Viola Jones, nous avons utilisé des échantillons positifs et négatifs en quantités différentes d'un modèle à l'autre .la formation de ces modèles déroule sur une machine de 8 Go et CPU i3 avec deux noyaux de 1.8 GHZ. Et chaque model a pris de 4 jusqu'à 7 jours selon le nombre des échantillons utiliser.

- **HOG:**

Pour former nos modèles de HOG nous avons utilisé Google Colaboratory (Colab). Le trainement de tous les modèles pris environ de 1 jour. Où nous avons formé 32 modèles avec des valeurs différentes des paramètres c et σ .

4.6. Résultats et discussion

L'objectif de cette section est de présenter les résultats obtenus à partir le test des modèles sur la partie test de l'ensemble de donnée et de les interpréter tout en faisant une comparaison. Pour choisir le bon modèle à notre problème. Les tests ont fait selon trois conditions pour chaque approche (viola-Jones ou HOG) où dans la première condition nous avons testé les modèles sur l'image entière, la deuxième sur le visage seulement et la dernière sur le deuxième quartier de visage.

4.6.1. Viola & Jones

- **Condition 1 : Image entière**

Dans ce test nous avons appliqué les modèles sur l'image entière et les résultats sont présentés dans le tableau. Ou les modèles de 1 jusqu'à 7 sont nos modèles et le model 8 est le modèle de opencv. A travers les valeurs du tableau 4.1, il apparaît clairement la supériorité de quatrième, sixième et septième modèle sur le modèle associer à opencv et les autres modèles.

Malgré cette supériorité, cela ne suffit pas car la valeur de FN est plus élevée et cela affectera sur le rappel et bien sur le F1-Score.

Numéro de modèle	Nombre d'échantillon positif	Nombre d'échantillon négatif	TP	FP	FN	Précision	Rappel	F1-Score
1	3500	10500	1308	1442	89	0.47	0.93	0.63
2	3500	17000	1308	2333	89	0.35	0.93	0.51
3	3500	3500	1290	577	107	0.69	0.92	0.79
4	3000	17351	1064	59	333	0.94	0.76	0.84
5	3500	7000	1216	2152	181	0.36	0.87	0.51
6	3500	1000	1167	282	230	0.80	0.83	0.82
7	2000	6000	941	22	451	0.97	0.67	0.80
8	-	-	836	298	561	0.73	0.59	0.66

Tableau 4.1: les résultats de test dans la première condition.

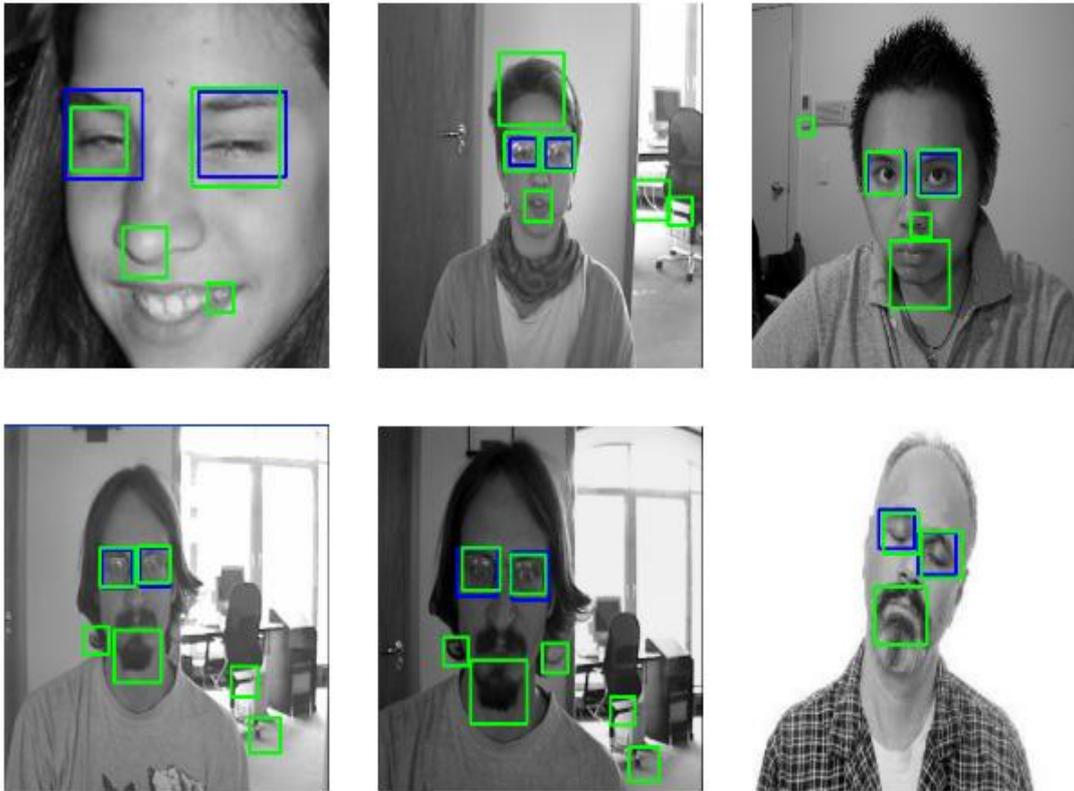


Figure 4.1 : Des exemples de résultats dont le premier cas où les cadres bleus représentent les class réel et les cadres verts représente les class prédit par le model numéro 5.

- **Condition 2: Visage seulement**

Dans ce test, nous avons d'abord détecté le visage, puis appliqué le modèle à cette zone uniquement et les résultats sont toujours affichés dans le tableau 4.2. Selon les valeurs du tableau, nous pouvons observer une amélioration significative des modèles 1, 2 et 5, où le score f1 a augmenté de 0,63, 0,51 et 0,51 à 0,73, 0,66 et 0,70. Cette amélioration est due de l'augmentation de précision et la diminution de FP.

Numéro de modèle	Nombre d'échantillon positif	Nombre d'échantillon négatif	TP	FP	FN	Précision	Rappel	F1-Score
1	3500	10500	1299	820	98	0.61	0.92	0.73
2	3500	17000	1294	1209	103	0.51	0.92	0.66
3	3500	3500	1296	386	101	0.77	0.92	0.84
4	3000	17351	1064	56	333	0.95	0.76	0.84
5	3500	7000	1329	1027	68	0.56	0.95	0.70
6	3500	1000	1156	199	241	0.85	0.82	0.84
7	2000	6000	943	18	454	0.98	0.67	0.79
8	-	-	821	271	576	0.75	0.58	0.65

Tableau 4.2: les résultats de test dans la deuxième condition



Figure 4.2: Des exemples de résultats dans le deuxième cas où les cadres bleus représentent les class réel et les cadres verts représente les class prédit par le model numéro 5.

- **Condition 3 : Deuxième quartier de visage**

Dans la plupart des cas, les yeux sont situés en haut du visage et plus précisément dans deuxième quartier de visage. Dans ce test nous avons utilisé ces information pour diminuer le nombre de FP et bien sûr pour augmente la précision.

Selon les valeurs de tableau, nous avons observé aussi une amélioration significative des modèles 1, 2 et 5, où le score f1 a augmenté de 0,73, 0,66 et 0,70 à 0,93, 0,92 et 0,94.

Numéro de modèle	Nombre d'échantillon positif	Nombre d'échantillon négatif	TP	FP	FN	Précision	Rappel	F1-Score
1	3500	10500	1297	68	100	0.95	0.92	0.93
2	3500	17000	1290	96	107	0.93	0.92	0.92
3	3500	3500	1296	45	101	0.96	0.92	0.94
4	3000	17351	1064	52	333	0.95	0.76	0.84
5	3500	7000	1321	70	76	0.94	0.94	0.94
6	3500	1000	1156	64	241	0.94	0.82	0.88
7	2000	6000	943	34	454	0.96	0.67	0.79
8	-	-	818	172	579	0.82	0.58	0.68

Tableau 4.3 : les résultats de test dans la troisième condition.

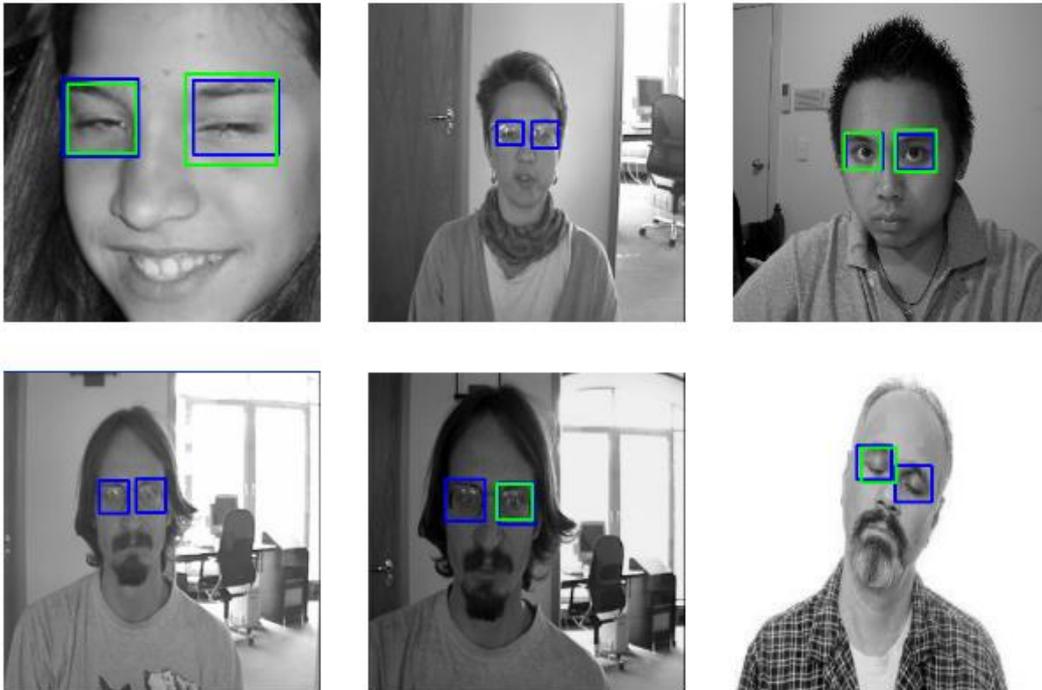


Figure 4.3 : Des exemples des résultats dans le troisième cas où les cadres bleus représentent les class réel et les cadres verts représente les class prédire par le model numéro 5.

4.6.2. Histogram of oriented gradients

- **Condition 1 : Image entière**

Après la formation des modèles HOG, nous avons aussi testé dans les trois conditions précédentes. Et nous avons commencé par le test sur l'image entière. Les résultats de ce test sont représentés dans les heatmaps de figures où l'heatmap plus haut gauche représente la précision et le plus haut droit représente le rappel et le bas gauche représente le F1-score. Tous ces paramètres sont varié en fonction de C et σ .

Et selon ces graphes nous observons que le f1-score est dans l'intervalle de 0.74-0.89, précision dans l'intervalle 1.0-0.96 et rappel dans l'intervalle 0.54-0.83. Nous avons vu aussi que les modèles ont formé dans le $c=1000$ ont le plus haut f1-score et les modèles ont formé dans $c=10$ ont le plus bas f1-score quoique sa précision soit 1.0.

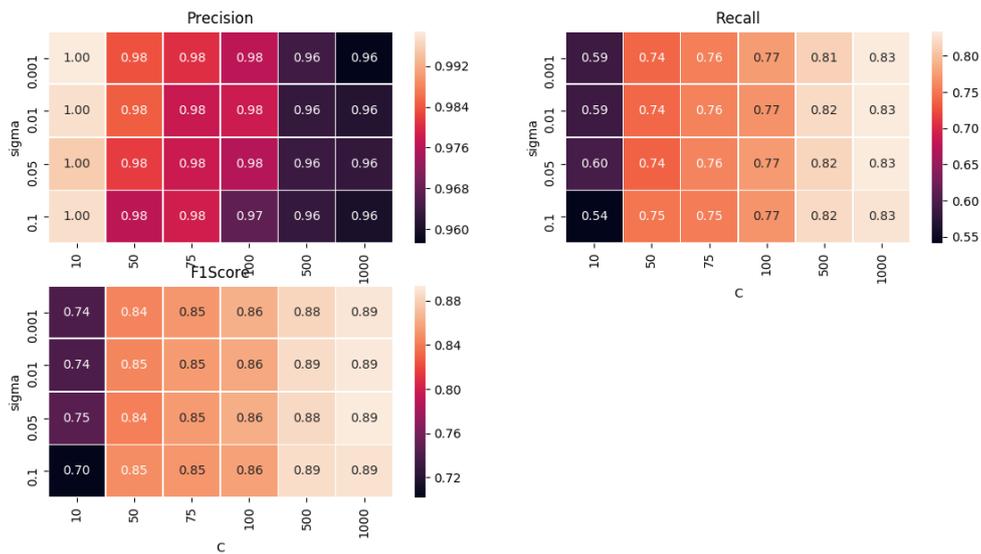


Figure 4.4 : les résultats de test des modèles HOG dans la première condition.

- **Condition 2: Visage seulement**

Dans ce test, nous avons appliqué les modèles uniquement au visage. Les résultats de ce test sont représentés dans les heatmaps. Comme nous le constatons, il y a une légère amélioration des valeurs de F1-scores, précision et rappel.

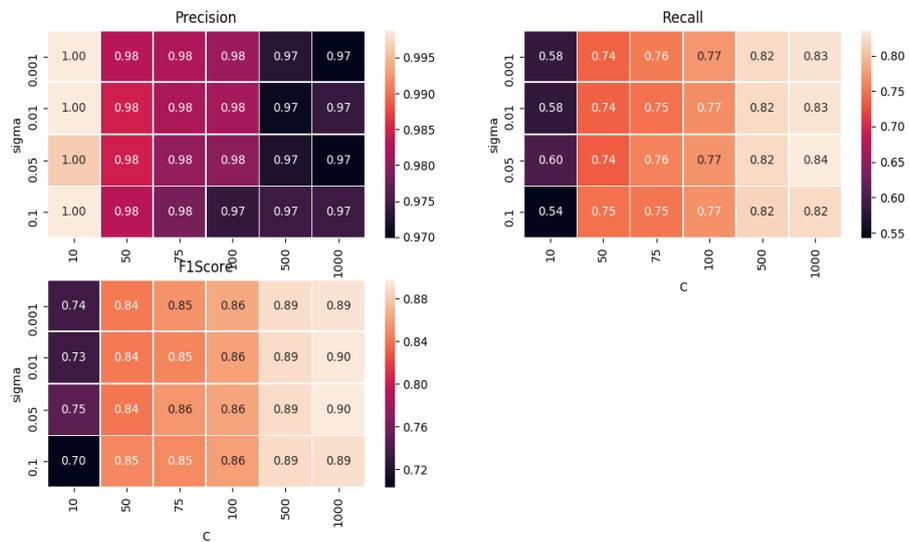


Figure 4.5: les résultats de test des modèles HOG dans la deuxième condition.

- **Condition 3: Deuxième quartier de visage**

Ce test n'a pas affecté positivement sur les valeurs où nous observons que les valeurs de f1-score, précision et rappel sont diminuées.

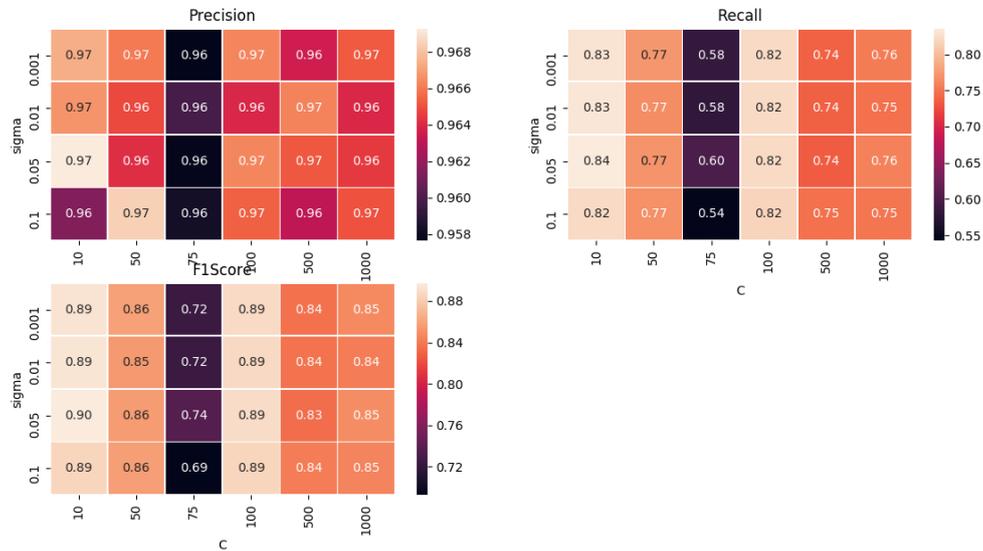


Figure 4.6 : les résultats de test des modèles HOG dans la deuxième condition



Figure 4.7 : Des exemples des résultats dans le troisième cas où les cadres bleus représentent les class réel et les cadres verts représente les class prédire par le model qui a $c=1000$, $\sigma=0.05$

4.7. Conclusion

La détection des yeux est une étape très importante, car elles contiennent plusieurs informations qui indiquent le statut du conducteur (sommolence, fatigue, Réveil). Et pour détecter le statut du conducteur, il est important que la méthode de détection des yeux établie soit précise et rapide. La principale préoccupation de ce chapitre est de former des modèles capable a Fonctionné bien dans tous la condition. Dans ce chapitre nous avons réalisé des modèles de violèrent-Jones et HOG. Et après le test dans les trois conditions, nous avons trouvé que les modèles HOG ont fonctionné bien dans la première condition, Contrairement aux modèles de viola-Jones. Cette situation s'est inversée dans la troisième condition où les modèles 1, 2,3 et 5 sont dépassés tous les modèles HOG.

Chapitre 5 : Détection de la somnolence

5.1. Introduction

Chaque jour, de nombreuses personnes meurent des suites d'accidents de la route. La cause de ces accidents est due à de nombreux facteurs, notamment le sommeil pendant le conduit, et pour éviter ce genre des accidents, il faut que le conducteur soit toujours dans l'état vigilant. Et ça le but général de notre projet. Où nous avons utilisé la mesure PERCLOSE pour détecter le statut de conducteur (sommolence ou Vigilant). L'entrer de notre system est une séquence des trames (vidéo) qui a acquis à partir d'une caméra pi avec 5 méga pixels et 1080p HD vidéo à 30 FPS. Puis, le système analyse chaque image séparément. Ce processus se fait dans 4 étapes principales : Premièrement, on détecte le visage dans chaque image. Après, on détecte les deux yeux (gauche et droit) du visage. Par la suite, on détecte l'état des yeux. Finalement on détecte l'état du conducteur avec le calcul de moyenne de PERCLOSE des deux yeux. Dans ce chapitre nous avons résumé notre système et le tester sur l'ensemble de donnée NTHU [49], Puis, nous l'avons comparé avec le système de Adrian [22].

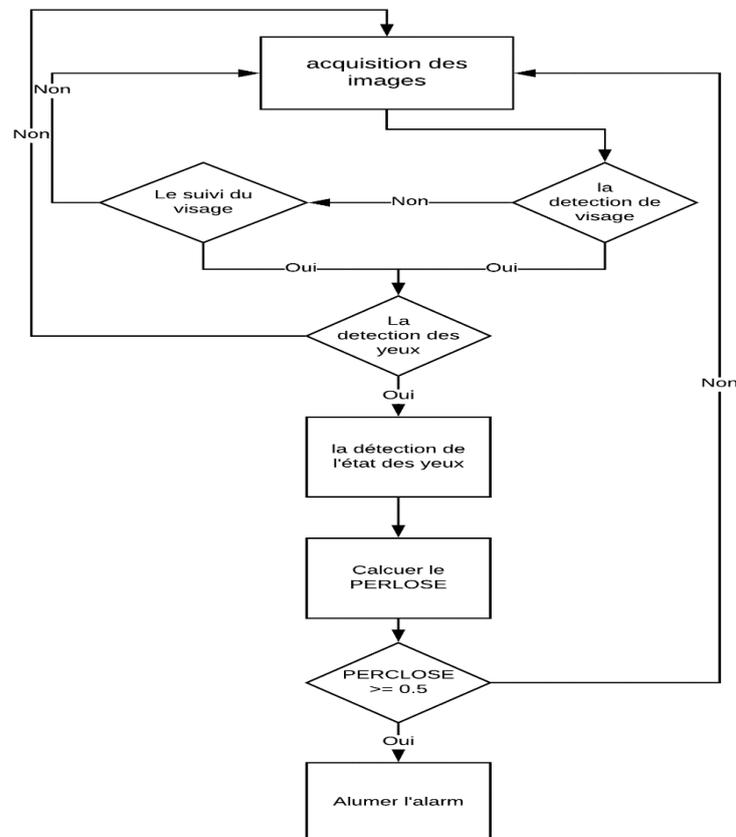


Figure 5.1 : L'organigramme de système proposé.

5.2. Détection et suivi du visage

Dans cette étape, nous avons utilisé les modèles qui nous avons discuté dans le chapitre 2 où nous avons utilisé l'approche de viola-Jones pour détecter le visage dans l'image actuelle et dans le cas où il est échoué dans la détection, nous avons utilisé l'approche de suivi dlib-correlation pour prédire la position de visage à partir de sa position dans l'image précédente.

5.3. Détection des yeux

Nous avons vu dans le chapitre 3 que les modèles viola-Jones à un F1-score plus élevé que les modèles HOG après l'utilisation des géométries de visage pour réduire le nombre de FP. Et aussi, nous avons vu dans le chapitre 2 que l'approche de viola-Jones plus rapide que l'approche HOG. C'est pour ça nous avons choisi les modèles de viola-Jones dans cette étape.

5.4. Etat des yeux

Comprendre les états de l'œil (ouverts et fermés) est une question fondamentale pour une vaste gamme de travaux de recherche appliquées au visage. Dans le système de surveillance du conducteur, la détection des états oculaires est une tâche difficile, car l'apparence des yeux peut être unique pour chaque visage. Afin de résoudre ce problème, nous avons utilisé les caractéristiques de HOG pour extraire les caractéristiques des yeux et le classificateur SVM pour classifier ces caractéristiques.

Dans la formation des modèles, nous avons utilisé l'ensemble de données réalisé dans le chapitre 3 pour trouver les meilleures valeurs des hyper-paramètres C et sigma du hyperplan de SVM dans lesquels sigma détermine la portée d'influence d'un seul échantillon sur la hyperplan de SVM et le paramètre C établit une comparaison entre la classification correcte des échantillon et la maximisation de la marge de la fonction de décision. Il est clair que la meilleure valeur de f1-score est 0.98. Donc nous avons utilisé une de ces modèles. Dans

l'évaluation de ces modèles, nous avons utilisé la même méthode que nous avons utilisée dans les chapitres précédents (matrice de confusions) la seule différence est que :

- TP (True positive) : le système prédire que l'œil ouvert et la classe réelle de l'œil est ouvert
- FP (False positive): le système prédire que l'œil fermé et la classe réelle de l'œil est ouvert
- FN (False négative): le système prédire que l'œil ouvert et la classe réelle de l'œil est fermé
- TN (True négative) : le système prédire que l'œil fermé et la classe réelle de l'œil est fermé

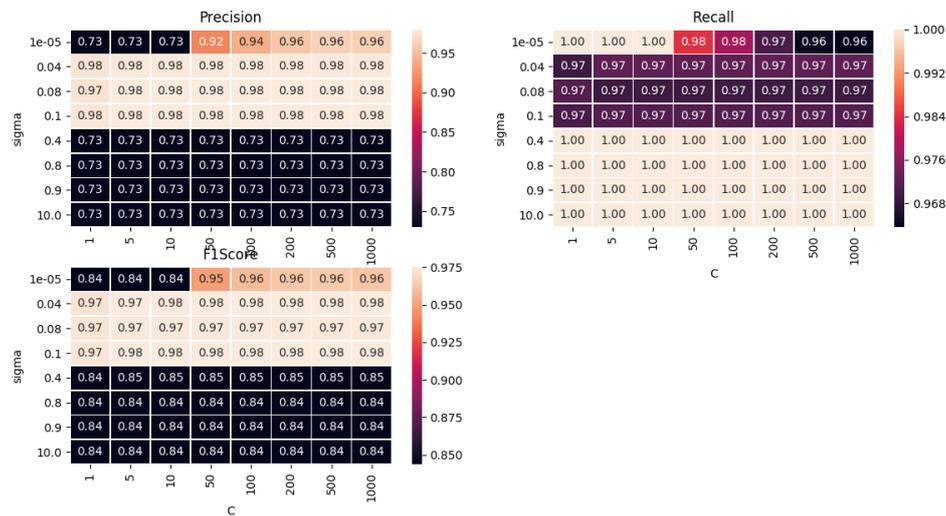


Figure 5.2 : les résultats de test des modèles de détection des yeux sur le 25% de l'ensemble de données de chapitre 3.

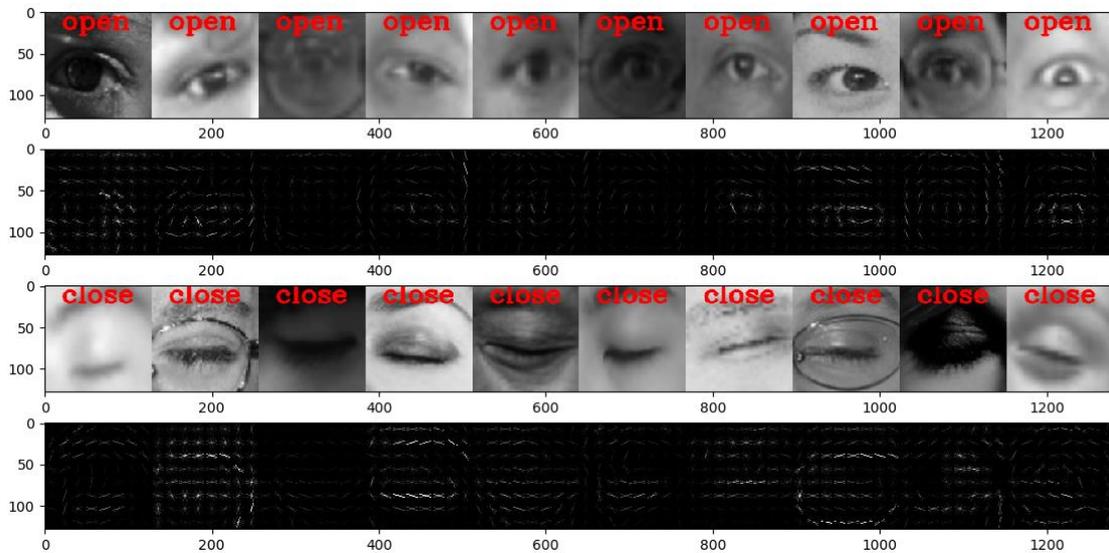


Figure 5.3 : Les résultats de test du modèle avec $\sigma=0.04$ et $C=1000$.

5.5. Pourcentage des yeux Fermés

Après la détection des yeux et son état, nous avons calculé le pourcentage des yeux Fermés dans un intervalle de temps pour détecter l'état de conducteur (somnolence ou vigilant). Ce pourcentage est la Mesure PERCLOSE et son formule est :

$$PERCLOSE = \frac{\text{nombre des yeux fermé}}{\text{nombre des yeux fermé} + \text{nombre des yeux ouvert}} \quad (5.1)$$

Cette formule permet de calculer PERCLOSE pour un œil. À cause de ça, nous avons calculé PERCLOSE de chaque œil puis nous avons pris la moyenne.

5.6. Evaluation

Comme nous avons dit précédemment, dans l'évaluation de notre approche proposée, nous avons utilisé l'ensemble de données NTHU. Cet ensemble de données comprend 20 vidéos avec diverses caractéristiques faciales. Ces vidéos sont divisées en cinq scénarios (avec des lunettes et avec un faible éclairage, sans lunettes et avec un faible éclairage, avec des lunettes et pendant le jour, sans lunettes et pendant le jour et avec lunettes de soleil). Et pour voir le comportement de notre approche proposée, nous avons évalué chaque scénario séparément.

Toujours nous avons utilisé la matrice de confusion dans l'évaluation avec:

- TP (True positive) : le système prédit que le conducteur est sommeil et dans le réel le conducteur est sommeil.
- FP (false positive): le système prédit que le conducteur n'est pas sommeil et dans le réel le conducteur est sommeil.
- FN (false négative): le système prédit que le conducteur est sommeil et dans le réel le conducteur n'est pas sommeil.
- TN (True négative) : le système prédit que le conducteur n'est pas sommeil et dans le réel le conducteur n'est pas sommeil.

5.7. Résultats et discussion

L'objectif de cette section est de présenter les résultats obtenus à partir du test de notre méthode proposé sur l'ensemble de donnée NTHU et de les interpréter tout en faisant une comparaison. Dans ce test nous avons utilisé un intervalle de 60 frames, et les résultats sont présentés dans les tableaux 5.1.

Scénario	TP	FP	TN	FN	Précision	Rappel	F1-Score
lunettes	15622	10053	6762	4824	0.60	0.76	0.67
Faible éclairage	13254	6294	4151	3426	0.67	0.79	0.73
Faible éclairage + lunettes	4545	4979	5779	1148	0.47	0.79	0.59
San lunettes	21327	11013	9592	3077	0.65	0.87	0.75
Lunettes de soleil	5642	3865	2139	3437	0.59	0.62	0.60

Tableau 5.1 : Les résultats de test de notre méthode sur l'ensemble des données NTHU.

Selon les valeurs de tableaux, nous voyons que notre approche est marchée bien dans le scénario de jour et sans les lunettes où on arrive à 0.75 de f1-score. Suivi de scénario de faible éclairage plus les lunettes avec 0.73 de f1-score. Et il vient à la dernière place le scénario de la lunette de soleil avec 0.51 de f1-score. Et Bien que ces valeurs semblent faibles, elles sont légèrement meilleures que les valeurs de la méthode proposé par Adrian [22], où son résultats sont présentées dans le tableau.

Scénario	TP	FP	TN	FN	Précision	Rappel	F1-Score
lunettes	14533	10058	6857	5913	0.59	0.71	0.64
Faible éclairage	12940	4533	5912	3740	0.74	0.77	0.75
Faible éclairage + lunettes	4670	6452	4306	1023	0.41	0.82	0.55
San lunettes	20312	12321	5442	1683	0.62	0.92	0.74
Lunettes de soleil	6475	8899	2558	1276	0.42	0.83	0.56

Tableau 5.2 : les résultats de test de la méthode de Adrian sur l'ensemble de données NTHU.

5.8. Conclusion

Dans ce chapitre, nous avons présenté le système de détection de somnolence que nous avons réalisé dans ce projet. Ce système est basé sur l'état des yeux (ouvert ou fermé) et compose de trois étapes de base: la détection du visage, le positionnement des yeux et enfin la reconnaissance de l'état des yeux.

Il existe plusieurs méthodes de détection du visage et des yeux. L'utilisation de ces méthodes dépend du problème à résoudre. Dans ce chapitre, nous avons expliqué les causes d'utilisation de la méthode de Viola et Jones dans la détection des yeux et du visage.

L'autre point que nous avons traité dans ce chapitre est la détection de l'état des yeux où nous avons utilisé le descripteur de HOG et le classificateur SVM.

La dernière partie de ce chapitre était consacrée à l'évaluation de ce système sur l'ensemble des données NTHU et la discussion de ces résultats.

Conclusion générale

Ce projet nous a permis de découvrir plusieurs méthodes de différents types qui permettent de détecter et surveiller l'état du conducteur pendant la conduite. Parmi ces méthodes, on peut distinguer des méthodes intrusives qui influencent les performances de conducteur et des méthodes non intrusives qui n'influence pas sur le conducteur.

Nous avons réalisé un système non intrusif qui surveille l'état des yeux pour détecter l'état du conducteur.

Ce système éventuellement être peu coûteux, car seulement trois équipements physiques est nécessaires dans la réalisation de ce système : une caméra, un ordinateur et une alarme.

Dans ce projet, certains aspects peuvent être améliorés. Après avoir testé notre système dans le cas où quelqu'un portant des lunettes, nous avons constaté que le système ne surveille pas bien l'état de conducteur.

Un autre aspect qu'on veut améliorer est le système de capture d'images où la caméra utilisée dans notre système est une caméra pi avec 5 méga pixels qui ne supporte pas l'infrarouge, donc, elle affecte négativement sur la précision de notre système lors de l'éclairage faible.

• **Perspectives :**

Plusieurs parties du projet nécessitent une série de développements, y compris l'analyse de la direction du regard, la détection de l'état de la tête, où la direction du regard et l'état de la tête varient selon l'attention du conducteur. Le traitement de ces problèmes fait partie de l'étude de détection de fatigue et de somnolence. Chacune de ces parties pourraient être des sujets de recherche. Plusieurs équipes s'affairent à trouver des solutions à ces problèmes.

Il y a d'autres perspectives, on peut citer par exemple:

- un système basé sur l'apprentissage en profondeur pour la détection de la somnolence et de la distraction du conducteur :

Une approche basée sur l'état des yeux du conducteur et la surveillance de la pose de la tête dans un flux vidéo. Cette approche doit être efficace pour localiser les yeux du conducteur, reconnaître leur état et détecter les différentes postures, pour estimer la distraction cognitive du conducteur.

- un système hybride de détection de somnolence :

Un système basé sur l'état des yeux, le bâillement et la tête pour surveiller l'état du conducteur. Ce système doit être efficace pour localiser le visage, les yeux, le bâillement et reconnaître leurs états.

Bibliographie

- [1] C. Solomon and T. Breckon, *Fundamentals of digital Image Processing*, Hoboken: wiley blackwell, 2011.
- [2] L. Diane, *cours de traitement d'images*, nice: Laboratoire I3S, 2004, p. 10.
- [3] G. Peyré., "le traitement numérique des images," *Archive ouverte HAL*, 2011.
- [4] M. Fatiha, "Détection et Suivi d'Objets en Mouvement Dans Une Séquence d'Images," *Université Mohamed Boudiaf , Oran*, 2011.
- [5] H. KAMELI and H. KALLAL, "Segmentation d'images satellitaire par Cuckoo," *UNIVERSITE ABDELHAMID IBN BADIS, MOSTAGANEM*, 2012.
- [6] "ROAD SAFETY FACTS," ASIRT, 2018. [Online]. Available: <https://www.asirt.org/safe-travel/road-safety-facts/>. [Accessed 1 5 2019].
- [7] " إحصاءات حوادث المرور المسجلة خلال سنة 2014," CNPSR, alger, 2014.
- [8] M. Bougueroua and L. Carnis, "Insécurité routière et développement économique : analyse du cas algérien," *RTS - Recherche Transports Sécurité*, vol. 2018, p. 13p, Apr 2018.
- [9] National Center for Statistics and Analysis, "Drowsy Driving 2015," Washington, 2017.
- [10] J. Krajewski, D. Sommer, U. Trutschel, D. Edwards and M. Golz, "Steering Wheel Behavior Based Estimation of Fatigue," *Driving assessment conference*, 2009.
- [11] P. Winston, Director, *Learning: Support Vector Machines*. [Film]. USA: Massachusetts Institute of Technology, 2010.
- [12] z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Annals of translational Medicine*, vol. 4, p. 218, 2016.
- [13] M. Inger, T. Åkerstedt, B. Peters, A. Anund and G. Kecklund, "Subjective sleepiness, simulated driving performance and blink duration: Examining individual differences," *Journal of sleep research*, vol. 15, pp. {47-53, 2006.
- [14] M. Gillberg, G. Kecklund and T. Åkerstedt, " Relations between performance and subjective ratings of sleepiness during a night awake," *Sleep: Journal of Sleep Research & Sleep Medicine*, vol. 17, pp. 236-241, 1994.
- [15] S. Hu and G. Zheng, "Driver Drowsiness Detection with Eyelid related Parameters by Support Vector Machine," *Expert Systems with Applications*, vol. 36, pp. 7651-7658, 2009.

- [16] P. Antoine, C. Sylvie and A. Caplier, "EOG-based drowsiness detection: Comparison between a fuzzy system and two supervised learning classifiers," *The International Federation of Automatic Control*, 2011.
- [17] M. Patel, S. K. Lal, D. Kavanagh and P. Rossiter, "Applying neural network analysis on heart rate variability data to assess driver fatigue," *Expert Systems with Applications*, vol. 38, p. 7235–7242, 2011.
- [18] S. Arun, S. Kenneth and M. Murugappan, "Detecting Driver Drowsiness Based on Sensors: A Review," *Sensors (Basel, Switzerland)*, vol. 12, pp. 16937-16953, 2012.
- [19] J. May and C. Baldwin, "Driver fatigue: The importance of identifying causal factors of fatigue when considering detection and countermeasure technologies," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 12, pp. 218-224, 2009.
- [20] I. García, S. Bronte, L. M. Bergasa, J. Almazán and J. Yebes, "Vision-based drowsiness detector for real driving conditions," *IEEE Intelligent Vehicles Symposium*, pp. 618-623, 2012.
- [21] W. Zhang, B. Cheng and Y. Lin, "Driver drowsiness recognition based on computer vision technology," *Tsinghua Science and Technology*, vol. 17, pp. 354-362, 2012.
- [22] A. Rosebrock, "Drowsiness detection with OpenCV," 08 May 2018. [Online]. Available: <https://www.pyimagesearch.com/2017/05/08/drowsiness-detection-opencv/>. [Accessed 29 May 2019].
- [23] G. Reza, G. Marnim and A. Vassilis, "A Realistic Dataset and Baseline Temporal Model for Early Drowsiness," *CoRR*, vol. abs/1904.07312, 2019.
- [24] M. Ochocki and D. Sawicki, "Yawning Recognition based on Dynamic Analysis and Simple Measure," in *Proceedings of the International Conference on Computer-Human Interaction Research and Applications*, in Funchal, Madeira, Portugal, pp. 111-117, 2017.
- [25] M. Omidyeganeh, S. Shirmohammadi, S. Abtahi, A. Khurshid, M. Farhan, J. Scharcanski, B. Hariri, D. Laroche and L. Martel, "Yawning Detection Using Embedded Smart Cameras," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, pp. 1-13, 2016.
- [26] T. H. Lam and N. T. Do, "Algorithm to detect driver's drowsiness based on nodding behavior," *International Journal of Soft Computing, Mathematics and Control (IJSCMC)*, vol. 5, 2016.
- [27] M. Omidyeganeh, A. Javadtalab and S. Shirmohammadi, "Intelligent driver drowsiness detection through fusion of yawning and eye closure," *IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurement Systems Proceedings*, pp. 1-6, 2011.
- [28] V. Esra, C. Mujdat, E. Aytul, L. Gwen, B. Marian and M. Javier, "Automated Drowsiness Detection For Improved Driving Safety," *International Conference on Automotive Technologies*, 2008.

- [29] B. Reddy, Y. Kim, S. Yun, C. Seo and J. Jang, "Real-Time Driver Drowsiness Detection for Embedded System Using Model Compression of Deep Neural Networks," 2017.
- [30] G. Nisha, N. Tejas and Y. Aditya, "Drowsy Driver Detection in Video Sequences using LSTM with CNN Features," 31 Des 2017. [Online]. Available: <https://github.com/nishagandhi/DrowsyDriverDetection>. [Accessed 21 July 2019].
- [31] B. Djamel Eddine, "Hypovigilance Detection and Assistance to Vehicle Drivers," UNIVERSITE BADJI MOKHTAR, ANNABA, 2018.
- [32] S. Abtahi, M. Omidyeganeh, S. Shirmohammadi and B. Hariri, "YawDD: A Yawning Detection Dataset," in *Proceedings of the 5th ACM Multimedia Systems Conference*, Singapore, Singapore, 2014.
- [33] P. Viola and M. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision*, vol. 57, pp. 137--154, 01 May 2001.
- [34] D. Mathew and Kurian, "Face Detection in Real Time Based on HOG Rekha N," in *International Journal of Advanced Research in Computer Engineering & Technology*, 2014.
- [35] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," *International Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 1886-893, jun 2005.
- [36] C. Zhaowei and V. Nuno, "Cascade R-CNN High Quality Object Detection and Instance Segmentation," *CoRR*, vol. abs/1906.09756, 2019.
- [37] R. Shaoqing, H. Kaiming, B. G. Ross and S. Jian, "Faster R-CNN Towards Real-Time Object Detection with Region Proposal Networks," *Computing Research Repository*, vol. abs/1506.01497, 2015.
- [38] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu and A. C. Berg, "SSD: Single Shot MultiBox Detector," *CoRR*, vol. abs/1512.02325, pp. 21--37, 2015.
- [39] J. Vidit and L. M. Erik, "FDDB: A Benchmark for Face Detection in Unconstrained Settings," University of Massachusetts, Amherst, 2010.
- [40] b. Gary, "Computer Vision Face Tracking For Use in a Perceptual User," *Computer Vision Face Tracking For Use in a Perceptual User Interface*, 1998.
- [41] D. Riahi, "Suivi multi-objets par la detection Application à la vidéo surveillance," Université de montréal, 2016, pp. 3-4.
- [42] D. S. Bolme, J. R. Beveridge, B. A. Draper and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010.
- [43] M. Danelljan, G. Häger, F. Khan and M. Felsberg, "Accurate Scale Estimation for Robust Visual Tracking," in *British Machine Vision Conference*, Nottingham, September 1-5, 2014.

- [44] F. Song, X. Tan, X. Liu and S. Chen, "Eyes Closeness Detection from Still Images with Multi-scale Histograms of Principal Oriented Gradient," *Pattern Recognitions*, 2014.
- [45] V. Le, J. Brandt, Z. Lin, L. Bourde and T. S. Huang, "Interactive Facial Feature Localization," *In ECCV*, 2012.
- [46] J. Oliver, K. Klaus J and F. Robert W, "Robust Face Detection," *In Int. Conf. on Audio- and Video-based Biometric Person Authentication*, pp. 90-95, 2001.
- [47] Z. Jianming, M. Shugao, S. Mehrnoosh, S. Stan, B. Margrit, L. Zhe, S. Xiaohui, P. Brian and M. Radomír, "Multi-Salient-Object (MSO)," 2015.
- [48] Y. L. Tsung, M. Michael, J. B. Serge, D. B. Lubomir, B. G. Ross, H. James, P. Pietro, R. Deva, P. Dollar and L. Z. C, "Microsoft COCO: Common Objects in Context," *CoRR*, vol. abs/1405.0312, 2014.
- [49] W. Ching-Hua, L. Ying-Hsiu and L. Shang-Hong, "Driver Drowsiness Detection via a Hierarchical Temporal Deep Belief Network," in *Asian Conference on Computer Vision Workshop on Driver Drowsiness Detection from Video*, Taipei, Taiwan, Nov.2016.