Université Mohammed V
Rabat

Faculté des Sciences
Rabat

Normandie Université

INSA ROUEN

Thèse de Doctorat
En cotutelle entre l'Université Mohammed V et l'INSA de Rouen

Présentée par:

Nawal ALIOUA

Discipline: Sciences de l'ingénieur
Spécialité: Informatique et Télécommunications

Extraction et analyse des caractéristiques faciales: Application à l'hypovigilance chez le conducteur

Soutenue le 28 mars 2015 devant la commission d'examen

Président :

M. Driss ABOUTAJDINE	Professeur de l'enseignement supérieur, Faculté des Sciences de Rabat, Maroc

Examinateurs :

M. Ahmed HAMMOUCH	Professeur de l'enseignement supérieur, Ecole Normale Supérieure de l'Enseignement Technique de Rabat, Maroc
M. Abdelaziz BENSRHAIR	Professeur des universités, Institut National des Sciences Appliquées de Rouen, France
M. Fabrice MERIAUDEAU	Professeur des universités, Institut Universitaire de Technologie du Creusot, France
M. Mohammed RZIZA	Professeur habilité à diriger les recherches, Faculté des Sciences de Rabat, Maroc
Mlle. Aouatif AMINE	Professeur habilité à diriger les recherches, Ecole Nationale des Sciences Appliquées, Kénitra, Maroc
M. Fawzi NASHASHIBI	Docteur habilité à diriger les recherches, Institut National de Recherche en Informatique et en Automatique, Paris-Rocquencourt, France

---

Travaux préparés en co-tutelle aux laboratoires : LRIT-CNRST (URAC-29), Faculté des Sciences de Rabat & LITIS, Institut National des Sciences Appliquées de Rouen, France

Avant-Propos

Les travaux présentés dans ce mémoire ont été effectués au Laboratoire de Recherche en Informatique et Télécommunications (LRIT-CNRST 29) de la Faculté des Sciences de Rabat (FSR), Université Mohammed V au Maroc sous la direction du Professeur **Mohammed RZIZA** et le co-encadrement du Professeur **Aouatif AMINE**, et au Laboratoire d'Informatique, du Traitement de l'Information et des Systèmes (LITIS), Institut National des Sciences Appliquées de Rouen (INSA de Rouen) en France sous la direction du Professeur **Abdelaziz BENSRHAIR** et le co-encadrement du Professeur **Alexandrina ROGOZAN** dans le cadre d'une thèse en cotutelle pour l'obtention du Doctorat National Marocain et du Doctorat de Normandie Université délivré par l'INSA de Rouen

Je commence par présenter ma plus vive gratitude à mon Directeur de thèse M. **Mohammed RZIZA**, Professeur Habilité à la FSR. Grâce à ses encouragements, sa pédagogie et ses précieux conseils, il a su me guider pour mener à bien mes travaux de recherche. J'exprime ici ma profonde gratitude à son égard et l'estime respectueuse que je lui porte.

Je tiens à exprimer mes remerciements à ma co-encadrante, Mlle. **Aouatif AMINE**, Professeur Habilité à l'Ecole Nationale des Sciences Appliquées (ENSA de Kénitra), pour ces années de soutien, pour ses précieux conseils scientifiques et humains, ainsi que pour ses encouragement.

Je veux exprimer toute ma reconnaissance et ma gratitude à mon directeur de thèse en France, Mr. **Abdelaziz BENSRHAIR** Professeur des Universités l'INSA de Rouen et co-directeur du LITIS. Je le remercie de m'avoir intégrée au sein de l'équipe Systèmes de Transport Intelligents (STI) et de m'avoir consacré tout ce temps et toute cette énergie malgré son emploi du temps très chargé. Sa gentillesse, ses encouragements et ses conseils m'ont permis de mener ce travail à son terme.

Je souhaite remercier profondément ma co-encadrante en France Mme. **Alexandrina ROGOZAN**, maître de conférence à l'INSA de Rouen, de m'avoir procuré ses conseils et ses idées scientifiques même quand de grandes distances nous séparaient. Sa motivation et son enthousiasme pour mener à bien ce travail m'ont été d'une très grande aide.

Je tiens à remercier Mr. **Driss ABOUTAJDINE**, Professeur de l'enseignement supérieur à la FSR, directeur du Centre National pour la Recherche Scientifique et Technique (CNRST) et du laboratoire LRIT-CNRST 29 d'avoir accepté de présider le jury de ma thèse. Je souhaite également lui présenter ma plus grande gratitude pour m'avoir intégrée au sein de son laboratoire et

d'avoir toujours été présent pour tous ses doctorants malgré ses occupations.

Mes remerciements vont à Mr. **Fabrice MERIAUDEAU**, Professeur des universités à l'Institut Universitaire de Technologie (IUT) du Creusot et directeur du laboratoire LE2I - UMR CNRS 6306 en France, pour avoir accepté de rapporter ce travail et de participer au jury. Je tiens également à remercier Mr. **Ahmed HAMMOUCH**, Professeur de l'enseignement supérieur à l'Ecole Normale Supérieure de l'Enseignement Technique de Rabat, Maroc, pour avoir accepté de rapporter ce travail et de participer au jury. Enfin, je voudrais remercier Mr. **Fawzi NASHASHIBI**, Docteur habilité à diriger les recherches et directeur de recherche équipe-projet RITS à l'Institut National de Recherche en Informatique et en Automatique (INRIA), Paris-Rocquencourt en France, pour avoir accepté d'examiner ce travail.

Au cour de cette thèse, j'ai bénéficié d'une bourse d'excellence octroyée par le CNRST dans le cadre du programme des bourses de recherche initié par le ministère de l'éducation national de l'enseignement supérieur, de la recherche scientifique et de la formation des cadres. J'ai également bénéficié d'un financement octroyé par l'ambassade de France à Rabat pour deux stages d'une durée d'un mois puis de deux mois à l'INSA de Rouen. Enfin, l'INSA de Rouen m'a accordé un financement d'une durée de huit mois pour mener à bien ma thèse pendant mon séjour en France.

Tout au long de ces années de thèse, j'ai eu l'occasion de rencontrer des personnes toutes aussi intéressantes les unes que les autres. A leur façon, ils ont tous contribué à mon apprentissage. Bien que je sois reconnaissante envers chacune de ces personnes, certaines d'entre elles méritent un remerciement particulier. Je tiens à remercier tous mes collègues du laboratoire LRIT-CNRST 29, particulièrement : Fadoua, Fadwa, Maryam, Meryem, Awatif, Safae, Chouaib, Said et Zakaria. Je remercie également les membres du LITIS, particulièrement Nadine, Abir, Alina, Yadu, Samia, Brigitte et Sandra.

Je termine par les personnes que je ne saurais jamais remercier assez. A mon père, ma mère, ma belle-mère, ma grand-mère et mon frère qui m'ont soutenus et encouragés à aller de l'avant. Je remercie aussi ma famille et ma belle famille pour leur soutien tout au long de mes études. Enfin, aucun mot ne peut exprimer toute la reconnaissance et la gratitude que je dois à mon bien aimé mari Othmane. Je n'aurais jamais tenu le coup sans ton aide.

Résumé

L'étude des caractéristiques faciales a suscité l'intérêt croissant de la communauté scientifique et des industriels. En effet, ces caractéristiques véhiculent des informations non verbales qui jouent un rôle clé dans la communication entre les hommes. De plus, elles sont très utiles pour permettre une interaction entre l'homme et la machine. De ce fait, l'étude automatique des caractéristiques faciales constitue une tâche primordiale pour diverses applications telles que les interfaces homme-machine, la science du comportement, la pratique clinique et la surveillance de l'état du conducteur. Dans cette thèse, nous nous intéressons à la surveillance de l'état du conducteur à travers l'analyse de ses caractéristiques faciales. Cette problématique sollicite un intérêt universel causé par le nombre croissant des accidents routiers, dont une grande partie est provoquée par une dégradation de la vigilance du conducteur, connue sous le nom de l'hypovigilance. En effet, nous pouvons distinguer trois états d'hypovigilance. Le premier, et le plus critique, est la somnolence qui se manifeste par une incapacité à se maintenir éveillé et se caractérise par les périodes de micro-sommeil correspondant à des endormissements de 2 à 6 secondes. Le second est la fatigue qui se définit par la difficulté croissante à maintenir une tâche à terme et se caractérise par une augmentation du nombre de bâillements. Le troisième est l'inattention qui se produit lorsque l'attention est détournée de l'activité de conduite et se caractérise par le maintien de la pose de la tête en une direction autre que frontale.

L'objectif de cette thèse est de concevoir des approches permettant de détecter l'hypovigilance chez le conducteur en analysant ses caractéristiques faciales. En premier lieu, nous avons proposé une approche dédiée à la détection de la somnolence à partir de l'identification des périodes de micro-sommeil à travers l'analyse des yeux. En second lieu, nous avons introduit une approche permettant de relever la fatigue à partir de l'analyse de la bouche afin de détecter les bâillements. Du fait qu'il n'existe aucune base de données publique dédiée à la détection de l'hypovigilance, nous avons acquis et annoté notre propre base de données représentant différents sujets simulant des états d'hypovigilance sous des conditions d'éclairage réelles afin d'évaluer les performances de ces deux approches. En troisième lieu, nous avons développé deux nouveaux estimateurs de la pose de la tête pour permettre à la fois de détecter l'inattention du conducteur et de déterminer son état, même quand ses caractéristiques faciales (yeux et bouche) ne peuvent être analysées suite à des positions non-frontales de la tête. Nous avons évalué ces deux estimateurs sur la base de données publique Pointing'04. Ensuite, nous avons acquis et annoté une base de données représentant la variation de la pose de la tête du conducteur pour valider nos estimateurs sous un environnement de conduite.

Mots clés : Hypovigilance chez le conducteur ; Analyse des caractéristiques faciales ; Estimation de la pose de la tête ; Pyramide orientable ; Apprentissage probabiliste ; Transformée de Hough Circulaire ; SVM

Abstract

Studying facial features has attracted increasing attention in both academic and industrial communities. Indeed, these features convey nonverbal information that plays a key role in human communication. Moreover, they are very useful to allow human-machine interactions. Therefore, the automatic study of facial features is an important task for various applications including robotics, human-machine interfaces, behavioral science, clinical practice and monitoring driver state.

In this thesis, we focus our attention on monitoring driver state through its facial features analysis. This problematic solicits a universal interest caused by the increasing number of road accidents, principally induced by deterioration in the driver vigilance level, known as hypovigilance. Indeed, we can distinguish three hypovigilance states. The first and most critical one is drowsiness, which is manifested by an inability to keep awake and it is characterized by microsleep intervals of 2-6 seconds. The second one is fatigue, which is defined by the increasing difficulty of maintaining a task and it is characterized by an important number of yawns. The third and last one is the inattention that occurs when the attention is diverted from the driving activity and it is characterized by maintaining the head pose in a non-frontal direction.

The aim of this thesis is to propose facial features based approaches allowing to identify driver hypovigilance. The first approach was proposed to detect drowsiness by identifying microsleep intervals through eye state analysis. The second one was developed to identify fatigue by detecting yawning through mouth analysis. Since no public hypovigilance database is available, we have acquired and annotated our own database representing different subjects simulating hypovigilance under real lighting conditions to evaluate the performance of these two approaches. Next, we have developed two driver head pose estimation approaches to detect its inattention and also to determine its vigilance level even if the facial features (eyes and mouth) cannot be analyzed because of non-frontal head positions. We evaluated these two estimators on the public database Pointing'04. Then, we have acquired and annotated a driver head pose database to evaluate our estimators in real driving conditions.

Keywords : Driver hypovigilance ; Faciale features analysis ; Head pose estimation ; Steerable pyramid ; Probabilistic learning ; Circular Hough Transform ; SVM

Liste des acronymes

AAM	modèles actifs d'apparence « Active Appearance Model »
ASFA	Association des Sociétés Françaises d'Autoroutes
BF	BestFirst
CARRS-Q	Center of Accident Research and Road Safety-Queensland
CCR	taux de Bonne Classification « Correct Classification Rate »
CFS	CorrelationFeatureSelection
CHT	Transformée de Hough Circulaire « Circular Hough Transform »
CV	validation croisée « Cross Validation »
DAC	Driver Alert Control
DF-SVM	Descriptors Fusion-SVM
ECG	électrocardiographie
ECT	Effective Control Transport
EEG	électroencéphalographie
EM	Expectation-Maximisation
FN	Faux Négatif
FP	Faux Positif
GPU	processeur graphique « Graphics Processing Unit »
GR	GainRatio
GS	GreedyStepwise
HMM	modèles de Markov cachés « Hidden Markov Models »
HOG	histogrammes des gradients orientés « Histograms of Oriented Gradients »
HT	transformée de Hough « Hough Transform »

IG	InformationsGain
LAAM	mémoire auto-associative linéaire « Linear Auto-associative Memory »
LARR	régression robuste localement ajustée « Locally Adjusted Robust Regressor »
LDA	analyse discriminante linéaire « Linear Discriminant Analysis »
LGO	histogramme des orientations des gradients localisés « Localized Gradient Orientation »
LoG	laplacien de gaussienne « Laplacian of Gaussian »
LPF	fonction de vraisemblance paramétrique « Likelihood Parametrized Function »
MAE	erreur angulaire moyenne « Mean Absolute Error »
ms	millisecondes
NHTSA	National Highway Traffic Safety Administration
OneR	OneRule
PCA	analyse en composantes principales « Principal Component Analysis »
PERCLOS	pourcentage de fermeture de l'œil en fonction du temps « Percentage of Eye Closure »
PLS	moindre carrée partiel « Partial Least Square »
POSIT	Pose from Orthography and Scaling with ITerations
PPG	photopléthysmogramme
RMS	racine de la moyenne du carré « Root Mean Square »
SF	filtres orientables « Steerable Filters »
SIFT	Scale Invariant Feature Transform
SP	pyramide orientable « Steerable Pyramid »
SURF	Speeded-Up Robust Features
SVM	machines à vecteurs de support « Support Vector Machine »
SVR	machines à vecteurs supports « Support Vector Regression »
RBF	fonction de base radiale « Radial Basis Function »
RF	ReliefF
Rk	Ranker

RP	projection aléatoire « Random Projection »
VN	Vrai Négatif
VP	Vrai Positif

Table des matières

Résumé iii

Abstract v

Liste des acronymes vii

Liste des figures xvi

Liste des tableaux 1

Chapitre 1 : Introduction générale 1
- 1.1 Contexte général 1
- 1.2 Problématique 2
- 1.3 Motivation 2
- 1.4 Objectif de la thèse 4
- 1.5 Hypovigilance chez le conducteur : état de l'art 5
- 1.6 Organisation de la thèse 11
- 1.7 Liste des publications 12

Partie I Analyse des caractéristiques faciales pour la détection de la fatigue et de la somnolence 13

Chapitre 2 : Analyse des caractéristiques faciales pour la détection de la somnolence et de la fatigue : état de l'art 15
- 2.1 Introduction 15
- 2.2 Détection du visage 16
- 2.3 État d'ouverture/fermeture des yeux 21
- 2.4 PERCLOS 23
- 2.5 Fréquence de clignement des yeux 23
- 2.6 Fréquence de bâillement 24
- 2.7 Conclusion 25

Chapitre 3 : Détection de la somnolence et de la fatigue basée sur la Transformée de Hough Circulaire 27

3.1 Introduction 27
3.2 Localisation des zones d'intérêt 28
3.3 Détection de la somnolence par l'analyse des yeux 32
3.4 Détection de la fatigue par l'analyse de la bouche 39
3.5 Schéma général du système 42
3.6 Résultats expérimentaux 44
3.7 Conclusion 53

Partie II Estimation de la pose de la tête du conducteur pour la détection de l'inattention 55

Chapitre 4 : Estimation de la pose de la tête et détection de l'inattention : état de l'art 57
4.1 Estimation de la pose de la tête 57
4.2 Estimation de la pose de la tête du conducteur 71
4.3 Conclusion 74

Chapitre 5 : Estimation de la pose de la tête basée sur la pyramide orientable et l'apprentissage probabiliste 77
5.1 Introduction 77
5.2 Modélisation de la pose de la tête par la pyramide orientable 78
5.3 Estimation de la pose de la tête par la fonction de vraisemblance paramétrique . 83
5.4 Estimation de la pose de la tête du conducteur par la pyramide orientable et la fonction de vraisemblance paramétrique 84
5.5 Résultats expérimentaux 86
5.6 Conclusion 92

Chapitre 6 : Estimation de la pose de la tête basée sur la classification et la fusion de descripteurs 95
6.1 Introduction 95
6.2 Vecteur caractéristique basé sur la fusion de descripteurs 96
6.3 Estimation de la pose de la tête du conducteur par des SVM multi-classes 100
6.4 Résultats expérimentaux 101
6.5 Conclusion 105

Conclusion et perspectives 107

Liste des publications 111

Bibliographie 113

Liste des figures

1.1 Nombre d'accidents à un seul véhicule pour cent au Maroc (2007-2011) (Benjelloun, 2013) . . . 3
1.2 Nombre de décès pour cent qui résulte des accidents à un seul véhicule au Maroc (2007-2011) (Benjelloun, 2013) . . . 3
1.3 Système d'acquisition et de transmission des signaux ECG et PPG à partir de capteurs placés sur le volant (Shin *et al.*, 2010) . . . 6
1.4 Capteur EEG placé sur la tête du conducteur (Khan et Aadil, 2012) . . . 7
1.5 Signal visuel émis par le système DAC en cas d'hypovigilance . . . 8
1.6 Signal visuel émis par le système Attention Assist en cas d'hypovigilance . . . 8
1.7 Système Driver's Mate de la société ECT . . . 10
1.8 Système multi-caméras Smart Eye Pro 5.10 . . . 10

2.1 Effet de la lumière infrarouge sur l'œil. (a) Effet sombre ; (b) Effet brillant (Ji et Yang, 2004) . . . 16
2.2 Hyperplan séparateur optimal qui maximise la marge dans l'espace de redescription. Les échantillons entourés correspondent aux vecteurs supports . . . 19
2.3 Extraction de l'état de l'œil (Zhang *et al.*, 2008). (a) Image de l'œil ; (b) Binarisation ; (c) Raffinement ; (d) Contour fin ; (e) Distance entre les paupières inférieure et supérieure . . . 22

3.1 Application de fdlib sur une frame réelle . . . 29
3.2 Étapes de la localisation de la limite inférieure et de la limite supérieure des yeux. (a) Image Gradient ; (b) Projection horizontale ; (c) Projection horizontale lissée ; (d) Projection horizontale traitée ; (e) Niveaux des yeux ; (f) Zone des yeux. . . . 29
3.3 Gradient de la zone des yeux . . . 30
3.4 Zones de recherche de l'œil gauche et de l'œil droit . . . 31
3.5 Projection verticale de la zone de recherche de l'œil gauche . . . 31
3.6 œil gauche localisé . . . 31
3.7 œil droit localisé . . . 31
3.8 Zone de recherche de la bouche . . . 32
3.9 Zone de la bouche limitée à gauche et à droite . . . 32
3.10 Bouche localisée . . . 32
3.11 Illustration de la détection du centre d'un cercle par la CHT . . . 35
3.12 Application des détecteurs de contours standards sur un œil ouvert . . . 35

3.13 Application des détecteurs de contours standards sur un œil fermé 35
3.14 Morphologie de l'œil 36
3.15 Illustration des contours gauche et droit de l'œil ouvert 37
3.16 Contours de l'iris par le détecteur proposé 37
3.17 Algorithme de la détection de l'iris par la CHT 38
3.18 Détection de l'iris par la CHT 38
3.19 Application des détecteurs de contours standards sur des bouches fermées, peu ouvertes et grandes ouvertes 40
3.20 Structure de la bouche lors du bâillement 40
3.21 Illustration des contours supérieur et inférieur de la bouche lors du bâillement . . 41
3.22 Contours du bâillement par le détecteur proposé 42
3.23 Détection de la grande ouverture de la bouche par la CHT 42
3.24 Schéma général de la détection de la fatigue et de la somnolence chez le conducteur 43
3.25 Les 18 séquences de la base de données personnelle pour détecter la fatigue et la somnolence 46
3.26 Web caméra avec un éclairage intégré 47
3.27 VP, FP, VN et FN par la méthode de l'analyse de l'œil 48
3.28 VP, FP, VN et FN par la méthode de l'analyse de la bouche 49
3.29 Acquisition des séquences de test dans une voiture 49
3.30 Résultats de l'analyse de l'œil par le système de détection de la fatigue et de la somnolence sous la lumière ambiante du jour 50
3.31 Résultats de l'analyse de la bouche par le système de détection de la fatigue et de la somnolence sous la lumière ambiante du jour 50
3.32 Résultats de l'analyse de l'œil par le système de détection de la fatigue et de la somnolence sous un éclairage artificiel pendant la nuit 52
3.33 Résultats de l'analyse de la bouche par le système de détection de la fatigue et de la somnolence sous un éclairage artificiel pendant la nuit 52

4.1 Représentation de la pose de la tête par trois degrés de liberté 58
4.2 Résultats de l'estimation de la pose de la tête par l'approche AAM + POSIT pour (a) le pitch, (b) le yaw et (c) le roll (Martins et Batista, 2008) 61
4.3 Résultats du suivi par le filtrage particulaire sur des points caractéristiques d'un modèle cylindrique de la tête (Aggarwal *et al.*, 2005), affichés par la grille cylindrique. Les 3-uplets correspondent aux orientations estimées (roll,pitch,yaw) . . . 62
4.4 Application d'un modèle cylindrique (ligne 1) et un modèle ellipsoïdal (ligne 1) sur la tête (Choi et Kim, 2009) 63
4.5 (a) CCR obtenus par l'application des arbres de décision combinés aux caractéristiques de la symétrie faciale sur la base FacePix ; (b) Frame de l'ensemble des données de l'université de Boston (en haut) ; graphe représentant l'estimation de la pose selon le yaw et la vérité terrain (en bas) (Dahmane *et al.*, 2012) 65
4.6 Le ratio de la similarité de la pose en variant la pose de la tête et l'orientation des filtres de Gabor. (a) variation du pitch avec yaw fixé à 90° ; (b) variation du yaw avec pitch fixé à 90° (Sherrah *et al.*, 2001) 67

4.7 Le ratio de la similarité de la pose en variant la pose de la tête et les dimensions du PCA. (a) : variation du pitch ; (b) : variation du yaw (Sherrah *et al.*, 2001) . 67
4.8 (En haut) Le testbed LISA-P utilisé par Murphy-Chutorian *et al.* (2007) pour la collecte et l'évaluation de leur estimateur de la pose de la tête ; (En bas) Vue du conducteur par le testbed LISA-P. 72
4.9 (Gauche) Modèle anthropométrique 3D utilisé pour le suivi. (Droite) Exemple du modèle retourné par l'algorithme du suivi (Murphy-Chutorian et Trivedi, 2010) . 73
4.10 L'unité d'acquisition Smart Eye AntiSleep composée d'une seule camera et de deux sources de lumières infrarouges (Bretzner et Krantz, 2005) 73

5.1 Application des SF choisis sur l'image d'un disque (Freeman et Adelson, 1991). (a) Image du disque ; (b) $f_1^{0^\circ}$; (c) $f_1^{90^\circ}$; (d) $f_1^{30^\circ}$; (e) $R_1^{0^\circ}$; (f) $R_1^{90^\circ}$; (g) R_1^{θ} avec $\theta = 30$. 80
5.2 Décomposition d'une image par la SP (Simoncelli *et al.*, 1992) 81
5.3 Poses frontales de la base Pointing'04 86
5.4 Temps d'exécution pour une image de test en variant nb_{filt} 88
5.5 Temps d'exécution pour une image de test en variant *level* et en considérant ($nb_{filt} = 3$, $step = 60^\circ$) 90
5.6 Acquisition de la séquence du conducteur 91
5.7 Frames du conducteur. (a) tête frontale (Pitch et Yaw) ; (b) Profil gauche (Yaw) ; (c) Profil droit (Yaw) ; (d) Tête haute (Pitch) ; (e) Tête basse (Pitch) 92

6.1 Dérivées partielles discrétisées $L_{yy}(X, \sigma)$ et $L_{xy}(X, \sigma)$ 98
6.2 Approximation de $L_{yy}(X, \sigma)$ et $L_{xy}(X, \sigma)$ par des filtres carrés 98

Liste des tableaux

2.1 Tableau récapitulatif des approches de détection de la somnolence et de la fatigue chez le conducteur . . . 26

3.1 Matrice de confusion de l'analyse de l'œil . . . 44
3.2 Matrice de confusion de l'analyse de la bouche . . . 45
3.3 Interprétation du coefficient κ . . . 46
3.4 Résultats de l'évaluation de l'analyse de l'œil basée sur la CHT . . . 48
3.5 Résultats de l'évaluation de l'analyse de la bouche basée sur la CHT . . . 49
3.6 Résultats de l'évaluation du système de détection de la fatigue et de la somnolence sous la lumière ambiante du jour . . . 51
3.7 Résultats de l'évaluation du système de détection de la fatigue et de la somnolence sous un éclairage artificiel pendant la nuit . . . 52
3.8 Comparaison avec divers systèmes de détection de l'hypovigilance . . . 53

4.1 Comparaison des MAE pour le modèle cylindrique et le modèle ellipsoïdal de la tête (Choi et Kim, 2009) . . . 62
4.2 Étude comparative entre SVM, SVR et LARR (Guo *et al.*, 2008) . . . 70
4.3 Tableau récapitulatif des estimateurs de la pose de la tête. (1) écart type, (2) RMS, (3) MAE, (4) CCR. (*) estimateur dédié au conducteur. MF (Modèle flexible) ; MG (Méthode géométrique) ; Cl (Classification) ; Rg (Régression) ; Sv (Suivi) ; TA (Template d'apparence) . . . 74

5.1 Pitch-MAE en variant nb_{filt} et *step* . . . 88
5.2 Yaw-MAE en variant nb_{filt} et *step* . . . 88
5.3 Pitch-MAE en variant *level* et en considérant les trois meilleurs valeurs pour nb_{filt} et *step* . . . 89
5.4 Yaw-MAE en variant *level* et en considérant les trois meilleurs valeurs pour nb_{filt} et *Step* . . . 89
5.5 Comparaison de l'approche SP-LPF avec la littérature en termes de pitch-MAE et yaw-MAE . . . 90
5.6 Matrice de confusion de SP-LPF pour la séquence du conducteur selon le pitch . 92
5.7 Matrice de confusion de SP-LPF pour la séquence du conducteur selon le yaw . . 92

6.1 Évaluation des descripteurs sans utiliser la sélection des variables . . . 102
6.2 Performance de DF-SVM en variant les techniques de sélection des variables . . . 103

6.3 Visualisation de la participation de chaque descripteur dans le vecteur caractéristique final 104
6.4 Comparaison de l'approche DF-SVM avec la littérature en termes de pitch-MAE et yaw-MAE 104
6.5 Matrice de confusion de l'estimation de la pose de la tête selon le pitch en utilisant l'approche DF-SVM pour la séquence du conducteur 105
6.6 Matrice de confusion de l'estimation de la pose de la tête selon le yaw en utilisant l'approche DF-SVM pour la séquence du conducteur 105

Chapitre

Introduction générale

Sommaire

1.1 Contexte général 1
1.2 Problématique 2
1.3 Motivation 2
1.4 Objectif de la thèse 4
1.5 Hypovigilance chez le conducteur : état de l'art 5
1.5.1 Étude des signaux physiologiques du conducteur 5
1.5.2 Étude du comportement du véhicule 7
1.5.3 Étude des signaux physiques du conducteur 9
1.6 Organisation de la thèse 11
1.7 Liste des publications 12

1.1 Contexte général

Les expressions faciales transmettent des signaux non verbaux qui jouent un rôle important dans la communication entre les êtres humains. Notre cerveau est capable d'effectuer une étude du visage en un fragment de seconde afin de déterminer les relations et les connexions complexes reflétant l'état d'esprit et le comportement d'un individu. Cette étude débute par une analyse des structures basiques du visage afin d'obtenir une impression générale, suivie par une analyse des formes, puis une reconnaissance de la disposition spatiale des caractéristiques faciales. Bien que nous possédons la capacité d'extraire et de reconnaitre les expressions faciales sans effort et sans délai, ces tâches présentent encore des défis pour les machines. En effet, par l'analyse de l'image, la machine devra « voir » et « comprendre » le visage présent dans une image de la même façon que le cerveau, afin d'émettre une décision sur le comportement et l'état d'esprit de l'individu. Grâce à ces propriétés décisionnelles, le domaine de l'analyse automatique des expressions faciales connait un grand essor et constitue une tâche primordiale pour diverses applications telles que la robotique, les interfaces homme-machine, la science du comportement, la pratique clinique et la surveillance de l'état du conducteur.

1.2 Problématique

Dans cette thèse, nous nous focalisons sur la surveillance de l'état du conducteur qui sollicite un intérêt universel, causé par le nombre croissant des accidents routiers. D'après « The European accident research and safety report 2013 » (Volvo, 2013) établi par Volvo Truck Corporation, environ 1.2 million de décès sont signalés chaque année à travers le monde suite à des accidents de la route. Selon ce même rapport, 90% de ces accidents sont dus à des erreurs humaines correspondant principalement à l'hypovigilance chez le conducteur. Un autre rapport publié en 2011 par le Center of Accident Research and Road Safety-Queensland (CARRS-Q) [1] a conclu que 30% des décès sur les routes sont causés par l'hypovigilance chez le conducteur (CARRS-Q, 2011). Ce taux peut atteindre 50% dans des cas bien particuliers tels que les accidents mortels impliquant un seul véhicule. L'Association des Sociétés Françaises d'Autoroutes (ASFA) [2] a affirmé qu'en 2010, l'hypovigilance était la première cause des accidents sur les autoroutes (1 accident sur 3), suivie de la conduite en état d'ivresse (1 accident sur 4) et l'excès de vitesse (1 accident sur 8) (ASFA, 2010). D'autres statistiques fournies par la National Highway Traffic Safety Administration (NHTSA) [3] affirment que 100.000 accidents sont liés à l'hypovigilance dont 1550 sont fatals et 40.000 occasionnent des blessures graves (NHTSA, 2010). Au Maroc (Benjelloun, 2013), les statistiques des accidents impliquant un seul véhicule entre 2007-2011 sont données par la figure 1.1 et les décès qui en résultent sont présentés dans la figure 1.2. La problématique des accidents à un seul véhicule est intimement liée à la diminution de la vigilance des conducteurs et correspond à environ 43% des accidents en 2011, ce qui est très conséquent. A partir de ces deux figures, nous pouvons conclure que ce type d'accidents était en diminution entre 2007 et 2010, puis a commencé à augmenter à partir de 2011 aussi bien au niveau des accidents (+0.28%), qu'au niveau des décès (+0.68%).

1.3 Motivation

Les technologies dédiées à l'assistance du conducteur sont largement étudiées par l'industrie automobile. La surveillance de l'état du conducteur est l'une de ces technologies qui sollicite un énorme intérêt pour l'industrie automobile, mais aussi pour plusieurs gouvernements. En effet, le taux alarmant de mortalité liée aux accidents provoqués par l'hypovigilance (voir section 1.2) prouve qu'il est impératif d'agir en développant des systèmes intelligents pour surveiller l'état du conducteur.

Il existe plusieurs déclencheurs de l'hypovigilance dont les plus importants sont : la fatigue, les troubles du sommeil, la prise de somnifères ou de drogues, une conduite de plus de deux heures sans repos ou dans un environnement monotone tel que les autoroutes. Contrairement

1. CARRS-Q est un centre Australien dédié à la recherche et l'éducation dans le domaine de la sécurité routière au niveau national et international. Il évalue les dégâts humains, économiques et matériels causés par les accidents routiers.

2. L'ASFA est une association professionnelle regroupant tous les acteurs du secteur de la concession, de l'exploitation d'autoroutes et des ouvrages routiers en France.

3. La NHTSA est une agence fédérale américaine des États-Unis chargée de la sécurité routière créée en 1970. Elle est chargée de définir et de faire appliquer les standards de construction des infrastructures routières et des véhicules.

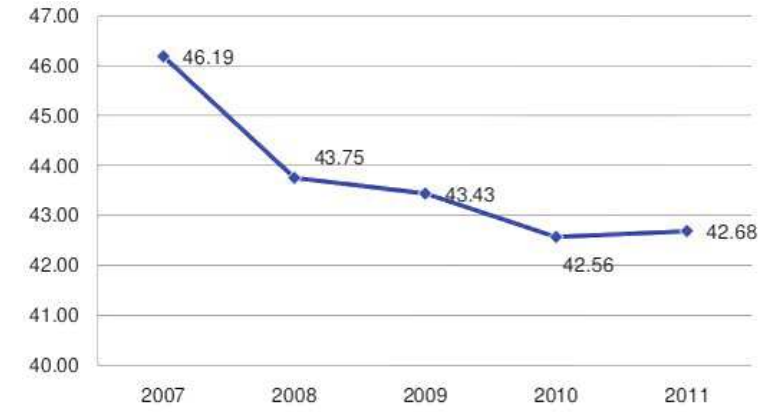


Figure 1.1 – Nombre d'accidents à un seul véhicule pour cent au Maroc (2007-2011) (Benjelloun, 2013)

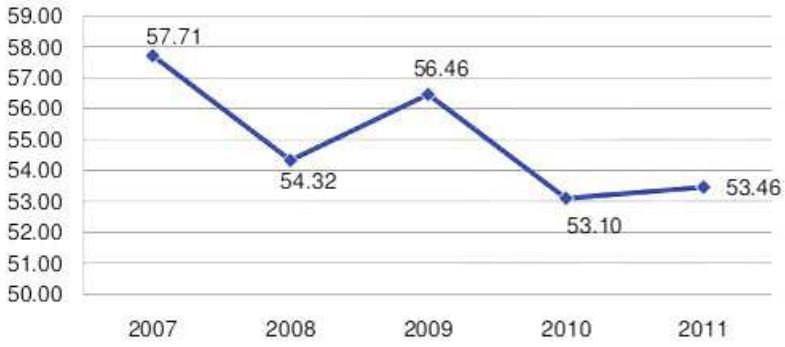


Figure 1.2 – Nombre de décès pour cent qui résulte des accidents à un seul véhicule au Maroc (2007-2011) (Benjelloun, 2013)

aux systèmes préventifs luttant contre les accidents provoqués par la vitesse excessive ou la consommation d'alcool, tels que les alcoomètres ou les radars, il est beaucoup plus compliqué de concevoir un outil préventif contre l'hypovigilance. En effet, il n'existe aucun outil standard pour mesurer le niveau de vigilance, la seule solution est d'observer les signes révélateurs de l'hypovigilance émis par le conducteur et de les analyser. Ces signes peuvent être divisés en signes comportementaux et physiologiques :

- Les signes comportementaux se manifestent par un comportement anormal du conducteur et sont représentés par :
  - une lenteur de réaction,
  - une inattention à l'environnement (panneaux de signalisation, obstacles, piétons, $\cdots$),
  - des erreurs de coordination,
  - une incapacité à maintenir une vitesse ou une trajectoire fixe.
- Les signes physiologiques apparaissent comme des expressions anormales principalement au niveau du visage du conducteur. Ils se manifestent par :
  - un picotement des yeux,
  - une raideur de la nuque ou des douleurs au dos,
  - des bâillements fréquents,

– une difficulté à maintenir les yeux ouverts et la tête en position frontale,
– des périodes de micro-sommeil (endormissement entre 2 à 6 secondes).

Quand un des signes apparaît, il est primordial de prendre une pose d'au moins 15 minutes avant de reprendre la conduite. Malheureusement, les conducteurs ont tendance à surestimer leur niveau de vigilance et ignorent très souvent ces signes. Selon quelques observations Dawson et Reid (1997); Lamond et Dawson (1999); Williamson et Feyer (2000), conduire entre 2h et 5h du matin, ou bien dormir moins de 5 heures par jour multiplie respectivement par 5 et par 3 le risque d'avoir un accident. De plus, il est précisé que conduire plus de 17 heures par jour équivaut à conduire avec un taux d'alcoolémie sanguine de 0.1 g/l (la normale étant de 0.03 g/l). Selon Stutts *et al.* (1999), 50% des conducteurs impliqués dans un accident provoqué par l'hypovigilance ont dormi moins de 6 heures la veille et 40% sont restés éveillés plus de 17 heures. Il a été noté que même si le trafic nocturne est de 10%, les accidents se produisant la nuit provoquent 37% de blessés graves et 45% de décès.

1.4 Objectif de la thèse

L'objectif de cette thèse est de proposer un système basé sur l'analyse des expressions faciales du conducteur afin d'estimer son niveau de vigilance. En premier lieu, il est nécessaire de distinguer entre les trois niveaux de l'hypovigilance qui correspondent à l'inattention, la fatigue et la somnolence. En général, ces termes sont considérés comme des notions bien définies, pourtant ils ne l'ont jamais clairement été en tant que concepts scientifiques (Regan, 2010). En effet, de nombreux travaux sur l'hypovigilance au volant ne définissent pas le concept même qu'ils étudient. L'absence de définitions bien établies est problématique car elle peut rendre les comparaisons entre différentes études délicates et peut aussi engendrer des estimations très variables du rôle des différents niveaux de vigilance dans les accidents de la route. Cependant, nous établissons une définition de chaque état du conducteur comme suit :

– **La somnolence** est l'incapacité à se maintenir éveillé. Elle se caractérise par un changement d'état de conscience et peut avoir des conséquences comportementales graves (Chaumet et Philip, 2007). La somnolence est caractérisée par les périodes de micro-sommeil correspondant à des endormissements de 2 à 6 secondes.
– **La fatigue** se définit [4] par la difficulté croissante à maintenir une tâche à terme et un effondrement des performances. Elle se manifeste par la diminution progressive de la vigilance physique et mentale qui conduit vers la somnolence. La fatigue est caractérisée par une augmentation du nombre de bâillements.
– **L'inattention** se produit lorsque l'attention est détournée de l'activité de conduite pour une raison non indispensable telle qu'un objet (téléphone portable, panneau publicitaire, nourriture), un évènement (accident, éclairages), passagers (enfant) ou autres usagers de la route (piéton, moto) (Regan, 2010). L'inattention se caractérise par le maintien de la pose de la tête en une direction autre que frontale.

4. Définition donnée par le professeur Pierre Philip, spécialiste de la neuropsychopharmacologique du sommeil, CHU Pellegrin Bordeaux

Il est évident à partir de ces définitions que la somnolence est l'état le plus critique de l'hypovigilance. Nous avons donc conçu en premier lieu une approche permettant de relever la somnolence chez le conducteur en détectant les périodes de micro-sommeil à partir de l'analyse des yeux. Ensuite, nous avons raffiné cette approche en proposant une étude de la fatigue à partir de l'analyse de la bouche afin de détecter les bâillements. Enfin, nous avons conçu un estimateur de la pose de la tête pour permettre à la fois de détecter l'inattention du conducteur et de déterminer son état même quand ses caractéristiques faciales (les yeux et la bouche) ne sont pas visibles. Pour faciliter par la suite une implémentation réelle du système, nous avons opté pour l'utilisation d'équipements à très faible coût, ainsi qu'une réduction de la complexité et du temps de calcul des algorithmes.

Dans la section suivante, nous allons présenter un bref état de l'art sur les diverses catégories de systèmes dédiées à la surveillance de l'état du conducteur.

1.5 Hypovigilance chez le conducteur : état de l'art

La conduite est une activité complexe qui implique la réalisation simultanée de nombreuses tâches : trouver son chemin, suivre la route, surveiller sa vitesse, éviter les obstacles, respecter le code de la route, maîtriser son véhicule, etc. (Regan, 2010). Il est donc évident que cette activité nécessite un niveau très élevé de vigilance afin d'éviter les accidents. Malheureusement, toutes les statistiques que nous avons présentées dans la section 1.2 montrent que les accidents liés à l'hypovigilance ne cessent de croitre. La plus récente de ces études (Volvo, 2013) estime que le nombre de décès sur les routes du monde atteint 1.2 million chaque année et que 90% de ces accidents sont dus principalement à des erreurs de la part des conducteurs. De ce fait, il est primordial de surveiller en permanence le niveau de vigilance des conducteurs afin de définir leur capacité à maintenir une conduite sûre et efficace.

Il existe divers travaux effectués pour développer des systèmes de surveillance de l'état du conducteur afin d'émettre des alarmes visuelles ou sonores quand son comportement est jugé anormal. Les avertissements émis peuvent être plus radicaux, tels que le déclenchement d'un mécanisme de vibration embarqué dans le siège du conducteur ou bien l'arrêt du véhicule sur le bord de la route. Nous pouvons distinguer trois catégories de systèmes dédiés à la surveillance de l'état du conducteur selon le type du signal utilisé pour déduire le niveau de vigilance. Nous allons présenter, dans ce qui suit, ces trois catégories qui permettent l'étude des signaux physiologiques (sous-section 1.5.1), l'analyse du comportement du véhicule (sous-section 1.5.2) et l'étude des signaux physiques (sous-section 1.5.3).

1.5.1 Étude des signaux physiologiques du conducteur

L'étude de l'hypovigilance à partir des signaux physiologiques consiste à mesurer la variation de signaux tels que les ondes cérébrales ou le rythme cardiaque en utilisant des capteurs spéciaux comme l'électroencéphalographie (EEG) (Berka *et al.*, 2007) ou l'électrocardiographie (ECG) (Shin *et al.*, 2010). L'EEG représente la technique la plus utilisée puisqu'il correspond à l'unique signal physiologique dont l'efficacité pour refléter l'état de vigilance a été prouvée (Berka *et al.*,

2007; Shen *et al.*, 2007; Ouyang et Lu, 2010; Shi et Lu, 2008). L'EEG mesure l'activité électrique des neurones par l'intermédiaire de plusieurs électrodes placées sur le cuire chevelu. La fréquence des ondes cérébrales dans l'EEG varie entre 1 Hz et 30 Hz et peut être divisée en quatre types en fonction des bandes de fréquences : delta (0-4 Hz), thêta (4-8 Hz), alpha (8-13 Hz) et bêta (13-20 Hz). Le rythme alpha représente un état de relaxation tandis que le rythme bêta est signalé pendant un état d'alerte. Les ondes delta correspondent à un état de somnolence et de sommeil. Le rythme thêta est associé à une variété d'états psychologiques impliquant une diminution du traitement de l'information.

Dans la plupart des travaux basés sur l'étude de la vigilance à partir du EEG, les signaux sont enregistrés pour plusieurs sujets et étiquetés par un expert. Ensuite, diverses techniques d'apprentissage sont utilisées pour établir la corrélation entre ces signaux et l'état observé du conducteur. Dans (Shi et Lu, 2008), il est noté que la majorité des approches exploitant les signaux EEG utilisent des techniques d'apprentissage supervisé pour estimer le niveau de vigilance. Toutefois, ce choix n'est pas judicieux puisqu'il n'existe aucun critère standard pour étiqueter les états de vigilance et les méthodes existantes dédiées à cette tâche sont complexes, couteuses et peu fiables. Shi et Lu (2008) optent donc pour une technique de clustering.

Shin *et al.* (2010) ont choisi d'utiliser des signaux autres que EEG. Ils décrivent une plateforme combinant entre un capteur ECG et un capteur photopléthysmogramme (PPG) embarqués dans le volant pour surveiller l'état du conducteur. Les signaux collectés par ces capteurs sont transmis à une station de base qui permettra de stocker, analyser et fournir les information concernant l'état du conducteur (Figure 1.3).

Figure 1.3 – Système d'acquisition et de transmission des signaux ECG et PPG à partir de capteurs placés sur le volant (Shin *et al.*, 2010)

Le laboratoire Fujitsu est en cours de développement d'un système de surveillance de la somnolence basé sur l'analyse des pulsations cardiaques, nommé Fujitsu Sleepiness Detection Sensor (Fujitsu, 2013). Le conducteur devra porter un capteur sans fil de pulsations cardiaques autour de l'oreille afin de mesurer et envoyer les données à un centre d'analyse. Une fois que les données révèlent un changement pouvant correspondre à une somnolence, le centre envoie un avertissement au conducteur.

Les signaux physiologiques permettent d'obtenir des résultats très satisfaisants pour l'esti-

mation de l'état du conducteur. Cependant, leur utilisation reste limitée à cause du prix des équipements et la nécessité d'embarquer des capteurs sur le corps humain (Figure 1.4) ou bien de les mettre en contacte avec celui-ci (Figure 1.3), ce qui est assez intrusif pour le conducteur. Cependant, les signaux physiologiques tels que les EEG sont fréquemment utilisés pour déterminer la vérité terrain afin de tester d'autres systèmes moins intrusifs pour la surveillance de l'état du conducteur.

Figure 1.4 – Capteur EEG placé sur la tête du conducteur (Khan et Aadil, 2012)

1.5.2 Étude du comportement du véhicule

La surveillance du comportement du véhicule peut révéler indirectement des actions anormales de la part du conducteur. Divers paramètres ont été étudiés tels que la force appliquée sur les pédales, le changement de vitesse, le mouvement du volant, le changement de voies, etc. Il existe un nombre limité de marques de véhicules qui proposent ce genre de système comme option pour quelques modèles. Volvo et Mercedes-Benz ont lancé pendant la même période des systèmes permettant de déterminer l'état du conducteur à partir du comportement du véhicule. En 2008, Volvo a conçu le premier dispositif en Europe permettant de détecter la fatigue et alerter le conducteur. Ce système, nommé Driver Alert Control (DAC) (DAC, 2008), se compose d'une caméra, de plusieurs capteurs et d'une unité de gestion. La caméra mesure en permanence le positionnement du véhicule par rapport aux marquages au sol. Les capteurs enregistrent les mouvements de la voiture. L'unité de gestion stocke les informations et calcule les risques de perte de contrôle du véhicule par le conducteur. Si le risque est estimé suffisamment élevé, le conducteur en est averti par un signal sonore. Un message textuel ainsi qu'un symbole représentant une tasse de café apparaissent sur l'écran d'information du véhicule (Figure 1.5), conseillant le conducteur de faire une pause.

En 2009, Mercedes-Benz a présenté le système Attention Assist (AttentionAssit, 2009) destiné à reconnaître une conduite influencée par l'hypovigilance, et encourage le conducteur à faire une pause dans ce cas. Ce système repose sur le fait que les conducteurs vigilants contrôlent constamment et inconsciemment la position de leur véhicule, et effectuent en permanence des petits ajustements de direction pour maintenir le véhicule sur la bonne voie. En cas d'hypovigilance, les interventions pour ajuster la direction sont moins fréquentes, et souvent accompagnées de corrections soudaines et exagérées lorsque l'attention est rétablie. Ainsi, Attention Assist

Figure 1.5 – Signal visuel émis par le système DAC en cas d'hypovigilance

Figure 1.6 – Signal visuel émis par le système Attention Assist en cas d'hypovigilance

identifie un type de direction caractéristique de l'hypovigilance qu'il combine avec d'autres informations telles que l'heure et la durée du trajet. Puisque le système ne peut fonctionner correctement en ville où le changement de voies est perturbé par le trafic, les 72 capteurs d'Attention Assist n'opèrent qu'à des vitesses comprises entre 80 et 180 km/h. Mercedes-Benz a effectué plusieurs tests, à l'aide de simulateurs et même sur la route, afin de déterminer le type de conduite caractéristique de l'hypovigilance. Les essais sur les routes ont été effectués dans des environnements extrêmes telles que le vent, la pluie, ou le brouillard, pour garantir la fiabilité des mesures et un EEG a été utilisé comme méthode objective pour évaluer le niveau de vigilance du conducteur. De cette manière, le niveau auquel Attention Assist émet une alerte a pu être réglé à un niveau approprié. Puisque chaque conducteur possède une conduite distincte, le système effectue un apprentissage pendant les 15 premières minutes de chaque trajet afin de déterminer la valeur de référence pour chaque paramètre impliqué dans le processus de surveillance du niveau de vigilance. Si une hypovigilance est identifiée, le système alerte le conducteur pour qu'il fasse une pause en affichant une tasse de café sur le tableau de bord (Figure 1.6), et en émettant un signal sonore. Le conducteur peut réagir à l'alerte et la faire disparaître de l'affichage. Si la pause n'est pas prise et que le style de conduite continue à indiquer une perturbation, l'alerte est répétée au bout de 15 minutes.

Contrairement aux systèmes basés sur l'analyse des signaux physiologiques, les systèmes basés sur l'analyse du comportement du véhicule ne sont pas intrusifs pour le conducteur. Ce-

pendant, ils sont limités par la dépendance au type du véhicule, l'expérience du conducteur, ainsi que les caractéristiques et les conditions de la route. En effet, le risque de fausses alarmes est très élevé en cas de généralisation pour divers types véhicules ou en cas de conditions météo extrêmes (vents très importants). De plus, le prix de ces systèmes est très important ; il faut compter environ 2000 $ pour le système DAC alors que Attention Assist est inclus dans la majorité des véhicules très haut de gamme sans proposer de prix pour ce système.

1.5.3 Étude des signaux physiques du conducteur

L'étude de l'hypovigilance à partir des signaux physiques repose principalement sur le traitement de la vidéo du conducteur pour mesurer le niveau de vigilance reflété par ses caractéristiques faciales. Il a été remarqué qu'en cas d'hypovigilance, le conducteur présente certains comportements visuels facilement observables à partir des changements des caractéristiques faciales telles que les yeux, la bouche et la pose de la tête (Momin et Abhyankar, 2012). Diverses études ont montré que l'activité des paupières est étroitement liée au niveau de vigilance. Le pourcentage de fermeture de l'œil en fonction du temps « Percentage of Eye Closure » (PERCLOS) a longtemps été la mesure la plus répandue pour détecter la somnolence chez le conducteur, puisqu'elle permet de déterminer la fermeture lente des yeux correspondant à l'assoupissement (Trutschel *et al.*, 2011). La fréquence de fermeture des yeux est aussi considérée comme un bon indicateur de la somnolence, puisqu'elle permet de détecter les périodes de micro-sommeil (Golz *et al.*, 2007). La fatigue peut être indiquée par une fréquence élevée du bâillement représentée par une grande ouverture de la bouche, tandis que l'inattention est souvent caractérisée par la pose de la tête ou la direction du regard (Momin et Abhyankar, 2012). La direction de la tête et/ou du regard possède la capacité de révéler l'inattention du conducteur et aussi sa condition mentale. La direction normale du regard du conducteur est frontale et le fait de maintenir d'autres directions pour de longues périodes peuvent indiquer une fatigue (fixité du regard) ou une inattention. La pose de la tête est étroitement liée à la direction du regard et doit aussi rester frontale le plus longtemps possible.

Il est vrai que la détection de l'hypovigilance à partir de l'analyse des changements physiques sollicite l'intérêt de plusieurs travaux de recherche. Néanmoins, il existe aussi des systèmes commerciaux appartenant à cette catégorie mais très peu de détails sont dévoilés sur ces produits. En 2009, la société québécoise Effective Control Transport (ECT) a lancé son système de prévention contre les accidents liés à l'hypovigilance pour les professionnels de la route, nommé Driver's Mate (Driver's Mate, 2009) et représenté par la figure 1.7. L'appareil correspond concrètement à une caméra dotée d'éclairages infrarouges qui permettent de traverser les verres fumés des lunettes. Il analyse 534 points du visage, principalement les yeux, afin de détecter les premiers signes de l'hypovigilance. Dès qu'une diminution du niveau de vigilance est détectée, le conducteur est alerté en temps réel, ce qui lui permet de planifier une pause. Toutefois, ce système ne semble pas être largement utilisé et aucun renseignement sur le prix n'est fourni.

La société suédoise SmartEye AG propose aussi divers systèmes mono ou multi caméras dédiés à la détection de l'hypovigilance. Leur produit Smart Eye Pro 5.10 (SmartEye 5, 2013), représenté par la figure 1.8, est un système multi-caméras qui permet le suivi du regard, l'estima-

Figure 1.7 – Système Driver's Mate de la société ECT

tion de la pose de la tête, la détermination de l'ouverture des paupières et la taille de la pupille. Ce système n'est pas dédié uniquement à l'industrie automobile mais peut être aussi utilisé pour l'aéronautique, les simulateurs et les tours de contrôle. Le nombre de caméras est configurable selon l'application et peut atteindre huit caméras. Toutefois, l'utilisation d'une seule caméra est souhaitable puisqu'elle permet une production industrielle plus facile et moins couteuse (Khan et Aadil, 2012).

Figure 1.8 – Système multi-caméras Smart Eye Pro 5.10

Les techniques de détection de l'hypovigilance à partir des caractéristiques faciales du conducteur ont l'avantage d'être non intrusives. Puisqu'elles se basent sur les caméras pour l'acquisition des signaux physiques, les équipements utilisés sont moins chers que ceux des autres catégories. Toutefois, il faut prendre en considération les limites de l'utilisation des caméras, à savoir la sensibilité aux changements d'éclairage et la nécessité d'utiliser des éclairages infrarouges pour permettre une acquisition dans des environnements obscurs.

Nous avons présenté dans cette section un bref état de l'art des différents types de techniques dédiées à la détection de l'hypovigilance chez le conducteur. Nous avons conclu que les approches basées sur l'analyse des signaux physiologiques (l'EEG par exemple) sont extrêmement intrusives et non tolérées par les conducteurs puisqu'elles nécessitent le déploiement de capteurs sur le corps. Cependant, les approches basées sur l'analyse du comportement du véhicule ne sont pas intrusives mais dépendent du type du véhicule et des conditions de la route. Ainsi, nous nous intéressons aux approches basées sur l'analyse des signaux physiques puisqu'elles ne sont pas intrusives et ne nécessitent pas d'équipements couteux. Nous proposons, en premier lieu, une

technique pour détecter l'état le plus critique de l'hypovigilance, qui correspond à la somnolence. Cette technique est basée sur l'analyse des yeux pour détecter les périodes de micro-sommeil. Ensuite, nous adaptons cette même technique pour détecter la fatigue à partir du bâillement. Enfin, nous proposons deux approches pour relever l'inattention à partir de l'estimation de la pose de la tête.

L'avantage de toutes nos contributions est qu'elles ne nécessitent aucun équipement spécial. En effet, une seule caméra bon marché qui coute dix euros est requise pour l'acquisition ainsi qu'un ordinateur pour les traitements. D'après les séries de tests que nous avons effectuées, ces traitements s'avèrent très prometteurs que ce soit au niveau de l'estimation de l'état du conducteur ou en ce qui concerne le temps d'exécution. Un apport majeur de nos travaux est la modélisation d'un système original composé d'un ensemble d'approches améliorant chacune les résultats obtenus dans la littérature. La première étape du système est l'estimation de la pose de la tête du conducteur afin de préciser son niveau d'inattention. Lorsque celui-ci est considéré attentif à la route suite à la présence d'une pose frontale, l'analyse de son état est alors approfondie pour rechercher la somnolence et la fatigue selon un scénario original que nous avons élaboré pour fournir une bonne estimation du niveau de vigilance, tout en réduisant au maximum le temps de calcul.

1.6 Organisation de la thèse

Dans ce manuscrit de thèse, nous débutons par la présentation des approches que nous avons élaborées pour la détection de la somnolence et de la fatigue chez le conducteur à partir des caractéristiques faciales incluses dans le visage (les yeux et la bouche). Du fait que nous avons eu recours aux mêmes techniques de traitement de l'image pour analyser ces caractéristiques, nous avons regroupé ces deux tâches dans la partie Partie I. Ainsi, nous présentons dans le chapitre 2 un état de l'art étendu sur la détection de la somnolence et la fatigue chez le conducteur. Ensuite, les approches que nous proposons pour détecter ces deux états sont détaillées dans le chapitre 3.

Nous consacrons la partie Partie II à la détection de l'inattention, qui se base sur l'estimation de la pose de la tête et nécessite des outils différents de ceux utilisés pour analyser les yeux et la bouche. Dans le chapitre 4, nous exposons un état de l'art détaillé sur les techniques l'estimation de la pose de la tête, notamment celles dédiées au conducteur. Par la suite, nous proposons dans le chapitre 5 et le chapitre 6 deux estimateurs de la pose de la tête du conducteur.

Finalement, nous présentons une conclusion générale portant sur l'ensemble des travaux de cette thèse, ainsi que des perspectives à court et à long termes.

1.7 Liste des publications

Revues internationales

N. Alioua, A. Amine and M., « Driver's Fatigue Detection Based on Yawning Extraction », ***International Journal of Vehicular Technology (IJVT)***, vol. 2014, Article ID 678786, 2014, Hindawi Publishing Corporation.

N. Alioua, A. Amine, M. Rziza, A. Bensrhair, « Estimating driver head pose using steerable pyramid and probabilistic learning », ***International Journal of Computer Vision and Robotics (IJCVR)***, in press, Inderscience Publishers, ISSN online : 1752-914X.

N. Alioua, A. Amine, A. Rogozan, A. Bensrhair, M. Rziza, « Driver head pose estimation using efficient descriptor fusion », submitted to *EURASIP Journal on Image and Video Processing*, Springer

Conférences internationales

N. Alioua, A. Amine, M. Rziza, D. Aboutajdine, « Eye State Analysis Using Iris Detection to Extract Micro-Sleep Periods », ***International Conference on Computer Vision Theory and Applications (VISAPP'11)***, Vilamoura, Protugal, 2011.

N. Alioua, A. Amine, M. Rziza, D. Aboutajdine, « Eye State Analysis using Iris Detection based on Circular Hough Transform », ***International Conference on Multimedia Computing and Systems, (ICMCS'11)***, Ouarzazate, Morocco, 2011.

N. Alioua, A. Amine, M. Rziza, D. Aboutajdine, « Fast Micro-Sleep and Yawning Detections to Assess Driver's Vigilance Level », ***the 6th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design (DA'11)***, California, USA, 2011.

N. Alioua, A. Amine, M. Rziza, D. Aboutajdine, « Driver's Fatigue and Drowsiness Detection to Reduce Traffic Accidents on Road », ***Computer Analysis of Images and Patterns (CAIP'11), Lecture Notes in Computer Science***, Volume 6855, pp 397-404, Sevilla, Spain, 2011.

N. Alioua, A. Amine, M. Rziza, A. Bensrhair, D. Aboutajdine, « Head pose estimation based on steerable filters and likelihood parametrized function », ***the 21st European signal Processing conference (EUSIPCO'13)***, Marrakech, Morocco, 2013.

A. Amine, **N. Alioua**, F. Zann, Y. Ruichek, N. Hmina, « Monitoring Drivers Drowsiness Using a Wide Angle Lens », ***The 16th IEEE International Conference on Intelligent Transportation Systems (ITSC'13)***, The Hague, The Netherlands, pp. 290-295, 2013.

Partie I

Analyse des caractéristiques faciales pour la détection de la fatigue et de la somnolence

---

Chapitre

Analyse des caractéristiques faciales pour la détection de la somnolence et de la fatigue : état de l'art

Sommaire

2.1 Introduction 15
2.2 Détection du visage 16
2.2.1 Machines à Vecteurs de Support (SVM) 17
2.2.2 Méthodes de détection du visage 20
2.3 État d'ouverture/fermeture des yeux 21
2.4 PERCLOS 23
2.5 Fréquence de clignement des yeux 23
2.6 Fréquence de bâillement 24
2.7 Conclusion 25

2.1 Introduction

Puisque les caractéristiques faciales (particulièrement les yeux et la bouche) sont un moyen naturel pour identifier la somnolence et la fatigue, plusieurs travaux les exploitent en utilisant des caméras pour l'acquisition et des techniques de traitement de la vidéo pour l'analyse. Cependant, la majorité des études se basent sur l'analyse des yeux puisqu'ils sont considérés comme des indicateurs puissants de l'état du conducteur, tandis que peu de travaux intègrent l'analyse de la bouche. En effet, plusieurs paramètres peuvent être étudiés pour l'analyse des yeux tels que le clignement des paupières, la fermeture de l'œil et la fixité du regard, alors que la bouche n'est analysée que pour retrouver le bâillement ou bien la parole.

L'analyse du comportement du conducteur à partir de l'image débute par l'acquisition de celle-ci. L'image doit avoir des propriétés photométriques consistantes sous différentes conditions climatiques et ambiantes. Elle doit également produire des caractéristiques distinguables afin de faciliter son traitement. Ainsi, nous pouvons distinguer entre deux types de caméras utilisées pour l'acquisition, à savoir les caméras à spectre visible et les caméras opérant sous un éclairage infrarouge. Les caméras infrarouges sont souvent utilisées pour produire un effet sombre ou brillant de la pupille afin de la mettre en évidence dans l'image, comme illustré par la figure 2.1. Ces caméras conviennent aux conditions stables d'éclairage mais quand celui-ci est très variable, la pupille disparait et la détection des yeux devient difficile. Les caméras à spectre visible sont bien adaptées à la conduite de jour et peuvent aussi opérer dans des environnements

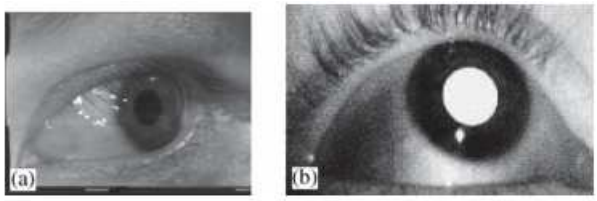


Figure 2.1 – Effet de la lumière infrarouge sur l'œil. (a) Effet sombre ; (b) Effet brillant (Ji et Yang, 2004)

peu éclairés. De plus, ces dernières sont généralement moins chères que les caméras infrarouges. Dans la suite du manuscrit, le terme « caméra » correspond à une caméra à spectre visible, sauf si nous mentionnons explicitement qu'il s'agit d'une caméra infrarouge.

Nous désirons spécifier que la majorité des travaux basés sur l'étude des caractéristiques faciales pour déterminer l'état du conducteur débutent par une détection du visage afin de limiter la zone de recherche de ces caractéristiques. Nous présentons donc dans la section 2.2 un bref aperçu sur la détection du visage. Jusqu'à présent, il n'existe aucun état de l'art complet sur les techniques permettant d'analyse les caractéristiques faciales pour la détection de la fatigue et de la somnolence chez le conducteur. Les travaux de recherche se contentent de lister quelques techniques existantes sans les organiser. Après une étude détaillée de l'existant, il nous est paru judicieux de catégoriser ces techniques selon les paramètres déduits des caractéristiques faciales. Nous avons relevé quatre paramètres fréquemment utilisés qui correspondent à la détection des yeux ouverts/fermés (section 2.3), le PERCLOS (section 2.4), la fréquence de clignement des yeux (section 2.5) et la fréquence de bâillement (section 2.6).

2.2 Détection du visage

La détection du visage en utilisant la vision par ordinateur intervient dans plusieurs domaines nécessitant la reconnaissance des individus ou la détermination de leur état. Son concept de base consiste à parcourir l'image avec une fenêtre, puis à comparer chaque fenêtre à une série de visages types présents dans une base de données. Un visage est défini comme étant une fenêtre dont la distance avec l'une des images de la base est suffisamment faible. Il est donc nécessaire que cette tâche soit basée sur des éléments stables et relativement descriptifs du visage humain tels que la forme du visage, la couleur de la peau, le contour des yeux, la forme du nez ou de la bouche, etc. En considérant les tailles, les orientations, les rotations et les éclairages, il faudrait comparer chaque fenêtre d'une image à des centaines de références ! De plus, si les expressions faciales (sourires, grimaces,...) sont prises en compte, la détection du visage devient un problème difficile à traiter. Ainsi, ce problème est souvent traité par l'analyse statistique et l'apprentissage automatique pour construire des machines capables de séparer les visages des non-visages. Les machines à vecteurs de support « Support Vector Machine » (SVM) appartiennent aux méthodes d'apprentissage automatique les plus souvent utilisées, que nous présentons dans la sous-section 2.2.1. Par la suite, nous citons dans la sous-section 2.2.2 quelques techniques de détection des visages et plus précisément celles utilisant des SVM.

2.2.1 Machines à Vecteurs de Support (SVM)

La capacité de généraliser des résultats obtenus à partir d'un nombre limité d'échantillons est l'enjeu majeur de l'apprentissage artificiel. En effet, il est bien connu que la minimisation du risque empirique ou l'erreur d'apprentissage n'est pas suffisante et ne garantit pas une faible erreur sur l'ensemble de test. Ainsi, des techniques de régularisation sont utilisées afin de réaliser un compromis entre la capacité du modèle à apprendre, liée à sa complexité, et son aptitude à généraliser. D'un point de vue conceptuel, la notion de risque structurel introduite par Vapnik (1995), fournit une borne de l'erreur de test en fonction de l'erreur d'apprentissage et de la complexité du modèle. D'un point de vue pratique, les SVM offrent un moyen opérationnel pour minimiser le risque structurel (Cortes et Vapnik, 1995), ce qui explique le grand intérêt que leur porte la communauté scientifique. Les SVM sont un ensemble de techniques d'apprentissage supervisé principalement conçues pour résoudre des problèmes de discrimination, permettant de décider à quelle classe appartient un échantillon. Cependant, ils peuvent aussi résoudre des problèmes de régression visant à prédire la valeur numérique d'une variable (Scholkopf et Smola, 2001). La résolution de ces deux problèmes se base sur la construction d'une fonction h qui, à un vecteur d'entrée x, fait correspondre une sortie y ($y = h(x)$). Si nous considérons un problème de discrimination à deux classes alors $y\epsilon\{-1, 1\}$.

Le principe théorique des SVM repose sur deux points fondamentaux :

- Le premier point est représenté par la transformation non linéaire Φ qui projette les exemples de l'espace d'entrée vers un second espace de grande dimension muni d'un produit scalaire, et nommé espace de redescription des données.
- Le second point permet de définir un hyperplan offrant une séparation linéaire optimale dans l'espace de redescription. Ce point est particulièrement utile pour traiter les cas où les données ne sont pas linéairement séparables. En effet, dans l'espace de redescription, la séparation est plus simple du fait que plus la dimension de l'espace est grande, plus la probabilité de trouver un hyperplan séparateur entre les exemples est élevée. Puisque le problème de recherche de l'hyperplan séparateur optimal possède une formulation duale, il est donc possible de résoudre ce problème par des méthodes d'optimisation quadratique standard.

D'un point de vue mathématique, la transformation non linéaire Φ est réalisée par une fonction noyau, ce qui a l'avantage de ne pas nécessiter la connaissance explicite de la transformation à appliquer pour le changement d'espace. De plus, la fonction noyau permet de transformer un produit scalaire dans un espace de grande dimension, ce qui est coûteux, en une simple évaluation ponctuelle d'une fonction. En pratique, quelques fonctions noyau paramétrables sont connues et il revient à l'utilisateur d'effectuer des tests pour déterminer celle qui convient le mieux à son application. Il s'agit donc de traduire le maximum de connaissances préalables dont nous disposons sur le problème étudié et sur les données.

Même dans les cas simples de problèmes linéairement séparables, il n'est pas évident de choisir l'hyperplan séparateur. En effet, il existe une infinité d'hyperplans séparateurs possédant certaines performances en apprentissage mais dont les performances en généralisation peuvent être très différentes. Selon Vapnik et Kotz (1982), il existe un unique hyperplan optimal permet-

tant de résoudre ce problème. Cet hyperplan optimal, illustré par la figure 2.2, est défini comme étant l'hyperplan qui maximise la marge entre les échantillons et l'hyperplan séparateur. Ce choix est justifié théoriquement par le fait que la capacité des classes d'hyperplans séparateurs diminue lorsque leur marge augmente. La marge est la plus petite distance entre les échantillons d'apprentissage et l'hyperplan séparateur qui satisfait la condition de séparabilité. Ces échantillons sont appelés vecteurs supports. Pour expliquer la condition de séparabilité, nous considérons un cas simple de fonction discriminante linéaire obtenue par une combinaison linéaire du vecteur d'entrée $x = (x_1, \cdots, x_N)^T$ de dimension N et d'un vecteur de poids $w = (w_1, \cdots, w_N)$. Ainsi, $h(x) = w^T x + w_0$ et x appartient à la classe 1 si $h(x) \geq 0$ ou à la classe -1 sinon. La frontière de décision $h(x) = 0$ est un hyperplan séparateur. Le but d'un algorithme d'apprentissage supervisé est d'apprendre la fonction $h(x)$ par le biais d'un ensemble d'apprentissage de taille p définie par l'équation 2.1 où les l_k représentent les étiquettes.

$$\{(x_1, l_1), \cdots, (x_p, l_p)\} \subset \mathbb{R}^N \times \{-1, 1\} \tag{2.1}$$

Si le problème est linéairement séparable, alors la condition de séparabilité est donnée par l'équation 2.2.

$$l_k(w^T x_k + w_0) \geq 0 \text{ Avec } 1 \leq k \leq p \tag{2.2}$$

Ainsi, l'hyperplan qui maximise la marge peut être défini par l'équation 2.3.

$$\arg\max_{w,w_0} \min_k \{\|x - x_k\| : x \epsilon \mathbb{R}^N, w^T x + w_0 = 0\} \tag{2.3}$$

Il s'agit donc de trouver w et w_0 remplissant les conditions de l'équation 2.3, afin de déterminer l'hyperplan séparateur exprimée par l'équation 2.4

$$h(x) = w^T x + w_0 = 0 \tag{2.4}$$

La distance entre un échantillon x_k et l'hyperplan est donnée par sa projection orthogonale sur l'hyperplan (voir l'équation 2.5)

$$\frac{l_k(w^T x_k + w_0)}{\|w\|} \tag{2.5}$$

L'hyperplan séparateur (w, w_0) de marge maximale est donc donné par l'équation 2.6

$$\arg\max_{w,w_0} \left\{ \frac{1}{\|w\|} \min_k \left[l_k(w^T x_k + w_0) \right] \right\} \tag{2.6}$$

Pour faciliter l'optimisation, w et w_0 sont normalisées pour que les échantillons à la marge (x^+_{marge} pour les vecteurs supports sur la frontière positive, et x^-_{marge} pour ceux situés sur la frontière opposée) satisfassent l'équation 2.7

$$\begin{cases} w^T x^+_{marge} + w_0 = 1 \\ w^T x^-_{marge} + w_0 = -1 \end{cases} \tag{2.7}$$

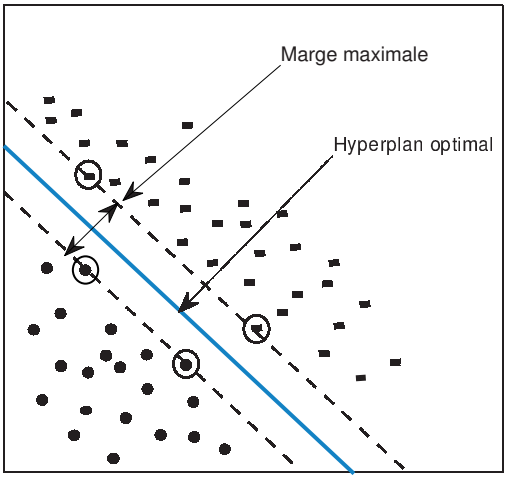


Figure 2.2 – Hyperplan séparateur optimal qui maximise la marge dans l'espace de redescription. Les échantillons entourés correspondent aux vecteurs supports

Ainsi, pour tous les échantillons $k = 1, \cdots, p$, nous obtenons l'inégalité 2.8

$$l_k(w^T x_k + w_0) \geq 1 \tag{2.8}$$

Avec cette mise à l'échelle, la marge est donnée par $\frac{1}{\|w\|}$, il s'agit donc de maximiser $\|w^{-1}\|$. La formulation dite primale des SVM s'exprime alors par l'équation 2.9 :

$$\text{Minimiser } \frac{1}{2}\|w^2\| \text{ sous les contraintes } l_k(w^T x_k + w_0) \geq 1 \tag{2.9}$$

Ceci peut être résolu par la méthode classique des multiplicateurs de Lagrange, où le lagrangien est donné par l'équation 2.10

$$L(w, w_0, \alpha) = \frac{1}{2}\|w^2\| - \sum_{k=1}^{p} \alpha_k \{l_k(w^T x_k + w_0) - 1\} \tag{2.10}$$

Toutefois, l'inconvénient de la forme classique des SVM est le coût élevé de la fonction de décision surtout pour les applications temps réels. En effet, la complexité temporelle d'une opération de classification SVM est influencée par deux paramètres. Premièrement, la complexité est linéaire au nombre de vecteurs supports. Deuxièmement, elle dépend du nombre d'opérations nécessaires pour le calcul de la similarité entre un vecteur support et l'entrée, ce qui correspond à la complexité de la fonction noyau. Lors de la classification d'images de taille $h \times w$, la fonction de décision nécessite un produit scalaire dimensionnelle $h \bullet w$ pour chaque vecteur support. Plus la taille du patch augmente, plus ces calculs deviennent couteux. Par exemple, l'évaluation d'images de taille 20×20 sur une image de 320×240 à 25 frames par seconde nécessite 660 millions opérations par seconde. Ainsi, plusieurs recherches permettant d'accélérer l'expansion du noyau ont été réalisées, principalement sur le premier paramètre visant à réduire le nombre

de vecteurs supports. Burges (1996) introduit une méthode qui pour une donnée SVM, crée un ensemble réduit de vecteurs supports approximant la fonction de décision. Cette approche a été appliquée avec succès dans le domaine de la classification d'images et a permis une accélération de l'ordre de 10 à 30 en conservant l'exactitude complète.

2.2.2 Méthodes de détection du visage

Les approches basées sur les SVM reposent essentiellement sur la théorie de décision pour résoudre les problèmes de classification. Osuna *et al.* (1997) ont développé une méthode efficace pour former un SVM pour des problèmes à grande échelle, et l'ont appliqué à la détection de visages. Kumar et Poggio (2000) ont incorporé un SVM dans un système pour l'analyse des visages en temps réel. Ils appliquent cet algorithme sur des régions segmentées de la peau dans les images d'entrée pour éviter le balayage approfondi. Karam *et al.* (2004) ont crée un système de détection de visages et d'extraction des caractéristiques faciales basé sur les SVM et appliqué sur des visages parlants dans des séquences vidéo. Une machine SVM est apprise sur des fenêtres après leur transformation dans le domaine d'ondelettes. Un modèle géométrique statistique est ensuite appliqué afin de lisser la sortie des SVM et d'affiner la détection. Un autre modèle probabiliste sur les distances aux frontières SVM permet plus de lissage et une meilleure sélection des composantes faciales.

Yang et Ahuja (1998) ont présenté une méthode pour détecter des visages humains à partir d'images en couleur. Un modèle de la couleur de peau humaine basé sur une analyse statistique multi-variante est construit pour capturer les propriétés chromatiques.

Schneiderman et Kanade (1998) décrivent deux détecteurs de visage basés sur la décision de Bayes présenté par l'équation 2.11

$$\frac{P(image/visage)}{P(image/Nonvisage)} > \frac{P(Nonvisage)}{P(visage)} \tag{2.11}$$

Si le rapport de probabilité (côté gauche) de l'équation 2.11 est plus grand que l'autre côté, alors un visage est considéré présent à l'endroit courant. L'avantage de cette approche est l'optimalité de la règle de décision de Bayes si les images sont précises.

Un des détecteurs du visage le plus connu est celui proposé par Viola et Jones (2001). L'un des points principaux de la méthode consiste à parcourir l'ensemble de l'image en calculant un certain nombre de caractéristiques pseudo-haar (entre 4 et 14) dans des zones rectangulaires qui se chevauchent. Ces caractéristiques sont déterminées par la différence des sommes de pixels de deux ou plusieurs zones rectangulaires adjacentes. Elles sont calculées à toutes les positions et à toutes les échelles dans une fenêtre de détection de petite taille. Afin de réduire le temps de calcul des caractéristiques, une image intégrale est utilisée. Il s'agit d'une image construite à partir de l'image d'origine, et de même taille qu'elle. Elle contient en chacun de ses points la somme des pixels situés au-dessus et à gauche du pixel courant. L'autre point important de la méthode Viola & Jones est la sélection par boosting des caractéristiques, qui consiste à utiliser plusieurs classifieurs faibles mis en cascade, plutôt que d'utiliser un seul classifieur fort. Dans le cas d'une mise en cascade de classifieurs dont le critère de sélection serait moins sévère, une

fenêtre est rejetée dès que l'un des étages estime qu'il n'y a pas de visage, ce qui permet un gain de temps considérable. Une étape préliminaire et non des moindres est l'apprentissage du classifieur. Il s'agit d'entraîner le classifieur afin de le sensibiliser aux visages, en présentant une grande quantité d'images de visages puis d'images non-visages.

Un autre détecteur de visages très efficace basé sur les SVM est proposé par Romdhani *et al.* (2001) et optimisée par Kienzle *et al.* (2005). Cette méthode a donné naissance à la bibliothèque libre fdlib, dont nous détaillons le fonctionnement dans la sous-section 3.2.1.

2.3 État d'ouverture/fermeture des yeux

Il existe des méthodes qui détectent la somnolence en comptant le nombre de fermetures consécutives des yeux à travers la séquence vidéo du conducteur. Si ce compteur dépasse un seuil de tolérance, une somnolence est détectée. D'Orazio *et al.* (2004) ont proposé un algorithme de détection des yeux du conducteur qui recherche l'œil dans l'image entière en supposant que l'iris est toujours plus sombre que la sclère. Ils localisent les candidats pouvant représenter l'œil par une approche géométrique et la Transformée de Hough Circulaire « Circular Hough Transform » (CHT), qui permet de retrouver les objets de formes circulaires dans l'image. Cependant, il est impossible de localiser les yeux quand ils sont fermés et dans ce cas, la première étape fournit des régions pouvant contenir l'œil fermé, mais avec plus de faux positifs. Pour résoudre ce problème, un réseau de neurones est appris pour distinguer entre deux classes d'images (œil et non œil). Cette étape permet de confirmer la détection de l'œil ouvert obtenue par la première étape et, le cas échéant, de différencier entre une région non œil et un œil fermé. Les tests effectués pour valider cette approche sur 6 séquences, pour une total de 1474 images, fournissent un taux de Bonne Classification « Correct Classification Rate » (CCR) de 93%. Cette technique est différente de la plupart des approches proposées dans la littérature, qui localisent la zone du visage avant de localiser et déterminer l'état des yeux.

Dans (Zhang *et al.*, 2008), une approche basée sur des templates de l'œil du conducteur est utilisée pour déterminer l'état des yeux à partir de leur degré d'ouverture. En supposant que le conducteur n'effectue que de faibles mouvements, la zone du visage est déterminée par le calcul de la différence entre la frame courante et la frame accumulée, sur laquelle une binarisation par seuillage adaptative est appliquée. La zone de recherche des yeux est donnée par les $\frac{2}{5}$ de la partie supérieure du visage, qui est divisée verticalement en deux pour représenter chaque œil. Les candidats pouvant représenter l'œil sont déterminés dans cette zone en fonction de leur rapport longueur sur largeur et de leur taille. Ensuite, les templates des yeux sont obtenus par le calcul de la distance entre la position des candidats dans la frame courante et la frame précédente. Si cette position est stable pour une certaine période, ce candidat est considéré comme un template de l'œil. Une normalisation et une accumulation des templates correspondant à l'œil gauche et l'œil droit sont appliquées pour retrouver le template final pour chaque œil. Par la suite, une mise en correspondance par coefficient d'intercorrélation entre les templates et la zone de recherche de l'œil dans la frame courante est calculée. La déduction de l'état de l'œil est effectuée par l'extraction de son squelette par une binarisation (Figure 2.3-b), suivie d'un raffinement (Figure 2.3-c) et une réduction pour obtenir un contour fin (Figure 2.3-d). Ensuite, la paupière inférieure

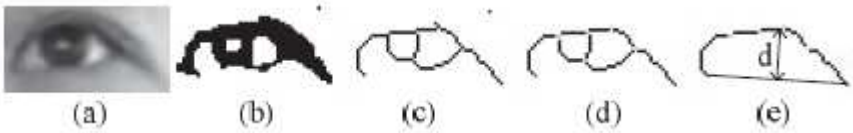


Figure 2.3 – Extraction de l'état de l'œil (Zhang *et al.*, 2008). (a) Image de l'œil ; (b) Binarisation ; (c) Raffinement ; (d) Contour fin ; (e) Distance entre les paupières inférieure et supérieure

est représentée par la ligne entre les 2 extrémités du squelette de l'œil. Finalement, le degré de fermeture est calculé comme étant la distance entre le point le plus élevé de la paupière supérieure et de la paupière inférieure (Figure 2.3-e). Si cette distance est inférieure à un seuil, l'œil est considéré fermé. Une étude a été effectuée par les auteurs pour valider la méthode sur une seule séquence. Le CCR s'élève à 93%. Cette méthode s'avère être efficace pour la détection de l'œil du conducteur mais très consommatrice en temps de calcul.

Horng *et al.* (2004) ont développé une approche de détection de la somnolence fondée sur le suivi et l'analyse de l'œil. Le visage du conducteur est localisé à partir de la première frame en utilisant les caractéristiques de la couleur de la peau, et les deux régions des yeux sont déterminées par la détection des contours. Par la suite, ces régions sont utilisées comme templates dynamiques pour le suivi des yeux. En cas d'échec du suivi, l'étape de détection du visage et des yeux est réitérée. La somnolence est déterminée par l'analyse de la couleur du globe oculaire, considérée plus sombre que la peau. Cette observation est utilisée pour déterminer si un pixel de la région de l'œil correspond à un pixel du globe oculaire. En cas d'absence des pixels du globe oculaire, l'œil est considéré fermé et si cette fermeture est observée pendant cinq frames consécutives, le conducteur est alerté contre la somnolence.

Tian et Qin (2005) ont aussi proposé une approche pour la détection de la somnolence débutant par des étapes de localisation du visage et de la région des yeux basées sur les informations de couleur. La somnolence chez le conducteur est déterminée par le calcul d'une fonction de complexité qu'ils ont défini par l'équation 2.12

$$com(K) = \sum_{j=1}^{m} \sum_{i=1}^{n-1} b(i,j) - b(i+1,j) * k(i,j) \tag{2.12}$$

où K est une image binaire de taille $m \times n$, $com(K)$ la fonction de complexité de K et $b(i,j)$ l'intensité du pixel (i,j). $k(i,j)$ est un coefficient de pondération dépendant du pixel, dont la valeur est grande pour les pixels intérieurs de l'image des yeux et faible pour les marges et l'espace entre les yeux. Puisque l'œil ouvert contient plus d'information (pupille et iris) que l'œil fermé, $com(K)$ est plus grande pour les yeux ouverts et la somnolence peut être détectée si cette fonction dépasse un seuil pour une certaine durée. Les tests effectués pour valider cette approche révèlent un CCR de 91%

L'avantage majeur du comptage des yeux fermés consécutifs à travers une séquence vidéo afin de détecter la somnolence est la simplicité de réalisation des techniques et leur adaptation aux conditions temps-réelles. Toutefois, il est nécessaire de bien délimiter les zones du visage et des yeux pour garantir le bon fonctionnement de l'algorithme d'analyse des yeux.

2.4 PERCLOS

L'une des mesures les plus récurrentes dans la littérature est le PERCLOS qui a été proposé pour quantifier le changement apparent sur le mouvement des paupières (Wierwille, 1994; Dinges et Grace, 1998). Cette mesure est basée sur le calcul du pourcentage de la fermeture de l'œil en fonction du temps et reflète la lourdeur des paupières au lieu du clignement des yeux. Le PERCLOS peut être obtenu par l'équation 2.13 (Jo *et al.*, 2011)

$$PERCLOS(k) = \frac{\sum_{i=k-n+1}^{k} c(i)}{n} \times 100 \tag{2.13}$$

où $PERCLOS(k)$ est le pourcentage des yeux fermés dans la frame k à travers la durée de mesure n. $c(i)$ correspond à l'ouverture ($c(i) = 0$) ou la fermeture ($c(i) = 1$) des yeux pour chaque frame i. Wierwille (1994) ont démontré qu'une valeur de PERCLOS excédant 80% pendant environ 3 minutes est révélatrice d'une lourdeur des paupières, et donc de somnolence.

Tous les travaux utilisant PERCLOS suivent les mêmes étapes qui consistent à délimiter la zone du visage et la région de chaque œil, puis à appliquer cette mesure pour détecter la somnolence. Qing *et al.* (2010) utilisent l'algorithme Adaboost pour localiser le visage et construisent des templates de chaque œil en se basant sur la position naturelle des yeux dans le visage. Ensuite, le suivi des templates est effectué et le PERCLOS est calculé sur toutes les frames de la vidéo. Le conducteur est alerté par une fatigue si le PERCLOS est supérieur à 40% et le temps de fermeture continue des yeux dépasse 3 secondes.

Grace (2001) a aussi utilisé le PERCLOS pour détecter la fermeture lente des paupières dans son système Copilot, construit pour la surveillance de l'état des conducteurs professionnels (chauffeurs de camion principalement). Dans ce travail, un éclairage infrarouge est utilisé pour produire l'effet brillant de la pupille et faciliter sa détection. Le PERCLOS est calculé sur des périodes de 3 minutes afin de détecter une fatigue modérée ($8\% \leq PERCLOS \leq 14\%$) ou sévère ($PERCLOS > 14\%$).

Après l'étude proposée par Dinges et Grace (1998), PERCLOS a longtemps été considérée comme une mesure standard, supérieure aux autres mesures (même l'EEG) et suffisante à l'estimation de l'hypovigilance. Cependant, Trutschel *et al.* (2011) ont effectué une étude détaillée et ont prouvé sur huit sessions de test, chacune durant une heure, que l'EEG (taux d'erreur de 13%) est meilleur que le PERCLOS (taux d'erreur de 35%). La conclusion que nous pouvons déduire de cette étude est qu'il est important d'utiliser plusieurs indicateurs afin de détecter l'hypovigilance et de ne pas se contenter d'une seule mesure même si elle est puissante.

2.5 Fréquence de clignement des yeux

La fréquence du clignement des yeux est considérée comme l'un des signes important de l'hypovigilance (Lal et Craig, 2001). Ce paramètre utilise généralement les dérivées temporelles de l'image pour la détection du mouvement, suivie par une analyse de l'image binaire pour retrouver l'état de l'œil (Benoit et Caplier, 2005). Plusieurs paramètres des clignements des yeux ont été étudiés. Noguchi *et al.* (2007) ont suggéré l'application des modèles de Markov cachés

« Hidden Markov Models » (HMM) sur les paramètres de durée, d'amplitude et de vitesse des clignements afin d'obtenir un classifieur à neuf niveaux de vigilance. Il s'agit également du cas de Omi *et al.* (2008), qui ont proposé une fonction d'estimation du niveau de vigilance, résultant de l'analyse par régression multiple sur différents paramètres tels que le PERCLOS, la fréquence du clignement et la durée de fermeture de l'œil.

Il existe des systèmes qui utilisent des mesures différentes de celles cités auparavant et puis d'autres qui fusionnent plusieurs paramètres pour obtenir de meilleurs résultats. Ji *et al.* (2004) proposent un système opérant sous éclairage infrarouge et combinant plusieurs paramètres extraits de l'œil. L'éclairage infrarouge facilite la détection de l'œil grâce à l'effet brillant de la pupille. Ensuite, l'œil est suivi en utilisant le filtre de Kalman. Son état est mesuré par le PERCLOS et la vitesse moyenne de fermeture de l'œil qui définit le temps nécessaire à sa fermeture/ouverture complète.

Dans (Senaratne *et al.*, 2007), quatre indicateurs de l'hypovigilance ont été combinés pour déterminer l'état du conducteur, à savoir : le PERCLOS, la fréquence d'inclinaison de la tête, la fréquence de courbure du dos et la fréquence d'ajustement de la posture. Les auteurs ont obtenu un taux d'erreur de 15.2% lors de l'utilisation de PERCLOS uniquement, contre un taux de 12.7% pour la fusion des quatre indicateurs.

Friedrichs et Yang (2010) effectuent une détection de l'hypovigilance basée sur le comportement des paupières. Le PERCLOS, le changement des distances entre les paupières et la fermeture des yeux sont utilisés pour cette fin.

2.6 Fréquence de bâillement

Le bâillement est un réaction typique induite par la fatigue. Il se traduit par une ouverture prolongée et incontrôlée de la bouche bien différente des autres déformations des lèvres. Mohanty *et al.* (2009) modélisent cette activité par une estimation du mouvement non rigide des lèvres et considèrent qu'un bâillement est observé si un mouvement particulier des lèvres est maintenu entre cinq et dix secondes. La technique proposée nécessite une série de pré-traitements de l'image (application d'un filtre médian, amélioration du contraste, binarisation, érosions et dilatations) afin de fournir une estimation initiale de la zone de recherche du contour de la bouche en état de bâillement. Ce contour est déterminé par un modèle de contour actif, qui correspond à une courbe déformable dans une image 2D. Pour distinguer entre le degré d'ouverture de la bouche pendant la parole et le bâillement, les auteurs ont effectué une étude expérimentale sur huit séquences où ces deux comportements sont simulés. Les résultats sont mesurés par le nombre de pixels de l'ouverture du contour de la bouche. Ils ont conclu que le degré d'ouverture de la bouche pendant le pique du bâillement excède trois ou quatre fois celui de la parole.

Fan *et al.* (2007) localisent et suivent le mouvement de la bouche du conducteur pour déterminer le bâillement. Ils détectent le visage en utilisant des templates de centre de gravité (Miao *et al.*, 1999), localisent les coins de la bouche par les projections du visage sur les plans horizontal et vertical de l'image, et extraient les caractéristiques de texture par les ondelettes de Gabor. L'analyse discriminante linéaire « Linear Discriminant Analysis » (LDA) est appliquée pour classifier les vecteurs caractéristiques de texture et détecter le bâillement. Le test effectué

sur 400 frames a fourni un CCR moyen de 95%.

Rongben *et al.* (2004) déterminent la zone du visage et de la bouche par l'analyse de la couleur de la peau. Ensuite, la bouche est détecté et les caractéristiques des lèvres sont calculées par l'analyse des composantes connectées. Le suivi de la bouche du conducteur est effectué par le filtre de Kalman. Les caractéristiques géométriques de la région de la bouche (la largeur et longueur maximale entre les extrémités de la bouche ainsi que la largeur entre les deux lèvres) sont utilisées pour la classification par réseaux de neurones à rétropropagation à trois états, à savoir l'état normal, la parole et le bâillement. Les résultats obtenus en testant cette approche sur un ensemble de 450 frames a fourni un CCR moyen de 96%.

Wang et Shi (2005) proposent de localiser la zone du visage par le détecteur de Viola-Jones (Viola et Jones, 2001), puis d'effectuer son suivi par le filtre de Kalman. Une méthode basée sur la projection niveau de gris est utilisée pour déterminer les deux coins de la bouches ainsi que les limites supérieure et inférieure des lèvres. Le degré d'ouverture de la bouche est ensuite calculé par le rapport de la longueur sur la largeur. Un bâillement est détecté si ce rapport dépasse un seuil de plus de 20 frames.

Saradadevi et Bajaj (2008) proposent une méthode pour localiser et suivre la bouche du conducteur en utilisant la cascade de classifieurs proposée par (Viola et Jones, 2001). Les SVM sont utilisées pour l'apprentissage du bâillement dans des images de la bouche afin de détecter la fatigue. Pour valider leur système, les auteurs ont acquis quelques vidéo et ont sélectionné environ 20 images de bâillement et plus de 1000 images normales pour le test. Pour chaque vidéo, 10 images représentant des individus en bâillement et 10 images normales sont données pour la cascade de classifieurs afin d'effectuer l'apprentissage pour la détection et le suivi de la bouche. Ces même images sont fournies aux SVM pour classifier les bouches. Les résultats obtenus en utilisant les images conservées pour le test ont produit un CCR de 83%.

Même si l'analyse des yeux reste le critère le plus utilisé pour déterminer l'état du conducteur puisqu'il permet de révéler la somnolence, l'utilisation du bâillement comme indicateur de la fatigue est très utile pour renforcer cette décision. En effet, l'analyse de la bouche permet de fournir de meilleurs résultats que l'analyse des yeux, puisque l'ouverture de la bouche liée au bâillement est beaucoup plus apparente dans une image.

2.7 Conclusion

Nous avons présenté dans ce chapitre diverses techniques utilisées pour la détection de l'état du conducteur à partir de l'analyse des yeux et de la bouche, dont nous résumons une grande partie dans le tableau 2.1. Nous pouvons déduire de cet état de l'art que la majorité de ces techniques suivent une architecture bien précise pour effectuer leurs décisions. En effet, nous distinguons quatre tâches nécessaires pour accomplir une détection de l'état du conducteur :

- Acquisition : capturer une vidéo du conducteur en utilisant une ou plusieurs caméra(s) à spectre visible et/ou sensible(s) à un éclairage infrarouge. Généralement, le matériel d'acquisition est fixé sur le tableau de bord en face du conducteur.
- Pré-traitement : extraire les données pertinentes à partir de la vidéo. Cette étape consiste très souvent à localiser la zone du visage, ainsi qu'à délimiter les régions des caractéris-

Tableau 2.1 – Tableau récapitulatif des approches de détection de la somnolence et de la fatigue chez le conducteur

Approches	**Critères**	**CCR**	**Données**
(D'Orazio *et al.*, 2004)	Ouverture/fermeture œil + CHT + Réseau de neurones	93%	1474 images
(Zhang *et al.*, 2008)	Templates œil + Degré d'ouverture	95%	245 images
(Horng *et al.*, 2004)	Suivi + Somnolence par couleur globe oculaire	98%	4 vidéos
(Tian et Qin, 2005)	Somnolence par mesure de complexité de l'œil	91%	156 images
(Qing *et al.*, 2010)	Template œil + PERCLOS	-	Vidéos
(Grace, 2001)	Éclairage IR + PERCLOS	-	Vidéos
(Senaratne *et al.*, 2007)	PERCLOS + Fréquence inclinaison tête + Fréquence courbure dos + Fréquence ajustement posture	88%	Vidéos
(Friedrichs et Yang, 2010)	Éclairage IR + PERCLOS + Distances entre paupières + Fermeture des yeux + Réseaux de neurones	83%	Vidéos
(Fan *et al.*, 2007)	Bâillement par suivi mouvement bouche + Ondelettes de Gabor + LDA	95%	400 frames
(Rongben *et al.*, 2004)	Bâillement par analyse des composantes connectées + suivi + Réseaux de neurones	96%	450 frames
(Saradadevi et Bajaj, 2008)	Bâillement par suivi + SVM	83%	1020 images

tiques à étudier (yeux et/ou bouche).

- Unité de diagnostique : analyser les caractéristiques faciales extraites dans l'étape précédente pour relever les différents états du conducteur. Cette étape est considérée comme le cœur du système.
- Décision : déterminer les états qui nécessitent l'émission des avertissements relatifs à l'hypovigilance du conducteur.

Ainsi, nous proposons dans le chapitre 3 deux approches qui respectent ces différentes étapes pour déterminer la somnolence et la fatigue chez le conducteur. Nous avons opté pour une conception basée sur l'étude de l'état d'ouverture/fermeture des yeux et de la bouche, puisque nous avons été motivé par la simplicité et l'efficacité de ce type d'approches. En effet, nous avons pensé à utiliser la technique de la CHT pour déterminer l'ouverture des yeux, mais aussi celle de la bouche. Nous désirons préciser que cette technique a été précédemment utilisée dans la littérature pour estimer l'état des yeux (D'Orazio *et al.*, 2004; Flores *et al.*, 2010; Devi *et al.*, 2011), mais n'a jamais été exploitée auparavant pour l'analyse de la bouche. Pour améliorer la performance de la CHT, nous avons conçu des détecteurs de contours originaux pour l'œil et pour la bouche basés sur leur morphologie. Les résultats obtenus par les deux approches sont satisfaisants, comme le prouve les séries de tests effectuées sur des séquences vidéo réelles.

Chapitre

Détection de la somnolence et de la fatigue basée sur la Transformée de Hough Circulaire

Sommaire

3.1 Introduction **27**
3.2 Localisation des zones d'intérêt **28**
3.2.1 Détection du visage par SVM 28
3.2.2 Localisation des yeux 29
3.2.3 Localisation de la bouche 32
3.3 Détection de la somnolence par l'analyse des yeux **32**
3.3.1 Transformée de Hough Circulaire 33
3.3.2 Transformée de Hough Circulaire pour la détection de l'iris 35
3.3.2.1 Détecteur du contour de l'iris 36
3.3.2.2 Application de la Transformée de Hough Circulaire 37
3.3.3 Détection de la somnolence chez le conducteur 39
3.4 Détection de la fatigue par l'analyse de la bouche **39**
3.4.1 Transformée de Hough Circulaire pour la détection du bâillement . . . 39
3.4.1.1 Détecteur du contour du bâillement 40
3.4.1.2 Application de la Transformée de Hough Circulaire 41
3.4.2 Détection de la fatigue chez le conducteur 42
3.5 Schéma général du système **42**
3.6 Résultats expérimentaux **44**
3.6.1 Mesures utilisées 44
3.6.2 Base de données personnelle pour détecter la fatigue et la somnolence 46
3.6.3 Évaluation de l'analyse des zones d'intérêt 47
3.6.4 Évaluation du système proposé 49
3.6.5 Résultats sur quelques systèmes existants 52
3.7 Conclusion **53**

3.1 Introduction

Dans le présent chapitre, nous détaillons la solution proposée pour détecter la somnolence et la fatigue chez le conducteur à partir de l'analyse des caractéristiques faciales. Avant de traiter ces problèmes, il est nécessaire d'isoler les zones d'intérêt à partir de l'image du conducteur, à savoir les régions des deux yeux et de la bouche. Ainsi, nous présentons dans la

section 3.2 la technique utilisée pour l'extraction du visage, suivie par celle permettant l'isolation des zones d'intérêt. Ensuite, nous développons les techniques proposées pour la détection de la somnolence (section 3.3) et de la fatigue (section 3.4) chez le conducteur. Enfin, nous présentons un ensemble de résultats expérimentaux dans la section 3.6

3.2 Localisation des zones d'intérêt

La procédure que nous utilisons pour localiser les régions d'intérêt des yeux et de la bouche opère sur une zone de visage en niveaux de gris, extraite en utilisant une bibliothèque basée sur les SVM et dont le fonctionnement est présenté dans la sous-section 3.2.1. Ensuite, nous détaillons notre procédure fondée essentiellement sur la structure générale du visage ainsi que sur la projection de ses pixels pour extraire à la fois l'œil gauche, l'œil droit (voir la sous-section 3.2.2) et la bouche (voir la sous-section 3.2.3).

3.2.1 Détection du visage par SVM

L'extraction du visage à partir d'une frame de la séquence vidéo du conducteur est la première étape que nous effectuons pour réduire la zone de recherche des caractéristiques faciales. Nous avons choisi d'appliquer une méthode de détection du visage conçue par Romdhani *et al.* (2001) et optimisée par Kienzle *et al.* (2005). Cette méthode basée sur les SVM, que nous avons présentés dans la sous-section 2.2.1, a donné naissance à la bibliothèque fdlib [1].

L'idée de base de la méthode représentée par la bibliothèque fdlib est d'appliquer une fenêtre sur toutes les positions, les échelles et les orientations de l'image. Ensuite, un SVM non linéaire est appliqué pour déterminer si un visage est contenu dans la fenêtre. Le SVM non linéaire compare le patch d'entrée à un ensemble de vecteurs supports qui peuvent être considérés comme des modèles de visage ou des modèles de non-visage. Un score est attribué au vecteur support par une fonction non linéaire pour chaque fenêtre. Un visage est détecté si la somme résultante dépasse un seuil.

Puisque l'espace de recherche est volumineux, un ensemble de vecteurs réduits est calculé à partir des vecteurs supports pour optimiser le temps de calcul des SVM. En effet, seul un sous-ensemble de vecteurs réduits est nécessaire pour éliminer les objets ne correspondant pas au visage. L'optimisation introduite par Kienzle *et al.* (2005) concerne l'ensemble des vecteurs supports remplacé par un ensemble réduit de points synthétisés de l'espace d'entrée. Contrairement aux méthodes qui réduisent cet ensemble par une optimisation sans contrainte, une contrainte structurelle est imposée sur les points synthétisés afin que les approximations résultantes puissent être évaluées par des filtres séparables. Ainsi, le rang défini par l'utilisateur, correspondant au nombre de filtres séparables dans lesquels les vecteurs réduits sont décomposés, fournit un mécanisme pour contrôler le compromis entre la précision et la vitesse de l'approximation.

Nous avons choisi d'utiliser cette méthode basée sur l'aspect pour son efficacité et sa capacité à atteindre un CCR de 95%, selon les auteurs. De plus, il n'est pas nécessaire de paramétrer

1. Le code source de fdlib est disponible sur le lien suivant `http://people.kyb.tuebingen.mpg.de/kienzle/facedemo/facedemo.htm`

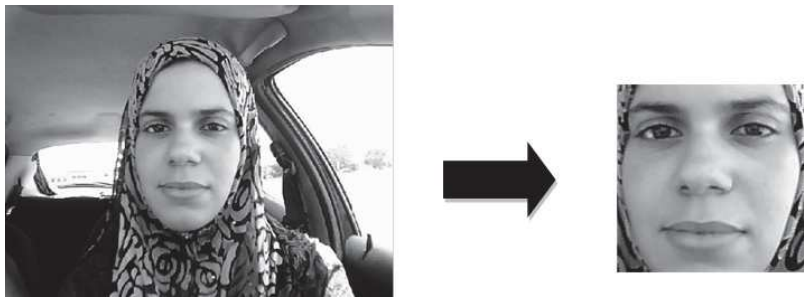


Figure 3.1 – Application de fdlib sur une frame réelle

manuellement la bibliothèque fdlib pour avoir de bons résultats. Même si fdlib permet de détecter plusieurs visages, nous allons nous restreindre à la détection d'un seul visage qui correspond au visage de dimension maximale représentant celui du conducteur. Avant d'appliquer fdlib, nous ajustons le contraste de la frame courante. La figure 3.1 présente le résultat obtenu sur une frame réelle de notre base de données, que nous détaillons dans la sous-section 3.6.2. Le résultat de cette étape est une zone contenant le visage du conducteur en niveaux de gris sur laquelle nous appliquons une méthode basée sur la géométrie du visage pour localiser la région de chaque œil, détaillée dans la sous-section 3.2.2, ainsi que la région de la bouche, présentée dans la sous-section 3.2.3.

3.2.2 Localisation des yeux

En premier lieu, nous réduisons la zone de recherche des yeux en fixant une limite inférieure et une limite supérieure au niveau du visage. Ceci est effectué par la procédure que nous proposons et qui se compose des sous-étapes suivantes :

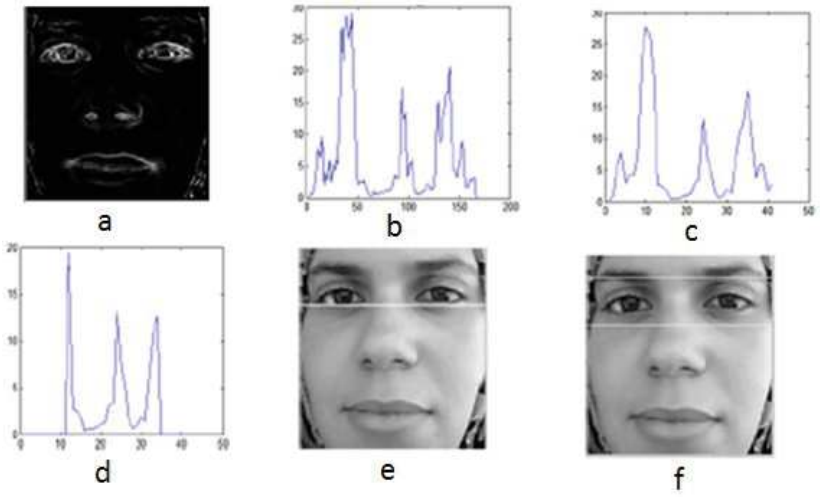


Figure 3.2 – Étapes de la localisation de la limite inférieure et de la limite supérieure des yeux. (a) Image Gradient ; (b) Projection horizontale ; (c) Projection horizontale lissée ; (d) Projection horizontale traitée ; (e) Niveaux des yeux ; (f) Zone des yeux.

– Nous calculons les gradients du visage selon les directions horizontale et verticale, notés G_x et G_y. Nous considérons un espacement important entre les points selon chaque direction

(fixé expérimentalement à 55), afin d'obtenir un contour fin. Par la suite, nous ferons référence au gradient du visage par :

$$G = \sqrt{G_x^2 + G_y^2}$$

L'image gradient ainsi obtenue est illustrée par la figure 3.2-a.

- Nous calculons la projection horizontale de l'image gradient (Figure 3.2-b). Un élément de la projection horizontale est la somme des pixels de la ligne i donnée par :

$$proj_h(i) = \sum_j (grad(i,j))$$

Où $grad$ symbolise l'image gradient et j représente l'indice d'une colonne.

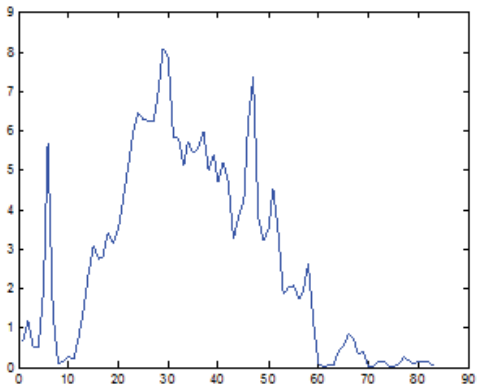
- Nous réduisons le nombre de pics de la projection horizontale par un lissage. Cela revient à remplacer chaque k éléments de la projection par leur moyenne :

$$proj_h(i) = \frac{\sum_{l=i}^{i+k} proj_h(l)}{k}$$

Ainsi, nous obtenons une réduction par k de la taille du vecteur de projection. La figure 3.2-c illustre un lissage avec $k = 4$, qui correspond au facteur que nous avons déterminé expérimentalement.

- Sachant que le niveau des yeux ne se situe ni dans l'extrémité supérieure ni dans l'extrémité inférieure du visage, nous avons proposé de procéder à la mise à zéro des éléments du vecteur de projection correspondants à ces parties du visage. Ainsi, les éléments situés avant la projection maximale du premier tiers du visage et les éléments situés après la projection maximale du dernier tiers sont annulés. Le résultat de ce traitement apparait dans la figure 3.2-d.
- Après les traitements effectués sur la courbe de projection horizontale, nous déterminons le niveau des yeux, illustré par la figure 3.2-e, comme étant le premier pic de la figure 3.2-d.
- La limite inférieure et la limite supérieure des yeux (Figure 3.2-f) sont obtenues en construisant un intervalle autour du niveau des yeux. Cet intervalle prend en considération la taille de l'image du visage.

L'étape de séparation de l'œil gauche et de l'œil droit s'effectue sur le gradient de la zone des yeux, illustré par la figure 3.3. Nous proposons d'effectuer cette étape comme suit :


0
1
2
3
4
5
6
7
8
9
0
10
20
30
40
50
60
70
80
90

3.2.3 Localisation de la bouche

La localisation de la bouche s'effectue sur la moitié inférieure du gradient du visage (Figure 3.8) et se compose des sous-étapes suivantes que nous avons élaborée :

3.3.1 Transformée de Hough Circulaire

- **Théorie de la Transformée de Hough**

Très peu d'attention a été portée dans la littérature à une définition convenable de la transformée de Hough « Hough Transform » (HT) (Duda et Hart, 1972). Il s'en suit une confusion fréquente sur le type d'opération effectivement réalisé, et une prolifération de termes imprécis pour qualifier des modifications mineures HT étendue, ou généralisée, ou modifiée, etc. Nous regroupons usuellement sous le nom de HT des transformations qui permettent de détecter dans des images la présence de courbes paramétriques appartenant à une famille connue, à partir d'un ensemble de points sélectionnés, appelés points caractéristiques. La HT utilise essentiellement l'information spatiale des points caractéristiques (leur position dans l'image), mais parfois, elle tient compte également de l'information contenue dans le signal de l'image lui-même, qui correspond à la valeur de la luminance en un point donné. Nous considérons que ce signal est une fonction scalaire : image en niveaux de gris, mais rien ne s'oppose à ce qu'il soit vectoriel : image couleur ou multispectrale. Nous désignerons par n la dimension de l'espace de définition de l'image. Soit $\mathbb{R}$ l'espace image, et $\mathbb{E}$ un ensemble de N points sélectionnés par un pré-traitement $\mathbb{E} = \{M_i, i = 1...N\}\epsilon\mathbb{R}$. Un point M de $\mathbb{R}$ est repéré par ses coordonnées x. Soit $\Omega \subset \mathbb{R}^p$ un espace de paramètres et $\mathbb{F}$ (équation 3.1) une famille de courbes dans $\mathbb{R}^n$ paramétrée par a.

$$\mathbb{F} = \{\{x : f(x, a) = 0, x\epsilon\mathbb{R}^n\}, : a\epsilon\Omega\} \tag{3.1}$$

La HT associée à la famille $\mathbb{F}$ est une transformation qui fait correspondre à l'ensemble $\mathbb{E}$ une fonction g définie sur Ω. Il existe donc de nombreuses HT, les deux principales sont les suivantes :

- **Transformation de m à 1** [2] :
  Soit m le nombre minimal de points de $\mathbb{R}$ définissant une courbe de $\mathbb{F}$. Soit $\mathbb{E}^{(m)}$ (équation 3.2) l'ensemble de tous les m-uplets issus de $\mathbb{E}$ avec $Card(\mathbb{E}^{(m)}) = C_N^m$.

$$\mathbb{E}^{(m)} = \{M_i^{(m)} = \{M_{i1}, M_{i2}, ..., M_{im} : M_{ik}\epsilon\mathbb{E}\}\} \tag{3.2}$$

  À tout m-uplet $M_i^{(m)}$ de $\mathbb{E}^{(m)}$ est associé une courbe de $\mathbb{F}$ de paramètre a_i. Soit $C(a)$ la fonction caractéristique de $\mathbb{R}^p$. La HT de m à 1 est définie par l'équation 3.3

$$g(a) = \sum_{M_i^{(m)}\epsilon\mathbb{E}^{(m)}} c(a - a_i) \tag{3.3}$$

- **Transformation de 1 à m** :
  Par tout point M_i de $\mathbb{R}^n$ passent m courbes de $\mathbb{F}$. Soit A_i l'ensemble des valeurs de a telles que $f(x_i, a) = 0 : A_i = \{a_k = f(x_i, a) = 0\}$. La HT de 1 à m est définie par l'équation 3.4

$$g(a) = \sum_{M_i^{(m)}\epsilon\mathbb{E}^{(m)}} \sum_{a_k\epsilon A^{(i)}} c(a - a_k) \tag{3.4}$$

2. L'expression HT de 1 à m et l'expression HT de m à 1 viennent de l'anglais « one to many » et « many to one ».

- En pratique la HT de 1 à m conduit à des calculs moins nombreux que la HT de m à 1, car elle évite une recherche combinatoire parmi les points. D'autre part, elle se prête bien à des implémentations rapides, par sa structure parallélisable. Pour ces raisons, elle correspond à la HT la plus utilisée, et le nom HT générique lui est souvent réservé.
- S'il existe dans l'image quelques représentants de $\mathbb{F}$, et si le pré-traitement a effectué une sélection judicieuse des points M, il est probable que la fonction g possédera pour quelques valeurs de a, soit des maximums marqués, soit une accumulation de valeurs. La tâche de reconnaissance de formes dans $\mathbb{R}^m$ est ainsi transformée en une tâche de recherche de maximums ou de recherche de nuages de valeurs, qui correspond à une tâche assez connue. Alors, la HT non seulement détecte la présence d'une ou plusieurs courbes mais également les identifie.
- En pratique, la fonction g est usuellement construite à partir d'une représentation quantifiée de l'espace Ω. À chaque cellule de quantification représentant une valeur de a, est associé un compteur. Ainsi, pour chaque m-uplet $M^{(m)}$, ou pour chaque point M, est incrémenté un ou plusieurs compteurs de Ω. En fin de transformation, la valeur de $g(a)$ est prise égale au compte associé à la cellule a .
- **Exemple : Détection de cercles**
  Un cercle de $\mathbb{R}^2$ est paramétré généralement par les coordonnées de son centre (a, b) et par son rayon r : $a = \{a, b, r\}; (x-a)^2 + (y-b)^2 = r^2$. Dans ce cas, $p = 3$ et l'espace de Hough Ω n'est pas aisément représentable. Dans la HT de m à 1, nous vérifions également que $m = 3$ et que le cardinal de $g(m)$ vaut $C_N^3 = \frac{1}{6}N(N-1)(N-2)$. La HT de m à 1 passe donc par la résolution de $O(N^3)$ équations du second degré, chacune donnant un centre et un rayon d'un cercle connaissant trois de ses points. Donc, il est nécessaire de disposer de trois points de $\mathbb{E}$ pour déterminer un point de Ω.

• Théorie de la Transformée de Hough Circulaire

La CHT est utilisée pour détecter les contours circulaires dans une image. Sachant que l'équation du cercle est donnée par : $r^2 = (x-a)^2 + (y-b)^2$, la représentation paramétrique du cercle peut être écrite sous la forme de l'équation 3.5. Afin de simplifier cette représentation, le rayon peut être considéré constant.

$$x = a + r\cos(\theta) \text{ et } y = b + r\sin(\theta) \tag{3.5}$$

Dans le but de déterminer des cercles dans une image, nous considérons que le pré-traitement à effectuer est la détection de contours et que tous les contours doivent être obtenus par un détecteur de contours efficace.

À chaque point du contour, un cercle est dessiné en gardant un rayon souhaité. Les coordonnées qui appartiennent au périmètre du cercle dessiné sont incrémentés et transmis par le biais d'un accumulateur. Après avoir dessiné des cercles pour chaque point du contour avec un rayon souhaité, les coordonnées correspondant au cercle sont incrémentés dans l'accumulateur. Ainsi, l'accumulateur contient des nombres correspondant au nombre de cercles passant par les coordonnées individuels. Les coordonnées qui apparaissent le plus souvent représentent le centre

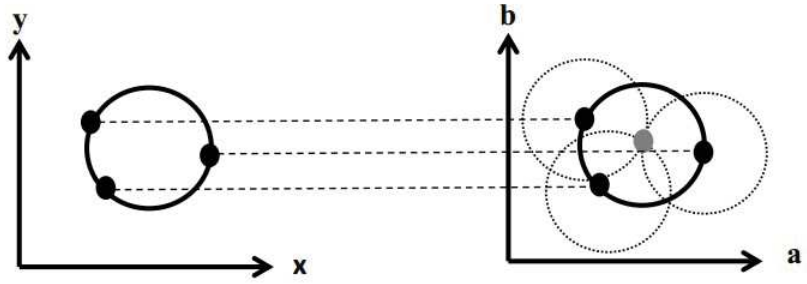
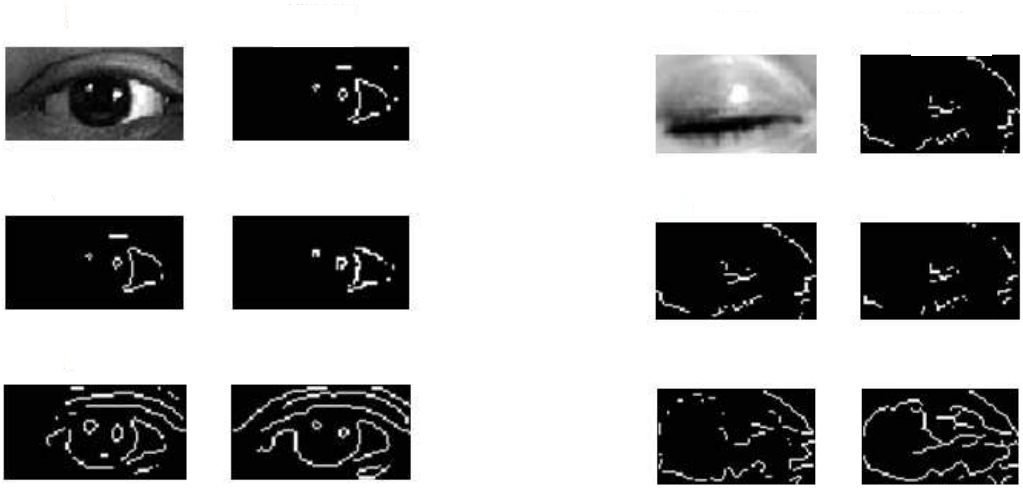


Figure 3.11 – Illustration de la détection du centre d'un cercle par la CHT

du cercle présent dans l'image. Le processus de détection du centre d'un cercle dans une image est illustré par la figure 3.11.

Avant d'appliquer la CHT, il est nécessaire de fixer la valeur estimée du rayon r du cercle recherché. Si la valeur du rayon n'est pas connue au préalable, il est possible de définir un intervalle de valeurs permises pour ce rayon.

3.3.2 Transformée de Hough Circulaire pour la détection de l'iris



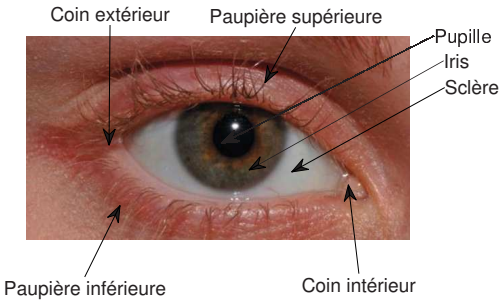


Figure 3.14 – Morphologie de l'œil

ouvert (Figure 3.12). Nous constatons également la présence de formes circulaires quand l'œil est fermé (Figure 3.13). Ainsi, nous déduisons que l'application de la CHT sur ce type de contours ne peut fournir de bons résultats. Pour cette raison, nous avons conçu un détecteur de contours adapté à la morphologie de l'œil, que nous présentons dans ce qui suit.

3.3.2.1 Détecteur du contour de l'iris

Le détecteur du contour de l'iris original que nous proposons se base sur la morphologie de l'œil, illustrée par la figure 3.14. En observant un œil ouvert, nous remarquons qu'il est constitué d'un disque interne sombre, qui correspond à la pupille, entouré par un disque plus grand et dont les intensités diffèrent selon la couleur des yeux de l'individu. Ce disque n'est autre que l'iris qui, à son tour, est entouré par la sclère dont la couleur est toujours plus claire que celle de l'iris. Finalement, le tout est entouré par les paupières inférieure et supérieure constituées de peau. Cette structure particulière de l'œil nous a permis de procéder à une extraction du contour de l'iris à partir des variations significatives entre les intensités de l'iris et de la sclère.

Pour construire le détecteur du contour de l'iris à partir de l'image de l'œil, nous considérons uniquement les pixels pouvant appartenir à l'iris, qui correspondent aux pixels x dont le niveau de gris est inférieur à un seuil (fixé expérimentalement à l'intensité moyenne des pixels). Pour chaque pixel x, nous déterminons un voisinage de n pixels à gauche et de n pixels à droite de x, puis nous calculons la différence entre x et ces voisins. Nous avons choisi de déterminer n par le nombre de colonnes de l'image de l'œil que nous divisons par 12.

- **Contour gauche :** si au moins $n-1$ voisins gauches produisent une différence supérieure à un seuil th_{sup} et au moins $n-1$ voisins droits fournissent une différence inférieure à un seuil th_{inf}, nous déduisons que le pixel x appartient au contour gauche de l'iris et nous le mettons à 1.
- **Contour droit :** si au moins $n-1$ voisins gauches fournissent une différence inférieure à th_{inf} et au moins $n-1$ voisins droits produisent une différence supérieure à th_{sup}, nous concluons que le pixel x appartient au contour droit de l'iris et nous le mettons à 1.
  La figure 3.15 illustre l'emplacement des contours gauche et droit dans une image de l'œil ouvert.
- **Interprétation :** Un pixel du contour gauche possède des voisins gauches d'intensité beaucoup plus grande, tandis que ses voisins droits ont une intensité presque identique à

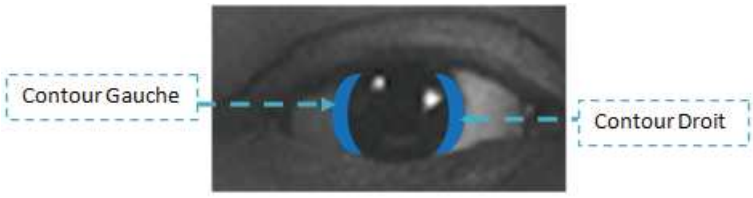


Figure 3.15 – Illustration des contours gauche et droit de l'œil ouvert

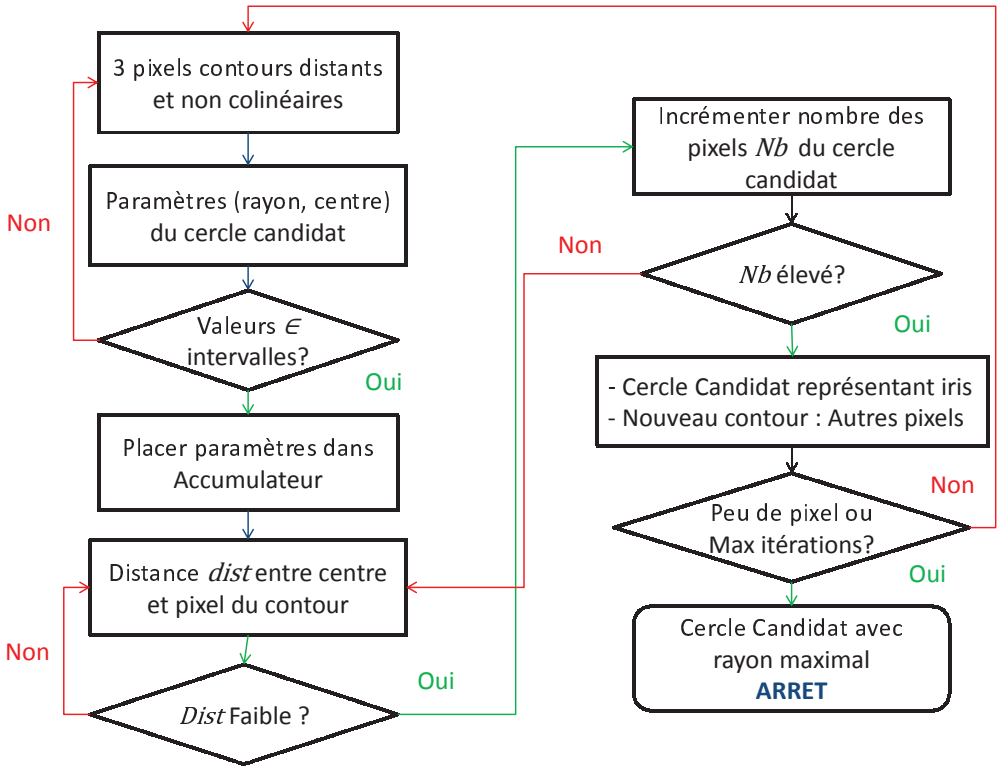
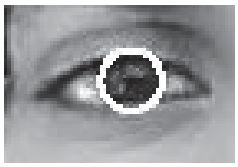


Figure 3.17 – Algorithme de la détection de l'iris par la CHT

s'ils sont compris entre des seuils prédéfinis qui détermineront la taille maximale et minimal du cercle candidat.

- Nous calculons la distance entre les coordonnées du centre et tous les pixels appartenant au contour. Si cette distance est inférieure à un seuil (égal à 3), nous incrémentons le compteur des pixels appartenant au cercle candidat. Si ce compteur est supérieur à un seuil ($\pi * rayon$), nous considérons que le cercle peut représenter l'iris. Nous retenons les pixels qui ne lui appartiennent pas comme nouvel ensemble des contours, à partir duquel nous sélectionnons trois nouveaux pixels pour représenter un nouveau cercle candidat.
- L'algorithme s'arrête quand l'ensemble des contours contient peu d'éléments ou quand le nombre maximal d'itérations toléré est atteint. Puisqu'on désire déterminer le cercle représentant l'iris, nous sélectionnons le cercle de plus grand rayon.

La figure 3.18 expose le résultats de la détection de l'iris par la CHT. Dans la sous-section 3.3.3, nous présentons notre algorithme de prise de décision en ce qui concerne la somnolence chez le conducteur.



3.3.3 Détection de la somnolence chez le conducteur

Puisque nous caractérisons la somnolence chez le conducteur par les périodes de micro-sommeil, nous déterminons la présence d'intervalles de micro-sommeil d'une durée minimale de 2 secondes comme suit :

- Pour réduire le temps de calcul, nous vérifions en premier lieu l'état de l'œil gauche. Si l'iris est présent au niveau de cet œil, nous effectuons l'étape de détection de la fatigue que nous détaillons dans la section 3.4.
- Si l'iris n'est pas détecté au niveau de l'œil gauche, nous déterminons l'état de l'œil droit.
- Si l'œil droit contient l'iris, nous considérons qu'une erreur de détection s'est produite au niveau de la détection de l'iris de l'œil gauche et nous effectuons l'étape de détection de la fatigue (section 3.4).
- Si l'œil droit ne contient pas d'iris, nous incrémentons le compteur du nombre de frames consécutives contenant les deux yeux fermés.
- Quand ce compteur dépasse un seuil (correspondant à 2 secondes de fermeture consécutive des deux yeux), nous émettons une alarme pour indiquer que le conducteur est somnolent.

Il est vrai que l'état le plus critique de baisse de vigilance est la somnolence. Toutefois, en général, la fatigue précède la somnolence. Ainsi, nous avons introduit la détection de la fatigue à partir de la fréquence de bâillement comme second critère pour prévenir le conducteur avant que la somnolence ne se produise.

3.4 Détection de la fatigue par l'analyse de la bouche

La fatigue chez le conducteur est un processus cumulatif entraînant des difficultés croissantes à poursuivre la conduite automobile et allant jusqu'à une baisse des performances. Puisque la fatigue est caractérisée par le fait de bâiller assez souvent, nous avons choisi de détecter les bâillements émis par le conducteur pour relever un état de fatigue. On sait que généralement, au cours du bâillement, la bouche reste grande ouverte entre 2 à 10 secondes. Ainsi, pour détecter un bâillement, nous appliquons la CHT sur l'image de la bouche afin de détecter une grande ouverture de celle-ci.

3.4.1 Transformée de Hough Circulaire pour la détection du bâillement

Comme nous l'avons défini dans la section 3.3, la CHT extrait les cercles en opérant sur une image obtenue après une détection des contours permettant de relever des points caractéristiques de la forme que nous désirons mettre en évidence. Nous avons ainsi testé plusieurs détecteurs de contours connus (Sobel, Prewitt, LoG et Canny) sur des bouches fermées, peu ouvertes et grandes ouvertes, comme le montre la figure 3.19. Cependant, nous avons obtenu des formes circulaires dans tous les cas. Pour cette raison, nous avons conçu un détecteur de contours pour distinguer la grande ouverture présente dans la bouche lors d'un bâillement. Ce détecteur de contours est très proche de celui que nous avons avons conçu pour détecter le contour de l'iris dans la sous-section 3.3.2.1.

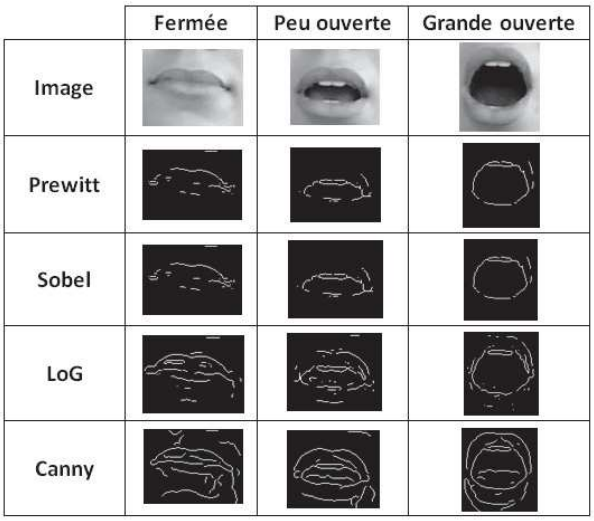


Figure 3.19 – Application des détecteurs de contours standards sur des bouches fermées, peu ouvertes et grandes ouvertes

3.4.1.1 *Détecteur du contour du bâillement*

Le détecteur du contour que nous proposons est original et se base sur la morphologie de la bouche lors du bâillement. Dans ce cas, la bouche présente une grande surface sombre possédant une forme pseudo-circulaire, comme illustré par la figure 3.20. Cette surface est délimitée par des régions de peau correspondant aux lèvres inférieures et supérieures, et éventuellement, une partie des dents. Ainsi, nous pouvons exploiter la grande variation des intensités entre la zone sombre du bâillement et la peau des lèvres (ou les dents) pour construire le détecteur du contour de la grande ouverture de la bouche. Comme pour le détecteur du contour de l'iris, nous considérons

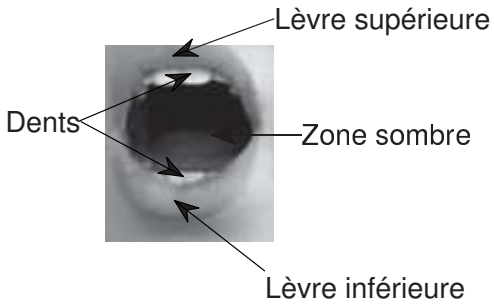


Figure 3.20 – Structure de la bouche lors du bâillement

uniquement les pixels x de l'image de la bouche susceptibles d'appartenir à la grande ouverture de celle-ci. Pour chaque pixel x, nous déterminons un voisinage de n pixels en haut et de n pixels en bas de x, puis nous calculons la différence entre x et ces n pixels voisins en haut et en bas. Le nombre de voisins n est déterminé par un cinquième du nombre des colonnes de l'image de la bouche.

- **Contour supérieur :** si au moins $n-1$ voisins hauts fournissent une différence supérieure à un seuil th_{sup} et si au moins $n-1$ voisins bas produisent une différence inférieure à un

seuil th_{inf}, nous déduisons que le pixel x appartient au contour supérieur de l'ouverture de la bouche et nous le mettons à 1.

– **Contour inférieur :** si nous disposons d'au moins $n-1$ voisins hauts produisant une différence inférieure à th_{inf} et d'au moins $n-1$ voisins bas fournissant une différence supérieure à th_{sup}, nous déduisons que le pixel x appartient au contour inférieur de l'ouverture de la bouche et nous le mettons à 1.
La figure 3.21 illustre les contours supérieur et inférieur de la bouche pendant le bâillement.

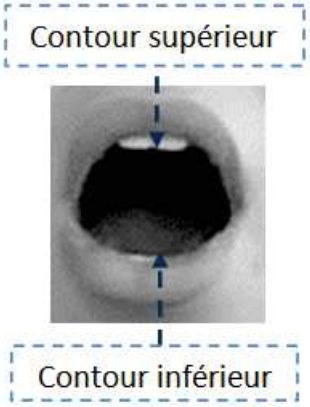


Figure 3.21 – Illustration des contours supérieur et inférieur de la bouche lors du bâillement

– **Interprétation :** Dans le cas d'un pixel du contour supérieur, nous remarquons que l'intensité de ses voisins hauts est très élevée par rapport à la sienne, tandis que l'intensité de ses voisins bas est presque identique à la sienne. Cette remarque se voit inversée pour un pixel du contour inférieur, dont l'intensité est très proche de celle de ses voisins hauts, tandis que l'intensité de ses voisins bas est très distincte de la sienne. Ainsi,le seuil th_{sup} doit être choisi afin de mettre en évidence la différence entre l'intensité des pixels de l'ouverture de la bouche et ses voisins qui appartiennent à la peau ou aux dents. Nous avons défini ce seuil par la différence entre la plus grande intensité et la plus faible intensité des pixels de l'œil divisée par 8. En ce qui concerne le seuil th_{inf}, il doit favoriser la ressemblance entre l'intensité des pixels appartenant à l'ouverture de la bouche. Nous avons donc fixé ce seuil à 10.

La figure 3.22 expose les résultats obtenus après l'application de notre détecteur de contours sur des bouches fermées, peu ouvertes et grandes ouvertes.

3.4.1.2 Application de la Transformée de Hough Circulaire

Après avoir déterminé le détecteur de contours adapté à notre problème de détection du bâillement, nous pouvons lui appliquer la CHT pour obtenir le rayon de l'ouverture à partir duquel nous décidons si la bouche est grande ouverte ou non. La CHT que nous appliquons est celle défini dans la sous-section 3.3.2.2. Puisqu'on désire déterminer le cercle représentant la grande ouverture de la bouche, nous sélectionnons le cercle de plus grand rayon pour représenter celle-ci. La figure 3.23 expose un exemple de la détection de la grande ouverture de la bouche par la CHT.

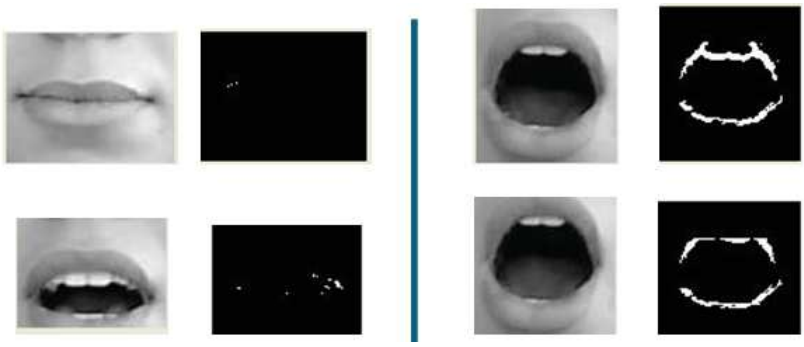


Figure 3.22 – Contours du bâillement par le détecteur proposé

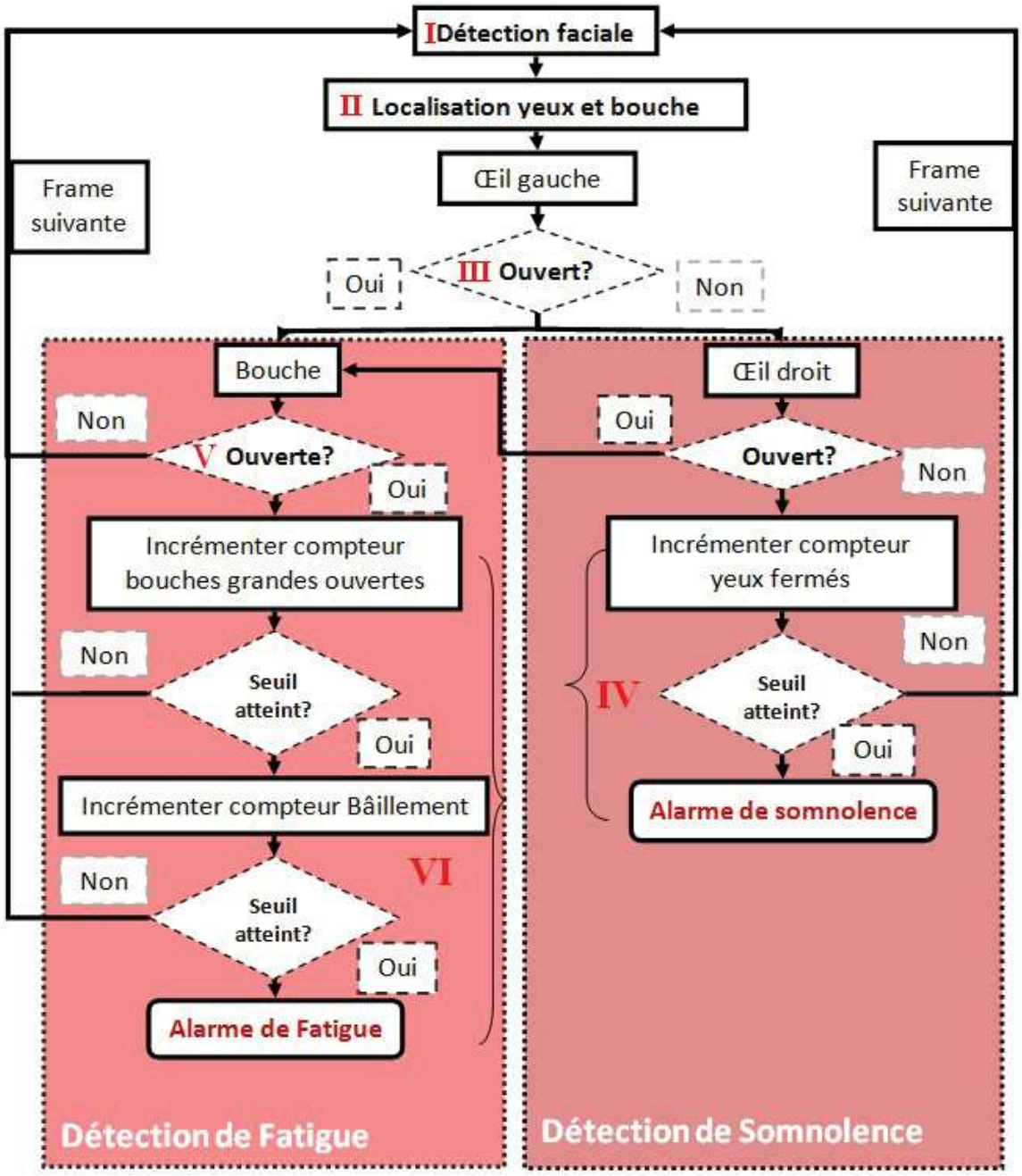


Figure 3.24 – Schéma général de la détection de la fatigue et de la somnolence chez le conducteur

I Détection du visage par la bibliothèque fdlib (sous-section 3.2.1).

II Localisation des yeux et de la bouche (sous-section 3.2.2 et sous-section 3.2.3).

III Vérification de l'état de l'œil gauche (sous-section 3.3.3) :

IV S'il est fermé, vérification de l'état de l'œil droit. S'il est aussi fermé, comptage du nombre de frames consécutives présentant les deux yeux fermés. Une fois ce compteur correspond à une fermeture de plus de 2 secondes, émission d'une alerte de somnolence.

V Si l'un des yeux est ouvert, vérification de l'état de la bouche (sous-section 3.4.2) :

VI Si elle est grande ouverte, comptage du nombre de bouches consécutives en état de bâille-

ment. Si ce compteur correspond à une durée supérieure à 2 secondes, incrémentation du nombre de bâillements et émission d'une alerte de fatigue quand ce nombre est important.

3.6 Résultats expérimentaux

Dans cette section, nous discutons les résultats obtenus en testant sous Matlab une implémentation non optimisée des méthodes proposées pour l'analyse des yeux et de la bouche par la CHT. Nous utilisons un processeur Intel Core2Duo pour tous les tests. Les résultats sont présentés par les mesures que nous définissons brièvement dans la sous-section 3.6.1.

3.6.1 Mesures utilisées

- **Matrice de confusion**

La matrice de confusion, dans la terminologie de l'apprentissage supervisé, est un outil servant à mesurer la qualité d'un système de classification. Chaque colonne de la matrice correspond au nombre d'occurrences d'une classe estimée, tandis que chaque ligne représente le nombre d'occurrences d'une classe réelle. Un des intérêts de la matrice de confusion est qu'elle permet de visualiser rapidement si le système parvient à effectuer une bonne classification des instances.

Dans le cadre du système que nous proposons pour la détection de la fatigue et de la somnolence chez le conducteur, nous disposons de deux méthodes de classification des images. La première méthode détecte les micro-sommeils en se basant sur l'analyse des yeux pour relever la présence de l'iris en utilisant la CHT (section 3.3). Pour évaluer cette méthode, nous utilisons la matrice de confusion présentée par le tableau 3.1. Cette matrice permettra de déterminer si chaque œil étudié est ouvert ou fermé. Les colonnes représentent les valeurs estimées par la méthode, tandis que les lignes correspondent à la vérité terrain. Nous expliquons, par les points

Tableau 3.1 – Matrice de confusion de l'analyse de l'œil

	Classe estimée		
Classe réelle	**œil ouvert**	**œil fermé**	**Total**
œil ouvert	VP	FN	P
œil fermé	FP	VN	N
Total	p	n	T

suivants, les valeurs présentées dans le tableau 3.1

- Vrai Positif (VP) : l'œil est ouvert et la méthode l'a détecté ouvert.
- Faux Négatif (FN) : l'œil est ouvert mais la méthode ne l'a pas détecté.
- Faux Positif (FP) : l'œil est fermé mais la méthode a détecte qu'il est ouvert.
- Vrai Négatif (VN) : l'œil est fermé et la méthode l'a détecté fermé.
- $P = VP + FN$: nombre total des yeux réellement ouverts.
- $N = VN + FP$: nombre total des yeux réellement fermés.
- $T = N + P$: nombre total des échantillons.
- $p = VP + FP$: nombre total des yeux ouverts selon la méthode.

– $n = VN + FN$: nombre total des yeux fermés selon la méthode.

La deuxième méthode que nous proposons permet de détecter les bâillements en analysant la bouche par la CHT, afin de déterminer la présence d'une grande ouverture de celle-ci. Nous utilisons la matrice de confusion présentée par le tableau 3.2 pour évaluer la méthode de détection du bâillement. Son interprétation est donnée par les points suivants :

Tableau 3.2 – Matrice de confusion de l'analyse de la bouche

	Classe estimée		
Classe réelle	**B. grande ouverte**	**B. fermée**	**Total**
B. grande ouverte	VP	FN	P
B. fermée	FP	VN	N
Total	p	n	T

– VP : la bouche est grande ouverte et la méthode l'a détectée.
– FN : la bouche est grande ouverte mais la méthode ne l'a pas détectée.
– FP : pas de grande ouverture mais la méthode en a détectée une.
– VN : pas de grande ouverture et la méthode n'en a pas détectée.
– $P = VP + FN$: nombre total de bouches présentant réellement une grande ouverture.
– $N = VN + FP$: nombre total de bouches réellement fermées.
– $T = N + P$: nombre total des échantillons.
– $p = VP + FP$: nombre total de bouches grande ouvertes estimées par la méthode.
– $n = VN + FN$: nombre total de bouches fermées estimées par la méthode.

- **Statistiques extraites de la matrice de confusion**

Taux de bonne classification

Le taux de Bonne Classification « Correct Classification Rate » (CCR) est la somme des bonnes détections représentées par VP et VN, divisée par le nombre total des échantillons T.

$$CCR = \frac{VN + VP}{T} \tag{3.6}$$

Coefficient kappa

En statistiques, le coefficient Kappa (κ) mesure l'accord entre les observateurs lors d'un codage qualitatif en catégories. Le calcul du coefficient κ se fait par l'équation 3.7

$$\kappa = \frac{P_0 - P_e}{1 - P_e} \tag{3.7}$$

où P_0 représente la proportion d'accord observé qui correspond au CCR et P_e représente la proportion d'accord aléatoire donnée par l'équation 3.8

$$P_e = \frac{1}{T^2}\left[(P \times p) + (N \times n)\right] \tag{3.8}$$

Nous avons toujours $-1 < \kappa < 1$. Le tableau 3.3 est utilisé pour interpréter le coefficient κ.

Tableau 3.3 – Interprétation du coefficient κ

Kappa	**Interprétation**
$\kappa \geq 0.81$	Accord excellent
$0.61 \leq \kappa \leq 0.8$	Accord fort
$0.41 \leq \kappa \leq 0.6$	Accord modéré
$0.21 \leq \kappa \leq 0.4$	Accord faible
$0.0 \leq \kappa \leq 0.2$	Accord très faible
$\kappa < 0$	Désaccord

3.6.2 Base de données personnelle pour détecter la fatigue et la somnolence

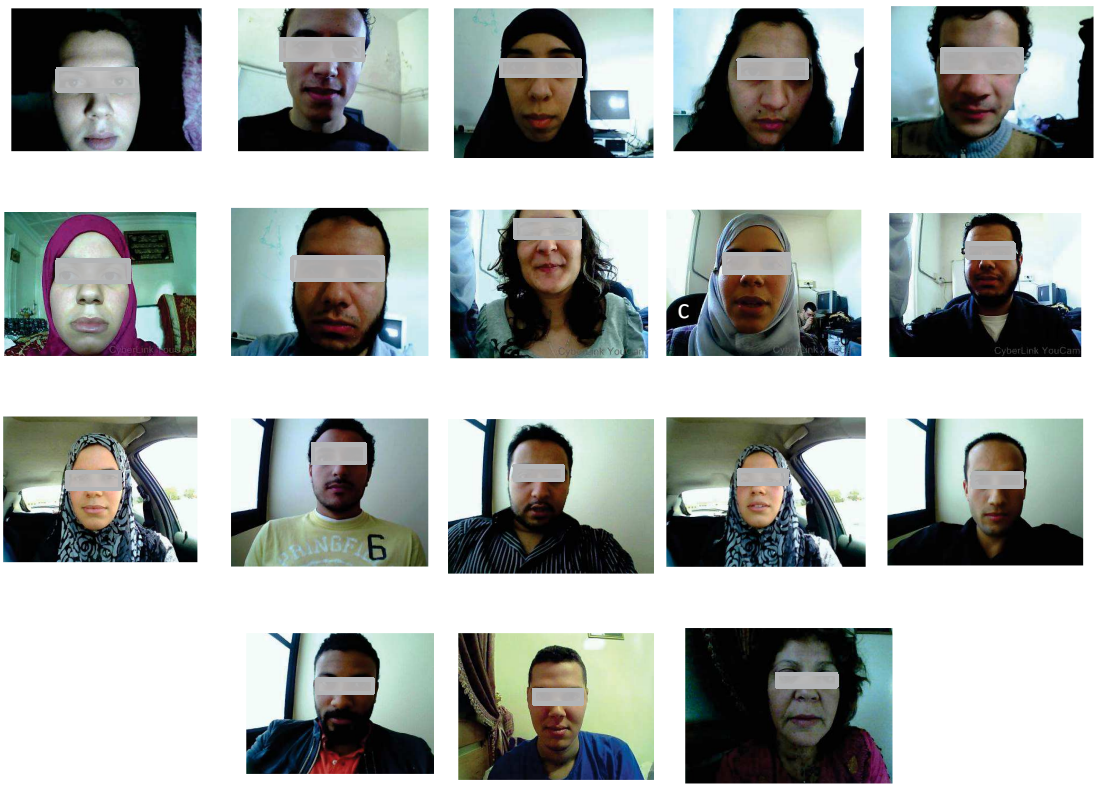


Figure 3.25 – Les 18 séquences de la base de données personnelle pour détecter la fatigue et la somnolence

Puisqu'il n'existe aucune base de données représentant les états de fatigue et de somnolence chez le conducteur, nous avons acquis et annoté 18 séquences vidéo de sujets simulant ces états sous différentes conditions d'éclairage. Toutes les séquences sont prises avec la même web camera à très faible coût fournissant des images de résolution 640×480, avec une cadence de 30 frames par seconde et un nombre total de frames environnant les 20000. La figure 3.25 affiche un exemple de chaque frame des 18 séquences de la base de données personnelle dédiée à la détection de la fatigue et de la somnolence. Pour des raisons de respect de la vie privée des sujets, nous

avons affiché les images en masquant la partie des yeux. Nous présentons ci-dessous quelques caractéristiques des différentes séquences de notre base de données :

- Douze sujets sont représentés dans notre base de données : huit sujets masculins (représentés par dix séquences) et quatre sujets féminins (représentés par huit séquences).
- Les séquences E, G, J, L, M et P présentent une pilosité faciale (barbe et/ou moustache) de différentes intensités.
- Les séquences A, Q et R sont acquises pendant la nuit en utilisant un éclairage artificiel apporté par de petites lampes embarquées sur la web caméra comme illustré par la figure 3.26, tandis que toutes les autres séquences sont acquises sous un éclairage ambiant pendant divers moments de la journée.

Figure 3.26 – Web caméra avec un éclairage intégré

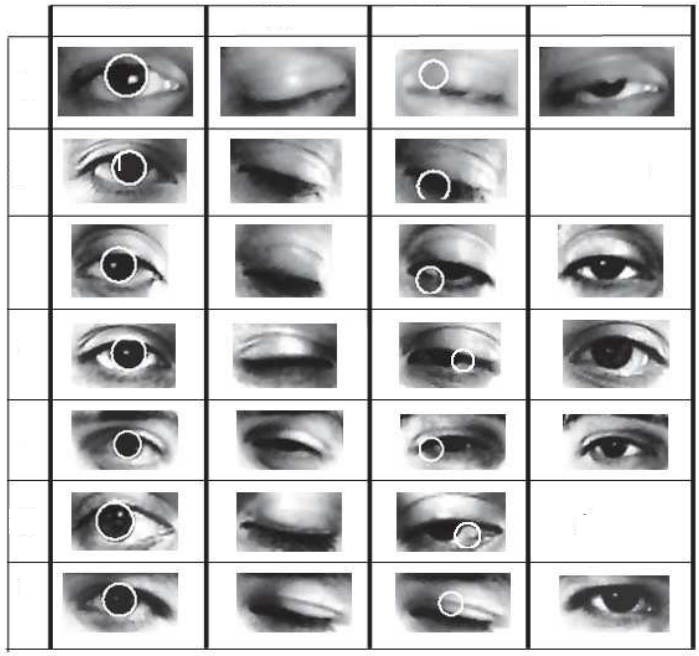
Dans la sous-section 3.6.3, nous présentons les premiers tests réalisés pour évaluer séparément les méthodes proposées basées sur la CHT pour l'analyse des zones d'intérêt, à savoir les yeux et la bouche. Ensuite, nous évaluons dans la sous-section 3.6.4 la détection de la fatigue et de la somnolence chez le conducteur.

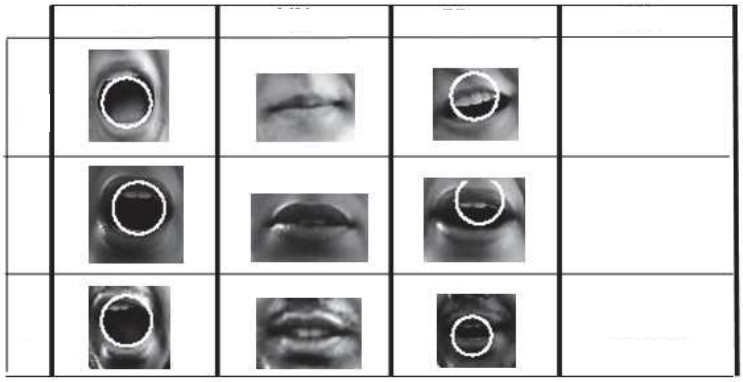
3.6.3 Évaluation de l'analyse des zones d'intérêt

- **Évaluation de l'analyse de l'œil**

Dans ce premier test, nous préférons effectuer une sélection manuelle des yeux pour évaluer uniquement la performance de la méthode proposée. Ainsi, nous étudions l'ouverture de la totalité des yeux gauches et droits sélectionnés manuellement à partir des frames des sept séquences considérées pour ce test. Nous exposons dans la figure 3.27 un exemple de l'œil pour chacun des VP, FP, VN et FN après l'application de la méthode d'analyse de l'œil sur les séquences de A à G de la figure 3.25. Le cercle blanc dans l'image de l'œil correspond à l'iris détecté par notre méthode. Nous détaillons les résultats du test dans le tableau 3.4. Ce tableau reporte les valeurs de la matrice de confusion (VP, FP, VN et FN), le nombre total d'images d'œil (T), ainsi que le CCR et le coefficient κ pour chaque séquence vidéo (Seq).

Dans la dernière ligne du tableau, nous calculons la moyenne (Moy) du CCR et du coefficient κ sur les sept séquences de test. Nous précisons que le nombre total des yeux étudiés dans ce test est de 13224 (6612 frames). D'après le tableau 3.4, le CCR moyen est de 98% et le coefficient κ moyen est de 88%. Cette dernière valeur signifie que l'accord est presque parfait entre les échantillons, comme précisé par le tableau 3.3. Ainsi, nous pouvons conclure que la méthode

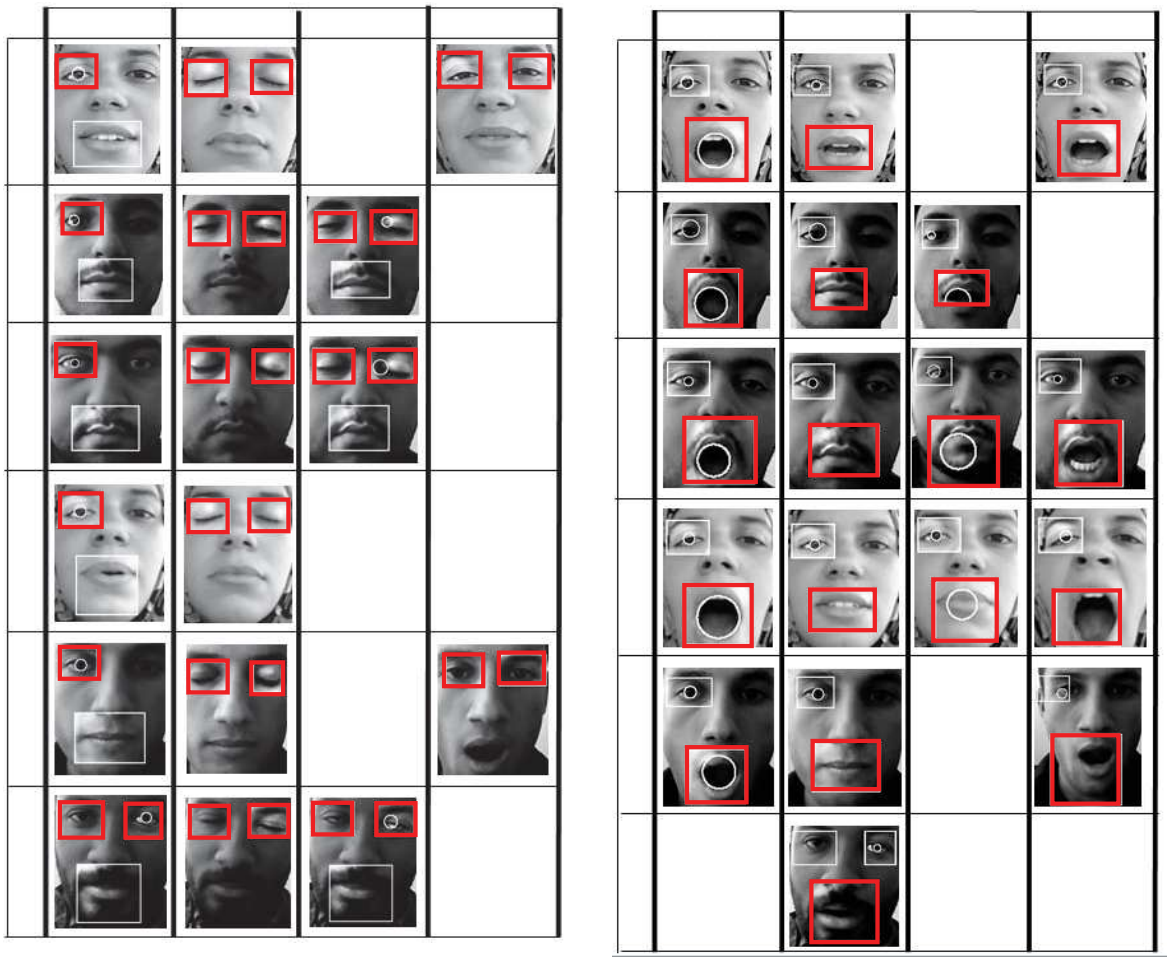




deux frames. Ce nombre de frames permet de conserver la quantité d'informations nécessaire à l'analyse de l'état du conducteur. Nous utilisons les séquences de K à P de la figure 3.25. Les séquences K et N sont acquises dans une voiture, comme illustré par la figure 3.29.

Nous présentons dans la figure 3.30 et la figure 3.31 respectivement, un exemple des VP, FP, VN et FN pour les méthodes de l'analyse de l'œil et de la bouche après l'application de notre système de détection de la fatigue et de la somnolence sur les 6 séquences de test acquises sous la lumière ambiante du jour, correspondant à environ 9150 frames. La zone d'intérêt de la caractéristique faciale impliquée dans chaque méthode est représentée en rouge dans la zone du visage et le cercle détecté est représenté en blanc.

Comme pour les tests présentés dans la sous-section 3.6.3, nous exposons les résultats de chaque séquence dans le tableau 3.6, qui contient les valeurs des VP, FP, VN et FN, le nombre total d'images (T), le CCR et le coefficient κ, ainsi que la durée de la séquence vidéo **DV** et le temps d'exécution de tout le système **TE** en secondes.



succession de frames dans lesquelles les deux yeux sont fermés pour au moins deux secondes, alors que la fatigue est reflétée par au moins deux bâillements, qui s'illustrent par des bouches grandes ouvertes présentes dans des frames successives pour la même durée. En d'autres termes, il n'est pas nécessaire de détecter la somnolence et la fatigue, mais il faudra plutôt évaluer l'efficacité des méthodes d'analyse de l'œil et de la bouche. En considérant ces observations, chaque colonne de 2 à 8 du tableau 3.6 est divisée en deux sous-colonnes, la première correspond aux résultats concernant l'analyse de l'œil et la seconde à ceux concernant l'analyse de la bouche. Dans l'avant dernière ligne du tableau, nous calculons la moyenne (Moy) du CCR et du coefficient κ pour chaque méthode du système, tandis que dans la dernière ligne, nous présentons la moyenne pour ces deux mesures sur tout le système.

Tableau 3.6 – Résultats de l'évaluation du système de détection de la fatigue et de la somnolence sous la lumière ambiante du jour

Seq	**VP**		**VN**		**FP**		**FN**		**T**		**CCR**		κ		**DV**	**TE**
1	99	30	23	68	0	0	1	1	124	99	0.98	0.99	0.94	0.97	61	57
2	82	31	11	52	2	1	0	0	95	84	0.98	0.99	0.90	0.97	47	48
3	68	8	19	59	1	1	0	1	89	69	0.98	0.97	0.94	0.87	41	40
4	109	37	36	68	0	2	0	2	145	109	1.00	0.96	1.00	0.92	72	68
5	57	18	18	39	0	0	2	0	77	57	0.97	1.00	0.93	1.00	40	39
6	73	0	14	90	3	0	0	0	90	0	0.96	1.00	0.88	1.00	44	45
								Moy.			0.97	0.98	0.93	0.95		
								Moy. totale			0.97		0.94			

Nous pouvons remarquer à partir du tableau 3.6 que le nombre des FP ne dépasse pas trois pour la méthode d'analyse de l'œil. Cette mesure est critique puisqu'elle représente le nombre des yeux fermés détectés ouverts par le système. Une valeur élevée des FP peut révéler la non détection des périodes de micro-sommeil, ce qui augmente la probabilité d'avoir un accident provoqué par la somnolence. Le nombre des FN ne dépasse pas deux, il correspondent aux yeux ouverts détectés comme fermés par la méthode. Cette mesure n'est pas aussi critique que les FP puisqu'elle ne produit que des fausses détections de micro-sommeils. Pour l'analyse de la bouche, le nombre des FP et des FN ne dépasse pas deux. Plus ces nombres augmentent, plus le risque d'avoir des non détections ou de fausses détections du bâillement est important. Nous remarquons aussi que le CCR moyen est 97% et que le coefficient κ moyen est 93% pour l'analyse de l'œil et de la bouche. Les temps d'exécutions sont très proches des durées des vidéos, ce qui implique que le système peut s'exécuter en temps-réel même si l'implémentation n'est pas optimisée. Dans la figure 3.32 et la figure 3.33, nous montrons un exemple des VP, FP, VN et FN pour les méthodes de l'analyse de l'œil et de la bouche après l'application de notre système de détection de la fatigue et de la somnolence sur deux séquences de test acquises sous une lumière artificielle pendant la nuit (Figure 3.25- Q et R), correspondant à environ 1700 frames. Comme pour le tableau 3.6, nous affichons dans le tableau 3.7 les valeurs des VP, FP, VN et FN, le nombre total d'images (T), le CCR et le coefficient κ. Chaque mesure est divisée en deux colonnes, la première correspond aux résultats concernant l'analyse de l'œil et la seconde à ceux concernant l'analyse de la bouche. Dans l'avant dernière ligne du tableau, nous calculons la

moyenne (Moy) du CCR et du coefficient κ pour chaque méthode du système, tandis que dans la dernière ligne, nous présentons la moyenne pour ces deux mesures sur tout le système.

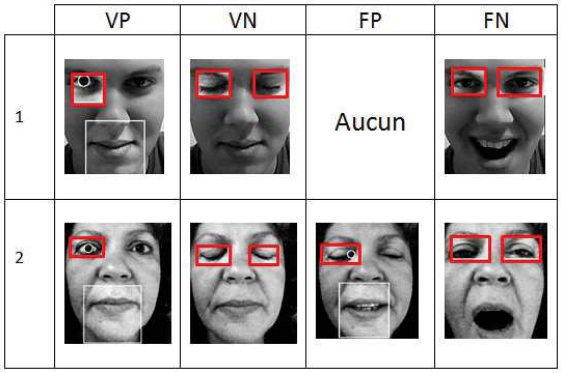


Figure 3.32 – Résultats de l'analyse de l'œil par le système de détection de la fatigue et de la somnolence sous un éclairage artificiel pendant la nuit

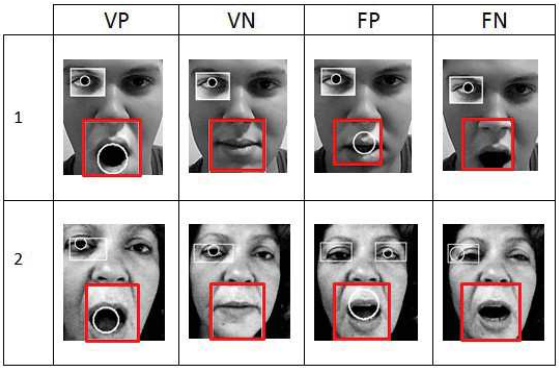


Figure 3.33 – Résultats de l'analyse de la bouche par le système de détection de la fatigue et de la somnolence sous un éclairage artificiel pendant la nuit

Nous remarquons que pendant la nuit et sous un éclairage artificiel (Tableau 3.7), les statistiques se dégradent comparées aux résultats obtenus sous la lumière ambiante du jour (Tableau 3.6). Ainsi, le CCR est réduit de 12% pour l'analyse de l'œil et de 6% pour l'analyse de la bouche, alors que le coefficient κ perd jusqu'à 27% et 12% pour ces deux méthodes, respectivement. Nous pouvons expliquer cette baisse des performances par l'influence de l'éclairage utilisé sur la qualité de l'image, et notamment sur les techniques d'extraction des caractéristiques faciales (voir section 3.2).

Tableau 3.7 – Résultats de l'évaluation du système de détection de la fatigue et de la somnolence sous un éclairage artificiel pendant la nuit

Seq	**VP**		**VN**		**FP**		**FN**		**T**		**CCR**		κ	
1	115	35	23	71	0	4	19	5	157	115	0.87	0.92	0.63	0.82
2	43	19	55	27	7	3	11	1	116	50	0.84	0.92	0.69	0.83
									Moy.		0.85	0.92	0.66	0.83
									Moy. totale		0.88		0.75	

3.6.5 Résultats sur quelques systèmes existants

Nous présentons dans le tableau 3.8 quelques résultats obtenus par d'autres systèmes de détection de la fatigue et/ou de la somnolence. Nous précisons que pour chaque méthode listée dans ce tableau, les données utilisées pour les tests sont personnelles et ne sont pas publiques. De plus, la reproduction des systèmes n'est pas faisable puisque les pré-traitements, les différents paramètres et la méthodologie complète ne sont pas tous détaillés dans les articles. De ce fait, la comparaison n'est pas équitable et est donnée uniquement à titre indicatif. Il est donc impossible d'estimer lequel des ces systèmes est le meilleur.

Tableau 3.8 – Comparaison avec divers systèmes de détection de l'hypovigilance

Approche	**Description**	**CCR**
D'Orazio *et al.* (2004)	Détection de l'œil par CHT dans l'image entière	93%
Horng *et al.* (2004)	Suivi des yeux + somnolence par l'analyse de la couleur du globe oculaire	88%
Senaratne *et al.* (2007)	PERCLOS + fréquence d'inclinaison de la tête + fréquence de courbure du dos + fréquence d'ajustement de la posture	88%
Zhang *et al.* (2008)	Construction d'un template pour chaque œil, utilisé pour une mise en correspondance avec les frames de test et calcul du degré d'ouverture.	95%
Friedrichs et Yang (2010)	PERCLOS + changement des distances entre les paupières + fermeture des yeux	88%
Rongben *et al.* (2004)	Suivi de la bouche par filtre de Kalman + extraction de ses caractéristiques géométriques et classification par réseaux de neurones	96%.
Fan *et al.* (2007)	Texture par Ondelettes de Gabor et classification par LDA	95%
Saradadevi et Bajaj (2008)	Détection de la bouche par la méthode de Viola-Jones + Classification du bâillement par SVM	83%

3.7 Conclusion

Dans le chapitre 3, nous avons proposé un système composé de deux critères pour la détection de la somnolence et de la fatigue chez le conducteur, basés sur l'analyse des yeux et de la bouche respectivement. Nous avons aussi présenté les résultats des tests obtenus sur une base de données que nous avons acquise pour valider le système. Nous rappelons qu'il n'existe à ce jour aucune base publique dédiée à l'analyse de la somnolence et de la fatigue, mais nous avons affiché les résultats obtenus par quelques systèmes sur leurs propres données, uniquement à titre indicatif. Nous avons prouvé la robustesse de notre système pour différentes séquences représentant différents sujets sous des conditions d'éclairage réelles. Nous avons obtenu en moyenne un CCR de 97% et un coefficient κ de 0.94 sous la lumière ambiante du jour, ce qui représente des valeurs très satisfaisantes. Quand les séquences sont acquises sous un éclairage artificiel pendant la nuit, ces valeurs sont réduites à 88% et à 0.75 respectivement, mais restent tout de même acceptables.

Un des avantages majeurs de ce que nous proposons est l'association de la rapidité de calcul à la robustesse des résultats. Toutefois, il existe quelques limitations à notre système :

– La première limitation est liée à l'extraction des caractéristiques faciales. En effet, il est impossible de détecter la somnolence et la fatigue dans le cas de non extraction de ces caractéristiques. Pour remédier à ce problème, nous introduisons une étape permettant de déterminer l'état du conducteur même quand les caractéristiques faciales ne sont pas visibles. Puisque la non-visibilité de ces caractéristiques est généralement causée par une position non frontale de la tête, il est évident que l'intégration d'une approche robuste pour l'estimation de la pose de la tête nous permettra de déterminer un état d'inattention

sans avoir à extraire les caractéristiques faciales. Nous présentons dans le chapitre 4 une étude des différentes techniques de l'estimation de la pose de la tête en général, ainsi que celles dédiées à l'analyse de l'état du conducteur. Dans le chapitre 5 et le chapitre 6, nous expliquons en détail deux approches que nous proposons pour résoudre le problème de l'estimation de la pose de la tête du conducteur pour détecter l'inattention.

– La seconde limitation est, quant à elle, engendrée par l'éclairage de la scène. Comme nous l'avons vu dans ce chapitre, quand les conditions d'éclairage sont adéquates, par exemple lors d'une conduite pendant le jour, le système atteint des performances maximales. Tandis que pendant la nuit, nous observons une nette réduction de celles-ci. Ainsi, l'une de nos perspectives à court terme est d'intégrer un système d'éclairage infrarouge qui ne sera activé que pendant la nuit afin d'améliorer l'acquisition de la scène. Il est à noter que ce changement entrainera la nécessité d'adapter notre système à ce nouveau type d'éclairage.

Partie II

Estimation de la pose de la tête du conducteur pour la détection de l'inattention

---

Chapitre

Estimation de la pose de la tête et détection de l'inattention : état de l'art

Sommaire

4.1 Estimation de la pose de la tête **57**
4.1.1 Introduction 57
4.1.2 Estimateurs de la pose de la tête basés sur les modèles 59
4.1.2.1 Modèles flexibles 60
4.1.2.2 Techniques géométriques 63
4.1.3 Estimateurs de la pose de la tête basés sur l'apparence 65
4.1.3.1 Template d'apparence 65
4.1.3.2 Classification 68
4.1.3.3 Régression 69
4.1.3.4 Suivi 70
4.2 Estimation de la pose de la tête du conducteur **71**
4.3 Conclusion **74**

Contrairement aux techniques existantes pour l'analyse de la somnolence et de la fatigue qui se basent sur l'étude de différentes caractéristiques faciales (Chapitre 2), l'analyse de l'inattention est fortement liée à l'estimation de la pose de la tête. Nous détaillons, en premier lieu, l'état de l'art sur les estimateurs de la pose de la tête en général (section 4.1), puis nous présentons des estimateurs dédiés à l'analyse de l'inattention chez le conducteur (section 4.2).

4.1 Estimation de la pose de la tête

4.1.1 Introduction

L'estimation de la pose de la tête se présente comme un processus permettant de déduire l'orientation de la tête relativement à la vue d'une caméra. Ce processus requiert une série de traitements pour transformer une représentation de la tête sous forme de pixels en un concept de directions de haut niveau. Nous distinguons deux types d'estimateurs de la pose de la tête. Le premier type permet de reconnaitre une ou peu d'orientations discrètes comme par exemple une vue frontale par rapport à une vue du profil gauche ou droit. Le deuxième type opère sur un niveau plus précis pour fournir une mesure angulaire continue à travers plusieurs degrés de liberté. Le rang normal du mouvement de la tête (Murphy-Chutorian et Trivedi, 2009) englobe une extension sagittal qui correspond au mouvement vers l'avant et l'arrière de la nuque entre

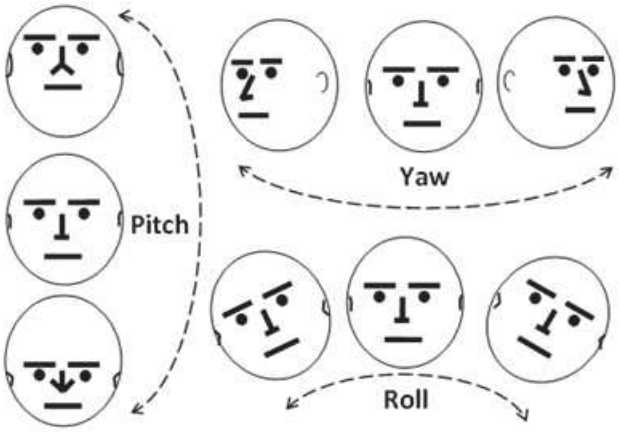


Figure 4.1 – Représentation de la pose de la tête par trois degrés de liberté

$-60.4°$ et $69.6°$, une flexion frontale latérale qui reflète le mouvement de gauche à droite de la nuque entre $-40.9°$ et $36.3°$, ainsi qu'une rotation axiale horizontale de gauche à droite de la tête entre $-79.8°$ et $75.3°$. Murphy-Chutorian et Trivedi (2009) ont aussi remarqué qu'une image acquise par une caméra déposée face à une tête tournée vers un côté n'est pas exactement semblable à une image d'une vue de la tête par ce même profil. Cependant, les combinaisons entre les rotations musculaires et les rotations relatives sont souvent ignorées, ce qui permet de considérer ces deux images comme identiques. Ainsi, la tête est souvent modélisée comme un objet rigide immatériel. En prenant en compte cette hypothèse, la pose de la tête peut être limitée, comme illustré par la figure 4.1, à trois degrés de liberté :

- le pitch : mouvement de haut vers le bas,
- le yaw : mouvement de gauche à droite,
- le roll : rotation.

Une autre hypothèse à tenir en compte lors de l'élaboration d'un estimateur de la pose de la tête est l'hypothèse de la similarité de la pose (Murphy-Chutorian et Trivedi, 2009). Cette hypothèse stipule que différents sujets sous la même pose sont plus similaires que le même sujet sous différentes poses. En d'autres termes, la pose est un indicateur plus puissant de la similarité que l'identité. Cependant, cette hypothèse n'est valable que pour des changements significatifs de la pose, et même dans ce cas, elle peut s'annuler à cause de la sensibilité de l'image aux variations de l'éclairage et de l'alignement de l'image par rapport à la caméra. Pour vérifier l'hypothèse de la similarité de la pose, l'image de la tête devra être transformée pour minimiser l'effet de ces variations et mettre en évidence les différences selon la pose en réduisant celles concernant l'identité.

Comme pour la plupart des systèmes dédiés à l'analyse du visage, un estimateur idéal de la pose de la tête devra être en mesure de démontrer une invariance face à plusieurs facteurs influençant l'image. Dans la littérature (Murphy-Chutorian et Trivedi, 2009; Liu *et al.*, 2009; Murphy-Chutorian et Trivedi, 2010), trois conditions sont établies pour définir un estimateur de la pose de la tête qui soit robuste et efficace :

(C1) Effectuer une estimation de la pose de la tête à partir d'une seule caméra. La précision peut être éventuellement améliorée en utilisant des techniques de stéréo vision. Dans ce cas, il faut tenir en compte des exigences additionnelles en termes de temps de calcul, de mémoire et de coût des équipements.

(C2) Assurer une autonomie de l'estimateur de la pose de la tête en évitant des initialisations ou des ajustements manuels.

(C3) Garantir une invariance de l'estimateur de la pose de la tête face aux changements de l'environnement et de l'identité.

Divers types de caméras ont été utilisés pour acquérir les séquences vidéos à partir desquelles la pose de la tête est estimée. Nous pouvons les diviser en trois catégories : caméras à spectre visible (Ricci et Odobez, 2009; Gourier *et al.*, 2007), caméra infrarouge (Bretzner et Krantz, 2005; Murphy-Chutorian *et al.*, 2007), caméra stéréo (Gurbuz *et al.*, 2012; Munoz-Salinas *et al.*, 2012) et la Kinect (Li *et al.*, 2012). Les caméras infrarouges sont particulièrement efficaces pour effectuer des acquisitions dans des environnements obscures, mais elles ne sont pas adaptées aux environnements sujets à une forte luminosité, puisque les images engendrées dans ce cas sont très brillantes. Les caméras stéréos correspondent à un système de plusieurs caméras qui peuvent être à spectre visible ou infrarouge mais leur coût est assez important. La Kinect est un capteur spécial de Microsoft dédié aux jeux vidéo, mais il est également utilisé dans des applications qui s'intéressent aux différents mouvements des personnes. Ce dispositif fournit, entre autre, des images couleurs et infrarouges, ainsi que des informations 3D. Toutefois, la Kinect n'est pas assez adaptée aux conditions réelles de conduite, puisqu'elle est conçue pour fonctionner à l'intérieur d'une habitation avec une distance minimal de 1.8 m de l'objectif. Parmi les différents outils d'acquisition, les caméras à spectre visible sont les moins chers puisqu'elles coûtent une dizaine d'euros. De plus, elles permettent de bonnes acquisitions à condition de disposer d'un minimum d'éclairage.

Divers estimateurs de la pose de la tête ont été proposés dans la littérature (Murphy-Chutorian et Trivedi, 2009). Nous distinguons entre deux types d'estimateurs : ceux basés sur les modèles et ceux basés sur l'apparence, que nous présentons respectivement dans la sous-section 4.1.2 et la sous-section 4.1.3.

4.1.2 Estimateurs de la pose de la tête basés sur les modèles

L'estimation de la pose basée sur les modèles de la tête requière l'usage de caractéristiques faciales spécifiques telles que les yeux, le nez et la bouche. Leur principe consiste à effectuer des comparaisons au niveau de ces caractéristiques plutôt qu'au niveau global de l'apparence. Ainsi, nous pouvons différencier entre deux types d'approches basées sur ce principe, les modèles flexibles et les techniques géométriques.

4.1.2.1 Modèles flexibles

Les modèles flexibles se basent sur l'ajustement d'un modèle non rigide à l'image de telle sorte à ce qu'il corresponde le mieux à la structure faciale de chaque individu. En plus d'un étiquetage des poses, des données d'apprentissage avec des caractéristiques faciales annotées sont requises. Le principe de cette technique est la création d'un modèle 3D qui peut représenter la tête. L'estimation de la pose est ensuite effectuée par la correspondance des points 2D de la tête (représentée dans le plan de l'image) avec des points 3D appartenant au modèle. Les modèles 3D les plus connus pour l'estimation de la pose de la tête sont les modèles actifs d'apparence « Active Appearance Model » (AAM) (Martins et Batista, 2008; Dornaika et Ahlberg, 2004), le modèle cylindrique de la tête (Aggarwal *et al.*, 2005; Cascia *et al.*, 2004) et le modèle ellipsoïdal de la tête (Choi et Kim, 2009).

Martins et Batista (2008) proposent une approche pour estimer les trois degrés de liberté (pitch, yaw et roll) de la pose de la tête en utilisant les AAM (Cootes *et al.*, 2001) et l'algorithme Pose from Orthography and Scaling with ITerations (POSIT) (DeMenthon et Davis, 1995). Le POSIT estime la pose en utilisant une correspondance entre un ensemble de points d'un modèle 3D et les projections 2D de l'image. Les auteurs utilisent un modèle anthropométrique statistique comme modèle 3D, acquis par un balayage frontal d'un modèle physique en utilisant un laser 3D. L'AAM est appliqué sur un visage détecté par l'algorithme Adaboost (Viola et Jones, 2001) pour extraire l'ensemble des points caractéristiques (yeux, sourcils, bouche, nez, etc.) et permettre une bonne correspondance entre les points 2D et les éléments 3D du modèle. L'objectif de l'AAM est de capturer les variations de forme d'un visage par l'analyse statistique d'un ensemble d'apprentissage en transférant la texture des exemples d'apprentissage vers un modèle de référence. Les auteurs ont acquis une séquence vidéo de 140 frames représentant un sujet sous différentes poses et à partir de la frame 95, la distance entre le sujet et la caméra a été modifiée. L'approche AAM + POSIT s'exécute en 5 frames par seconde en utilisant des images 1024×768, sur un processeur Intel P4 à 3.4 Ghz. La figure 4.2 affiche pour chaque frame les angles réels, les angles estimés ainsi que l'erreur angulaire moyenne « Mean Absolute Error » (MAE) pour (a) le pitch, (b) le yaw et (c) le roll. Les résultats graphiques montrent une certaine corrélation entre les MAE produites au niveau du pitch et du yaw, due à la différence entre le sujet et le modèle anthropométrique utilisé.

Aggarwal *et al.* (2005) proposent une approche basée sur un modèle cylindrique de la tête et un filtrage particulaire (Doucet *et al.*, 2001) pour le suivi des poses du visage dans une vidéo selon trois paramètres de translation et trois paramètres de rotation (pitch, yaw et roll). Le visage est modélisé par la surface courbée d'un cylindre pouvant effectuer des translations et des rotations arbitrairement. Le choix d'un modèle cylindrique par les auteurs est justifié par le travail de Cascia *et al.* (2004) qui prouve que ce type de modèle est plus robuste aux perturbations liées aux paramètres du modèle, comparé à un modèle 3D complexe du visage. Les auteurs supposent que le modèle cylindrique est très proche de la structure 3D d'un visage, et ainsi les problèmes liés aux changements de pose et aux auto-occultations sont systématiquement évités. Ensuite, la position des points caractéristiques du modèle cylindrique sont déterminés en utilisant des calculs géométriques. À partir du modèle structurel et ces points caractéristiques, Aggarwal *et al.* (2005)

effectuent une estimation de la pose par le filtrage particulaire afin d'estimer l'état dynamique d'un système à partir d'un ensemble d'observations bruitées. L'utilisation de cette approche statistique pour le suivi permet de fournir une robustesse face aux occultations modérées et aux variations de l'éclairage. Pour prouver l'efficacité de leur approche face à l'occultation, le changement des expressions faciales et les poses extrêmes, les auteurs ont effectué des tests sur trois ensembles de données (Honda/UCSD dataset (Lee *et al.*, 2003), BU dataset (Cascia *et al.*, 2004) et Li dataset (Li et Chellappa, 2001)). Ces ensembles de données présentent plusieurs séquences sous différents changements d'éclairage, d'expressions et de poses de la tête. La figure 4.3 affiche quelques résultats du suivi représenté par une grille cylindrique. La première ligne de cette figure affiche la capacité de l'approche à estimer les poses extrêmes, tandis que la seconde illustre l'efficacité en cas d'occultation. Un inconvénient de cette approche est que le suivi requiert un grand nombre de particules pour une bonne performance, ce qui la rend inadaptée aux applications nécessitant une exécution en temps-réel.

Choi et Kim (2009) proposent un suivi 3D de la tête basé sur le suivi par filtrage particulaire et un modèle ellipsoïdal de la tête. Leur choix s'est porté sur ce modèle au lieu du modèle

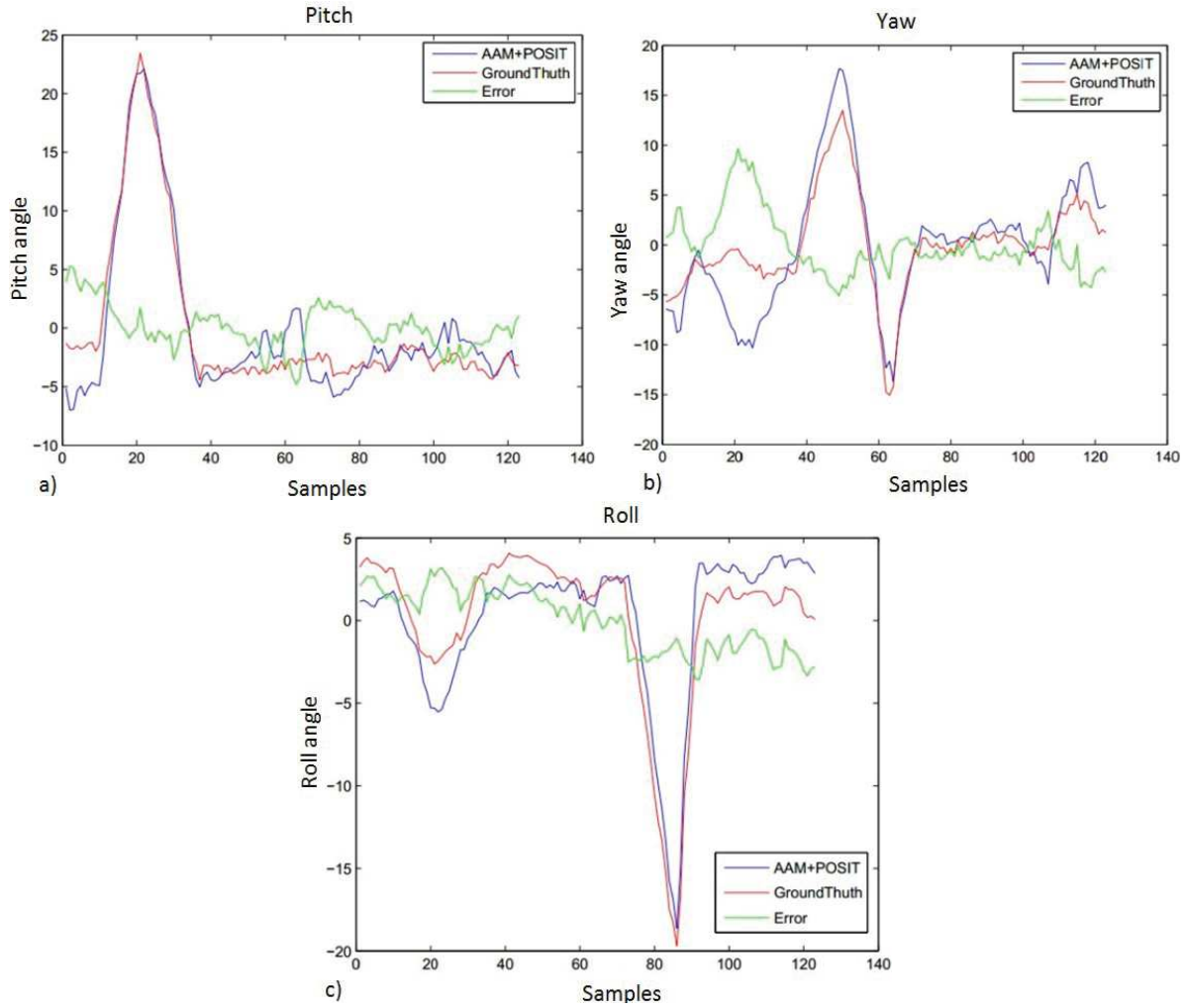


Figure 4.2 – Résultats de l'estimation de la pose de la tête par l'approche AAM + POSIT pour (a) le pitch, (b) le yaw et (c) le roll (Martins et Batista, 2008)

Tableau 4.1 – Comparaison des MAE pour le modèle cylindrique et le modèle ellipsoïdal de la tête (Choi et Kim, 2009)

Modèle	Pitch	Yaw	Roll
Cylindrique	4.43°	5.19°	2.45°
Ellipsoïdal	3.92°	4.04°	2.82°

cylindrique puisque ce dernier n'assure pas une bonne performance du suivi quand la rotation s'effectue selon le pitch. La défaillance du suivi associé au modèle cylindrique peut s'expliquer par deux points. Le premier point est que la texture de la tête change significativement pour les rotations selon le yaw et le roll. Cependant, le changement le plus significatif associé au pitch concerne plutôt l'échelle que la texture. Ainsi, il est difficile de différencier entre la translation de la tête selon l'axe des z (roll) et sa rotation autour de l'axe des x (pitch). Le second point est que le modèle cylindrique ne permet pas une bonne approximation du front à cause de sa forme courbée. Toutefois, le modèle ellipsoïdal s'adapte parfaitement à cette forme, ce qui permet un meilleur suivi quand la rotation se fait autour de l'axe des x. Pour permettre une adaptation aux changements de la pose à court et à long termes, un modèle d'apparence en ligne (Jepson *et al.*, 2003) est introduit pour construire un modèle d'observation stable et adaptative pour le filtrage particulaire. Les auteurs ont effectué trois différents tests pour valider leur approche sur un processeur Intel Core 2 duo à 2.4GHz. Avec 50 particules, le suivi s'exécute en 14 frames par seconde. Le premier test consiste à comparer le suivi par un modèle cylindrique et un modèle ellipsoïdal sur l'ensemble de données de l'université de Boston (Valenti et Gevers, 2009), comme illustré par la figure 4.4.

Le tableau 4.1 affiche les MAE pour le pitch, le yaw et le roll pour cet ensemble de données. Le modèle ellipsoïdal affiche une MAE pour le pitch réduite en moyenne de 15% de celle du modèle cylindrique.

Les limitations des estimateurs basés sur les modèles flexibles de la tête concernent princi-

Figure 4.3 – Résultats du suivi par le filtrage particulaire sur des points caractéristiques d'un modèle cylindrique de la tête (Aggarwal *et al.*, 2005), affichés par la grille cylindrique. Les 3-uplets correspondent aux orientations estimées (roll,pitch,yaw)

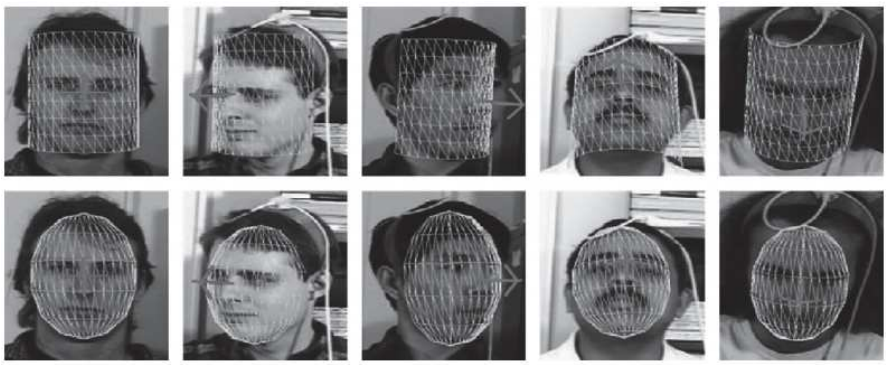


Figure 4.4 – Application d'un modèle cylindrique (ligne 1) et un modèle ellipsoïdal (ligne 1) sur la tête (Choi et Kim, 2009)

palement les erreurs liées à l'extraction des caractéristiques faciales nécessaires à la construction du modèle. De plus, ces estimateurs ne sont pas adaptés à des applications temps-réel opérant sur des images à faibles résolutions.

4.1.2.2 Techniques géométriques

Les techniques géométriques sont particulièrement intéressantes puisqu'elles exploitent directement les propriétés connues de la tête telles que sa forme et la configuration précise des caractéristiques faciales pour estimer la pose. Les premiers estimateurs exploitant les techniques géométriques ont utilisé uniquement les positions d'un ensemble de caractéristiques faciales, supposées connues au préalable, pour estimer la pose de la tête. La configuration des caractéristiques faciales peut être manipulée de différentes manières pour estimer la pose. Par exemple, Gee et Cipolla (1994) utilisent cinq caractéristiques (les coins extérieurs de chaque œil, les coins extérieurs de la bouche et le bout du nez) pour déterminer l'axe de la symétrie faciale en traçant une ligne entre le centre de l'axe des yeux et le centre de l'axe de la bouche. Ensuite, ils déterminent la direction de la tête en utilisant des techniques de la géométrie perspective. La MAE obtenue pour l'estimation du yaw par cet technique est de 15°.

Horprasert *et al.* (1996) proposent une autre estimation de la pose en utilisant un autre ensemble de cinq points (les coins internes et externes de chaque œil et le bout du nez). Sous l'hypothèse que les quatre points des yeux sont coplanaires, le yaw peut être déterminé à partir de la différence observable entre la taille de l'œil gauche et droit grâce à la distorsion projective des paramètres connus de la caméra. Le roll est retrouvé à partir de l'angle entre cette ligne et l'horizon. Le pitch est déterminé en comparant la distance entre le bout du nez et la ligne des yeux à un modèle anthropométrique. Contrairement au système proposé par Gee et Cipolla (1994), cette technique ne présente pas une solution pour améliorer l'estimation de la pose pour les vues proches-frontales. Ces configurations sont appelées angles dégénératives puisqu'ils requièrent une précision très élevée pour une bonne estimation de la pose avec ce modèle.

Wang et Sung (2008) proposent une autre approche géométrique en utilisant le coin interne et externe de chaque œil et les coins de la bouche qui sont détectés automatiquement à partir de l'image. Les auteurs ont observé que les lignes entre les coins externes de l'œil, les coins internes, et la bouche sont parallèles. Toute déviation observée du niveau parallèle dans l'image plane est

un résultat de la distorsion de perspective. Le point où ces lignes se croisent dans le plan de l'image peut être calculé en utilisant les moindres carrées. Ce point peut être utilisé pour estimer l'orientation 3D des lignes parallèles si le rapport des tailles est connu. Il peut aussi être utilisé pour estimer la position 3D absolue de chaque point caractéristique si la taille des lignes est connue. Pour améliorer l'efficacité de leur technique géométrique, les auteurs ont introduit une phase d'adaptation de l'estimation de la pose à chaque individu. Cette tâche est effectuée par l'algorithme Expectation-Maximisation (EM) qui consiste à estimer la probabilité a posteriori d'une pose lorsque les caractéristiques faciales sont fournies. Les tests sont effectués sur des données réelles dont la vérité terrain est déterminée par le système de suivi de l'orientation InertiaCube2 [1]. Les résultats sont exprimés en termes de la racine de la moyenne du carré « Root Mean Square » (RMS) et correspondent à 1.7°, 2.6° et 3.6° pour le pitch, yaw et roll respectivement. L'inconvénient du système proposé par Wang et Sung (2008) est que la pose ne peut être estimée que si elle est assez proche d'une vue frontale pour voir toutes les lignes du visage.

Une autre méthode géométrique plus récente, proposée par Dahmane *et al.* (2012), sélectionne un ensemble de caractéristiques à partir des parties symétriques du visage. Ils effectuent un apprentissage par un modèle d'arbre de décision afin de reconnaitre la pose de la tête en se basant sur les zones de symétrie faciales. Contrairement à la plupart des approches géométriques, cette méthode ne nécessite pas l'extraction des caractéristiques faciales (yeux, bouche,...) mais se base uniquement sur l'extraction des caractéristiques de la symétrie faciale. Les auteurs considèrent que la taille de la zone de symétrie bilatéral est un bon indicateur de la rotation selon le yaw. En effet, quand la tête est face à la caméra, la symétrie entre ses deux parties gauche et droite est bien apparente et la ligne qui relie les deux yeux au bout du nez définie l'axe de symétrie. Cependant, plus l'angle de mouvement de la tête est grand, moins on dispose de pixels symétriques. Les auteurs utilisent la base de données FacePix (Black *et al.*, 2002) pour l'apprentissage et l'évaluation en considérant uniquement sept poses pour le yaw variant entre $-45°$ et $+45°$ avec un pas de 15°. Ce rang de poses ne peut être élargi puisque la symétrie bilatérale risque de disparaitre du plan de l'image. Pour la classification basée sur les arbres de décision, quatre classes discrètes sont utilisées : classe 1 = $\pm 45°$, classe 2 = $\pm 30°$, classe 3 = $\pm 15°$ et classe 4 = $0°$. Les poses gauches et droites sont groupées dans une même classe puisqu'elles présentent la même symétrie et donc la même information. La figure 4.5 (a) montre les CCR pour chaque classe obtenus par les arbres de décision, la moyenne pondérée de ces taux est de 81.1%. Un test est effectué sur l'ensemble des données de l'université de Boston (Valenti et Gevers, 2009) afin d'estimer la pose de la tête. Pour localiser la région du visage, le détecteur de Viola-Jones est utilisé (Viola et Jones, 2001) et ensuite les caractéristiques de la symétrie faciale sont extraites. La figure 4.5 (b) montre une frame de l'ensemble des données de l'université de Boston ainsi que le graphe représentant la pose estimée et la vérité terrain après l'application des arbres de décision combinés aux caractéristiques de la symétrie faciale. L'avantage de la méthode proposée par Dahmane *et al.* (2012) est qu'elle peut opérer sur des images à faible ou à grande résolution. Toutefois, l'estimation de la pose est restreinte au yaw et ne peut concerner qu'un intervalle

1. http ://www.intersense.com/pages/18/55/

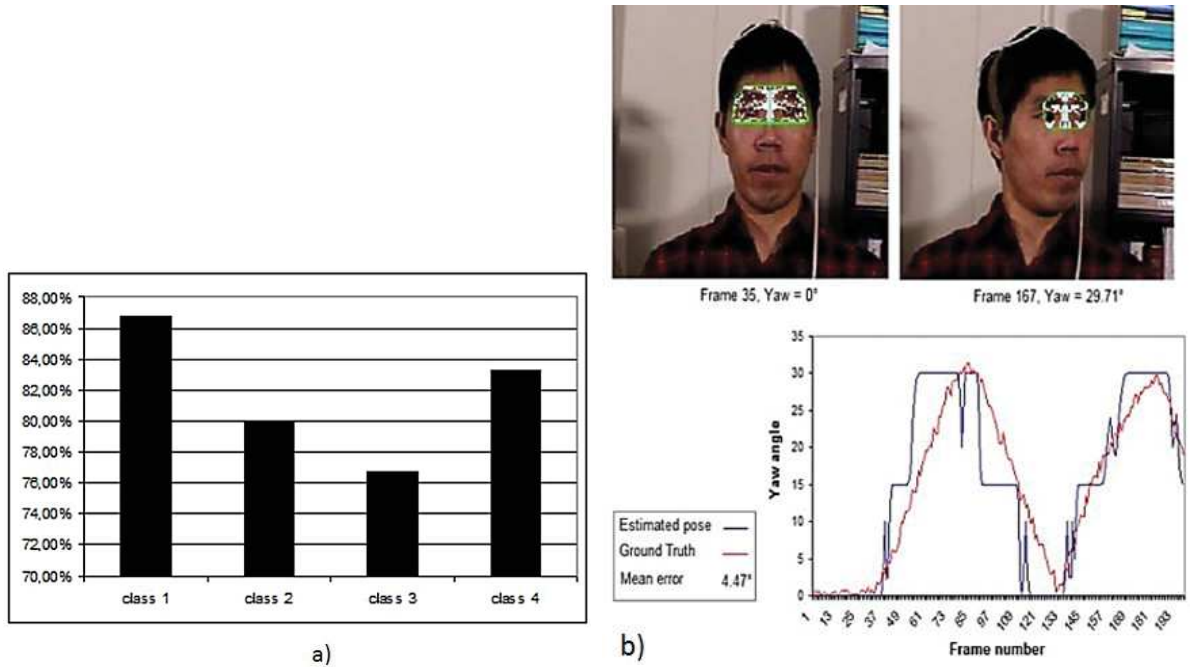


Figure 4.5 – (a) CCR obtenus par l'application des arbres de décision combinés aux caractéristiques de la symétrie faciale sur la base FacePix ; (b) Frame de l'ensemble des données de l'université de Boston (en haut) ; graphe représentant l'estimation de la pose selon le yaw et la vérité terrain (en bas) (Dahmane *et al.*, 2012)

limité de mouvements (entre $-45°$ et $+45°$).

En général, les estimateurs de la pose de la tête basés sur les techniques géométriques sont rapides et simples. En effet, avec seulement quelques caractéristiques faciales, une estimation acceptable de la pose de la tête peut être obtenue. La difficulté réside dans la détection de ces caractéristiques avec une grande précision et exactitude. Il est évident qu'une grande distance entre la tête et la caméra est problématique, car la résolution de l'image peut rendre difficile, voire impossible la détermination précise de l'emplacement des caractéristiques. De plus, les techniques géométriques sont généralement plus sensibles à l'occultation que les méthodes basées sur l'apparence qui utilisent les informations de toute la région faciale.

4.1.3 Estimateurs de la pose de la tête basés sur l'apparence

Les estimateurs basés sur l'apparence opèrent sous l'hypothèse stipulant que la pose 3D de la tête et quelques propriétés de l'image 2D de la tête sont liées par une certaine relation (Ma *et al.*, 2013). Cette relation peut être définie par quelques techniques de vision par ordinateur bien connues telles que la correspondance avec des templates d'apparence, la classification, la régression ou même le suivi.

4.1.3.1 Template d'apparence

Les estimateurs basés sur les templates d'apparence utilisent des métriques de comparaison basées sur l'image pour faire correspondre une vue de la tête à un ensemble de templates labellisés par une pose discrète. Ces méthodes ont certains avantages comparées à des méthodes plus

complexes. Ils sont utilisables pour des images à haute ou à faible résolution. De plus, les templates d'apparence ne nécessitent pas l'apprentissage des exemples négatifs ou l'extraction des caractéristiques faciales. Le nombre de templates peut être augmenté à n'importe quel moment afin d'accroitre les poses discrètes à estimer. En effet, la création d'un ensemble de templates d'apprentissage requiert simplement l'extraction des images représentant les têtes et l'attribution d'un label à chacune d'elles. Cependant, il existe plusieurs inconvénients liés à l'utilisation des méthodes basées sur les templates d'apparence. La région de la tête est supposée être localisée alors que les erreurs de localisation peuvent dégrader la précision de l'estimation. Ces méthodes peuvent aussi présenter un problème d'efficacité du fait que plus le nombre de templates augmente, plus le temps d'exécution est important. Pour résoudre ces deux problèmes, Ng et Gong (2002) ont effectué l'apprentissage d'un ensemble de SVM pour détecter et localiser le visage et ont utilisé par la suite les vecteurs supports comme des templates d'apparence pour estimer la pose de la tête. Toutefois, le problème majeur des templates d'apparence est qu'ils opèrent selon l'hypothèse admettant qu'une similarité dans l'espace de l'image peut correspondre à une similarité de la pose. Pour illustrer ce problème, on considère deux images de la même personne dans des poses faiblement différentes et deux images de deux personnes différentes dans la même pose. Dans ce scénario, l'identité produit plus de différence dans l'image que la variation de la pose, ce qui fait que l'algorithme du template matching fait correspondre une pose inappropriée à l'image. Toutefois, cet effet peut être considérablement réduit en choisissant soigneusement la technique de mise en correspondance, ainsi qu'en appliquant de bonnes transformations sur l'image. Par exemple, un filtre LoG peut être appliqué à l'image pour accentuer les contours les plus communs du visage tout en supprimant les variations de texture spécifiques à chaque individu. Ainsi, plusieurs travaux ont adopté des améliorations sur les méthodes basées sur les templates d'apparence afin de les rendre plus performantes (Sherrah *et al.*, 2001; Ricci et Odobez, 2009; Morency *et al.*, 2010).

Dans (Sherrah *et al.*, 2001), les templates d'apparence sont utilisés au lieu d'un modèle 3D afin d'effectuer simultanément une détection de la tête et une estimation de sa pose en se basant sur le suivi. Des mesures de similarité de deuxième ordre sont utilisées comme technique de correspondance. En effet, les auteurs ont effectué plusieurs tests pour déterminer les contraintes qui satisfont l'hypothèse de la similarité de la pose (voir sous-section 4.1.1). Ils ont étudié, pour une certaine pose, la transformation de l'image la plus adaptée pour mettre en évidence les différences de la pose et ignorer celles de l'identité. Ensuite, ils ont analysé la séparation angulaire minimale permettant de vérifier l'hypothèse. Ils ont déduit intuitivement que les filtres basés sur les orientations tels que les filtres de Gabor sont les plus adaptés pour déterminer les caractéristiques spécifiques à la pose. Cependant, ces filtres ne produisent qu'une faible invariance face à l'identité, qui peut être améliorée par l'application de la méthode de l'analyse en composantes principales « Principal Component Analysis » (PCA). Pour évaluer l'effet de ces transformations sur l'estimation de la pose, un ratio basé sur l'hypothèse de la similarité de la pose est défini, plus ce ratio est petit plus la performance est grande. La figure 4.6 montre le ratio de la similarité de la pose pour (a) le pitch (aussi appelé tilt) et (b) le yaw après la variation des orientations des filtres de Gabor en fixant en premier le yaw à 90° et ensuite le pitch à 90°. Le ratio varie selon les orientations des filtres mais aussi selon la pose, ce qui implique que les filtres

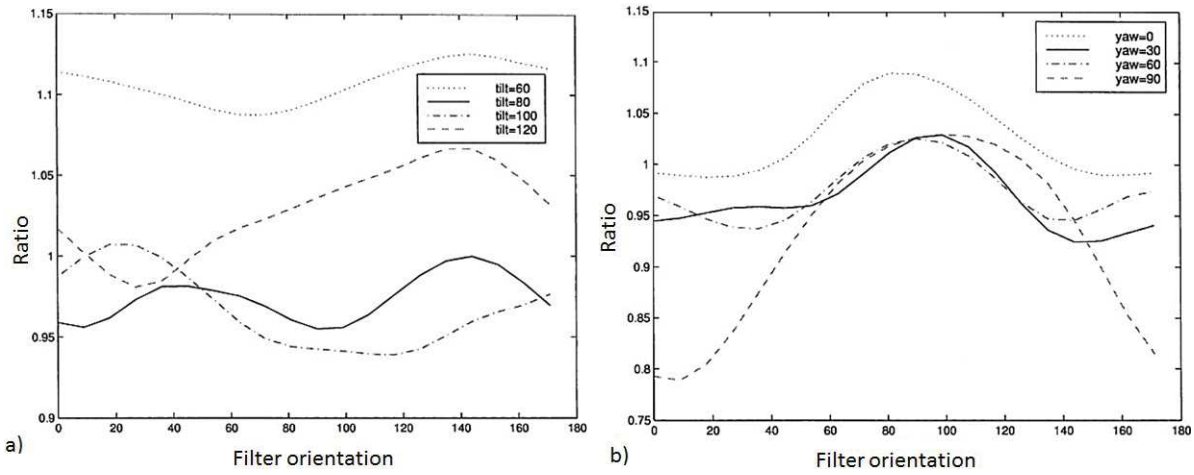


Figure 4.6 – Le ratio de la similarité de la pose en variant la pose de la tête et l'orientation des filtres de Gabor. (a) variation du pitch avec yaw fixé à 90° ; (b) variation du yaw avec pitch fixé à 90° (Sherrah *et al.*, 2001)

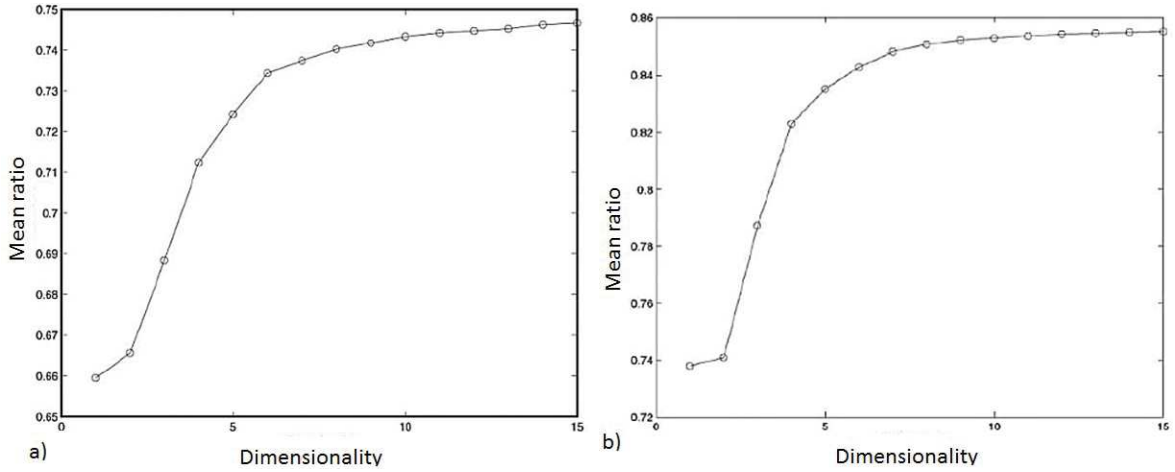


Figure 4.7 – Le ratio de la similarité de la pose en variant la pose de la tête et les dimensions du PCA. (a) : variation du pitch ; (b) : variation du yaw (Sherrah *et al.*, 2001)

de Gabor révèlent des caractéristiques orientées dépendantes de la pose. La figure 4.7 montre le ratio de la similarité de la pose pour (a) le pitch et (b) le yaw après la variation du nombre de dimensions du PCA. En moyenne, les ratios obtenus pour le second test sont inférieurs à ceux du premier. Ceci implique que le PCA ne fait pas qu'annuler l'effet de l'identité mais il renforce aussi la différence de pose. A partir d'un test sur la séparation angulaire minimale, les auteurs ont déduit que chaque pose doit être différente d'une autre par une valeur variant entre 10° et 20° selon la pose.

Dans un travail plus récent proposé par Ricci et Odobez (2009), une approche combinant l'estimation de la pose et le suivi de la tête est présentée afin de permettre une estimation simultanée de la position, la taille et l'orientation de la tête. Les auteurs ont utilisé l'algorithme des histogrammes des gradients orientés « Histograms of Oriented Gradients » (HOG) (Dalal et Triggs, 2005) pour extraire les caractéristiques de texture de l'image. Le HOG a l'avantage d'être rapide et en même temps, il fournit une robustesse aux variations de l'éclairage et à la

différence de la pose. En plus de la texture, des informations de couleur sont extraites pour différencier entre les pixels de peau et les pixels non-peau. Ces informations ne sont pas très représentatives de la pose, mais elle sont utiles pour la localisation de la tête. Ensuite, un algorithme de filtrage particulaire à états mixtes est appliqué pour calculer simultanément la position ainsi que la taille et l'orientation de la tête. La contribution majeure de ce travail est la définition de la fonction de vraisemblance, utilisée par le filtrage particulaire, comme une fonction paramétrique dont les paramètres sont déterminés à partir d'un ensemble d'apprentissage en utilisant une approche discriminatoire. La fonction de vraisemblance paramétrique « Likelihood Parametrized Function » (LPF) mesure la compatibilité entre une nouvelle entrée et le template correspondant à une certaine pose. Pour valider leur approche, les auteurs effectuent en premier lieu des tests sur la base de données Pointing'04 (Gourier *et al.*, 2004) représentant 30 séries d'images, chacune contenant 93 poses différentes. Les images sont divisées en deux ensembles, le premier pour l'apprentissage des templates et des paramètres de la LPF, et le second pour le test. Ils obtiennent pour ce test une MAE de 10.5° pour le pitch et de 9.1° pour le yaw. Dans le cas où les paramètres de la LPF ne sont pas appris, ces valeurs sont égales à 14.2° et 13.7°.

4.1.3.2 Classification

Les techniques de classification peuvent être utilisées pour estimer la pose de la tête quand nous disposons d'un grand nombre de données d'apprentissage. Dans ce cas, la relation entre la pose 3D et l'image 2D est établie en construisant une séparation efficace entre les différentes poses à estimer. Les estimateurs basés sur la classification sont très similaires à ceux basés sur les templates d'apparence puisqu'ils opèrent tous les deux directement sur un patch de l'image. Toutefois, au lieu de comparer une image à un large ensemble de templates, celle-ci est évaluée par un classifieur supervisé ayant effectué un apprentissage sur un grand nombre d'images. L'avantage majeur des estimateurs basés sur la classification est qu'ils emploient des algorithmes d'apprentissage qui sont capables d'ignorer la variation de l'apparence qui ne correspond pas au changement de la pose. Cependant, si le classifieur est utilisé pour détecter la tête et estimer sa pose, il est nécessaire d'inclure suffisamment d'exemples de non-visage au moment de l'apprentissage. Un autre inconvénient est que le nombre de poses à estimer ne peut être augmenté facilement comme pour les templates d'apparence, puisque cette action nécessite d'effectuer un ré-apprentissage du classifieur.

Il est possible d'estimer la pose de la tête par la classification en procédant de deux manières différentes. La première solution consiste à apprendre plusieurs classifieurs, chacun étant dédié à une classe de poses. Cette architecture présente trois inconvénients majeurs. Le premier inconvénient concerne la prise de décision finale puisque différents classifieurs peuvent répondre positivement. Lorsqu'une mesure de confiance peut être associée à la sortie de chaque classifieur, il est possible d'appliquer des règles simples de décision telles qu'une moyenne pondérée ou la règle du « winner takes all ». Cette dernière règle attribue la pose finale au classifieur ayant répondu avec le plus de certitude (Wu *et al.*, 2004; Gourier *et al.*, 2007). Le deuxième inconvénient réside dans le fait qu'à chaque classification, il est nécessaire de faire appel à tous les classifieurs, ce qui est coûteux en termes de temps de calcul. Le partitionnement de l'ensemble d'apprentissage constitue le troisième inconvénient de cette architecture. En effet, il n'est pas

évident de déterminer les frontières des différentes classes puisqu'en réalité, les changements de pose ne sont pas discrets mais continus. Gourier *et al.* (2007) proposent un estimateur basé sur cette architecture et définissent autant de réseaux de neurones que de classes de poses. Chaque réseau est une mémoire auto-associative linéaire « Linear Auto-associative Memory » (LAAM) qui apprend à synthétiser en sortie une image semblable à l'entrée. Si l'entrée possède une orientation proche de celle des visages d'apprentissage, l'entrée et la sortie seront similaires. La pose est alors définie par le réseau possédant l'erreur de reconstruction la plus faible. Les tests effectués sur la base de données Pointing'04 ont fournis des MAE de 10.3° et de 15.9° pour le pitch et le yaw respectivement.

La deuxième solution pour estimer la pose de la tête en utilisant la classification consiste à utiliser des classifieurs multi-classes. Pour cette architecture, tous les exemples sont utilisés pour apprendre simultanément toutes les classes de poses. Toutefois, le problème de la prise de décision reste ouvert lorsque plusieurs sorties du classifieur répondent positivement. Dans (Munoz-Salinas *et al.*, 2012), une estimation de la pose de la tête multi-vues basée sur des SVM multi-classes est proposée. En premier lieu, l'image est filtrée par le filtre de Sobel pour relever la magnitude du gradient, et la dimension est réduite en appliquant le PCA. Ensuite, un SVM multi-classes est appris sur un espace discret de la pose. La distribution de la probabilité des votes des classes est préférée à la sélection de la classe la plus votée par le SVM pour estimer la pose de la tête. Par la suite, cet estimateur est étendu afin d'opérer avec multiples caméras par la fusion de leurs informations. Les auteurs ont effectué des tests sur la base de données PEIS-Home (Saffiotti et Broxvall, 2005) utilisant six caméras. Plus le nombre de caméras est important plus les résultats sont meilleurs. Cependant, le temps de calcul augmente considérablement en fonction du nombre de caméras.

Dans (Jain et Crowley, 2013), l'image est transformée en lui appliquant une pyramide orientable « Steerable Pyramid » (SP) (Simoncelli *et al.*, 1992) composée de dérivées de second ordre d'une gaussienne. L'utilisation de cette transformation permet de mettre en évidence les contours orientés de l'image. Pour réduire la taille du vecteur caractéristique obtenu, les auteurs appliquent le PCA. Ensuite, deux SVM multi-classes sont appris pour estimer la pose de la tête selon le pitch et le yaw. Le test de cette méthode sur la base de données Pointing'04 a procuré une MAE de 8° pour le pitch et de 6.9° pour le yaw avec un temps de traitement de 108 ms par image.

4.1.3.3 Régression

L'estimation de la pose de la tête par les techniques de régression a pour objectif d'apprendre une relation fonctionnelle entre l'apparence de la tête et sa pose. Cette relation est construite à partir d'un ensemble d'apprentissage et permet de fournir une estimation continue de la pose. Toutefois, cette procédure est complexe car elle nécessite d'approximer une fonction fortement non-linéaire dans un espace de grande dimension. Dans (Al-Haj *et al.*, 2012) par exemple, le vecteur caractéristique de la pose est construit par une pyramide de HOG à trois niveaux, et une régression par moindre carrée partiel « Partial Least Square » (PLS) est utilisée pour déterminer les coefficients modélisant la relation entre la tête et sa pose. Pour renforcer la régression, le PLS est associé à un noyau fonction de base radiale « Radial Basis Function » (RBF) qui permet de

Tableau 4.2 – Étude comparative entre SVM, SVR et LARR (Guo *et al.*, 2008)

Technique	Pitch-MAE	Yaw-MAE
SVM	4.73	59.91
SVR	9.37	7.84
LAAR	7.69	9.23

produire de bons résultats au détriment de la complexité. Les résultats obtenus par l'application du PLS associé au RBF sur la base de données Pointing'04 correspondent à des MAE de 6.61° et 6.56° pour le pitch et le yaw respectivement. Quand le PLS linéaire est appliqué (sans noyau), les MAE pour le pitch et le yaw sur la même base de données correspondent à 10.52° et 11.29°.

Plusieurs estimateurs combinent entre une régression par machines à vecteurs supports « Support Vector Regression » (SVR) et un SVM. Dans (Guo *et al.*, 2008), les auteurs ont effectué une étude comparative entre SVM, SVR et la combinaison des deux, nommée régression robuste localement ajustée « Locally Adjusted Robust Regressor » (LARR). Pour une image d'entrée, le principe de la LARR consiste à utiliser deux SVR (l'un pour le pitch et l'autre pour le yaw) et de calculer deux distances entre les angles estimés et toutes les poses possibles. Ces distances sont ensuite triées selon un ordre ascendant et les classes les plus proches sont utilisées pour l'ajustement local par SVM. L'étude est effectuée sur les images de la tête extraites manuellement de la base de données Pointing'04. Le tableau 4.2 affiche les résultats obtenus. Nous en déduisons que le SVM est meilleur pour le pitch mais beaucoup moins performant pour le yaw, alors que le SVR procure une assez bonne performance pour les deux angles. Les MAE du LARR sont proches de ceux du SVR.

Dans Ho et Chellappa (2012), une combinaison entre le SVM et le SVR est aussi proposée. Les auteurs construisent le vecteur caractéristique de la pose par l'application du descripteur Scale Invariant Feature Transform (SIFT) sur des points réguliers appartenant à une grille de l'image au lieu de l'appliquer uniquement sur les points caractéristiques. Puisque la dimension de ce vecteur est importante, une réduction est effectuée par une projection aléatoire « Random Projection » (RP) sur un sous-espace de dimension inférieure. Ce procédé de réduction de l'espace est utilisé pour les vecteurs caractéristiques de très grande dimension, et remplace le PCA qui entraine de lourds calculs quand les dimensions sont importantes. Par la suite, une architecture combinant SVM et SVR est proposée pour l'estimation de la pose de la tête. En premier lieu, l'espace des configurations possibles des poses est divisé en un nombre fixe de classes étiquetées et un SVM multi-classes est appris pour ces classes. Les images appartenant à une même classe sont utilisées pour l'apprentissage d'un SVR pour raffiner l'estimation de la pose, ce qui signifie qu'il faudrait apprendre autant de SVR que de classes. Les auteurs ont testé leur approche sur la base de données Pointing'04 obtenant ainsi des MAE de 5.84° pour le pitch et de 6.05° pour le yaw.

4.1.3.4 Suivi

L'estimation de la pose de la tête basée sur le suivi opère en poursuivant le mouvement relatif de la tête entre frames consécutives d'une séquence vidéo. L'avantage majeur des approches

basées sur le suivi réside dans leur capacité à estimer la pose avec précision en détectant les faibles changements d'orientations. Toutefois, la difficulté de ces approches est liée à la nécessité de fournir une localisation et une pose initiale de la tête. Sans cette étape initiale, il est impossible de découvrir le changement relatif des poses entre les frames.

Le suivi par filtrage particulaire est le plus utilisé dans la littérature pour l'estimation de la pose de la tête et est souvent associé à d'autres approches basées sur l'apparence ou bien le modèle de la tête. Dans (Ricci et Odobez, 2009), un suivi par filtres particulaires à états mixtes est associé à des templates d'apparence combinant entre la texture et la couleur afin de permettre la localisation de la tête et l'estimation de sa pose (voir la sous-section 4.1.3.1 pour plus de détails sur ce travail). Dans les travaux de Aggarwal *et al.* (2005) et Choi et Kim (2009), détaillés dans la sous section 4.1.2.1, les auteurs ont aussi utilisé un suivi par filtres particulaires mais ils l'ont combiné respectivement à des modèles cylindrique et ellipsoïdal de la tête.

Nous avons exposé dans cette section un aperçu sur les techniques utilisées en générale pour l'estimation de la pose de la tête, nous allons consacré la section 4.2 à la présentation de quelques techniques dédiées à l'analyse de l'inattention chez conducteur.

4.2 Estimation de la pose de la tête du conducteur

Un grand intérêt est porté par les chercheurs aux systèmes d'aide à la conduite basés sur la pose de la tête du conducteur comme caractéristique révélant le centre d'attention visuelle d'une personne et son état de vigilance (Bretzner et Krantz, 2005; Bergasa *et al.*, 2006; Trivedi *et al.*, 2007; Murphy-Chutorian et Trivedi, 2008). Le laboratoire de recherche le plus connu dans ce domaine est le LISA, CVRR, University of California, USA [2]. Cette équipe a proposé plusieurs approches pour l'étude de la pose de la tête du conducteur afin de déterminer son niveau d'attention (Huang et Trivedi, 2003; Trivedi *et al.*, 2007; Murphy-Chutorian *et al.*, 2007; Murphy-Chutorian et Trivedi, 2008, 2010; Martin *et al.*, 2012; Tawari *et al.*, 2014).

Dans (Murphy-Chutorian *et al.*, 2007), un estimateur de la pose de la tête du conducteur opérant sur une caméra sensible à un éclairage visible et infrarouge est proposé. Une variante de l'algorithme SIFT appelée histogramme des orientations des gradients localisés « Localized Gradient Orientation » (LGO) est utilisée pour construire le vecteur caractéristique de la pose. Ensuite, ces vecteurs sont utilisés pour l'apprentissage d'un SVR pour le pitch et un autre pour le yaw afin de déterminer la pose de la tête. Pour tester cette architecture, les auteurs ont utilisé leur propre testbed LISA-P illustré par la figure 4.8 (en haut). LISA-P est constitué d'un véhicule embarquant à la fois une caméra sensible aux éclairages visible et infrarouge, une source de lumière infrarouge et un système de capture de mouvements doté de cinq capteurs placés sur la tête du conducteur (voir la figure 4.8 (en bas)). Ce système de capture permet de déterminer une vérité terrain précise pour la pose de la tête. La caméra est utilisée pour capturer des vidéos du conducteur de jour comme de nuit. Ainsi, la source de lumière proche infrarouge est placée devant le conducteur pour permettre une acquisition dans des environnements obscures. Les données d'apprentissage sont acquises par le système de capture des mouvements. Deux séries

2. Laboratory for Intelligent and Safe Automobiles, Computer Vision and Robotics Research Laboratory, University of California, San Diego, USA `http://cvrr.ucsd.edu/LISA/index.html`

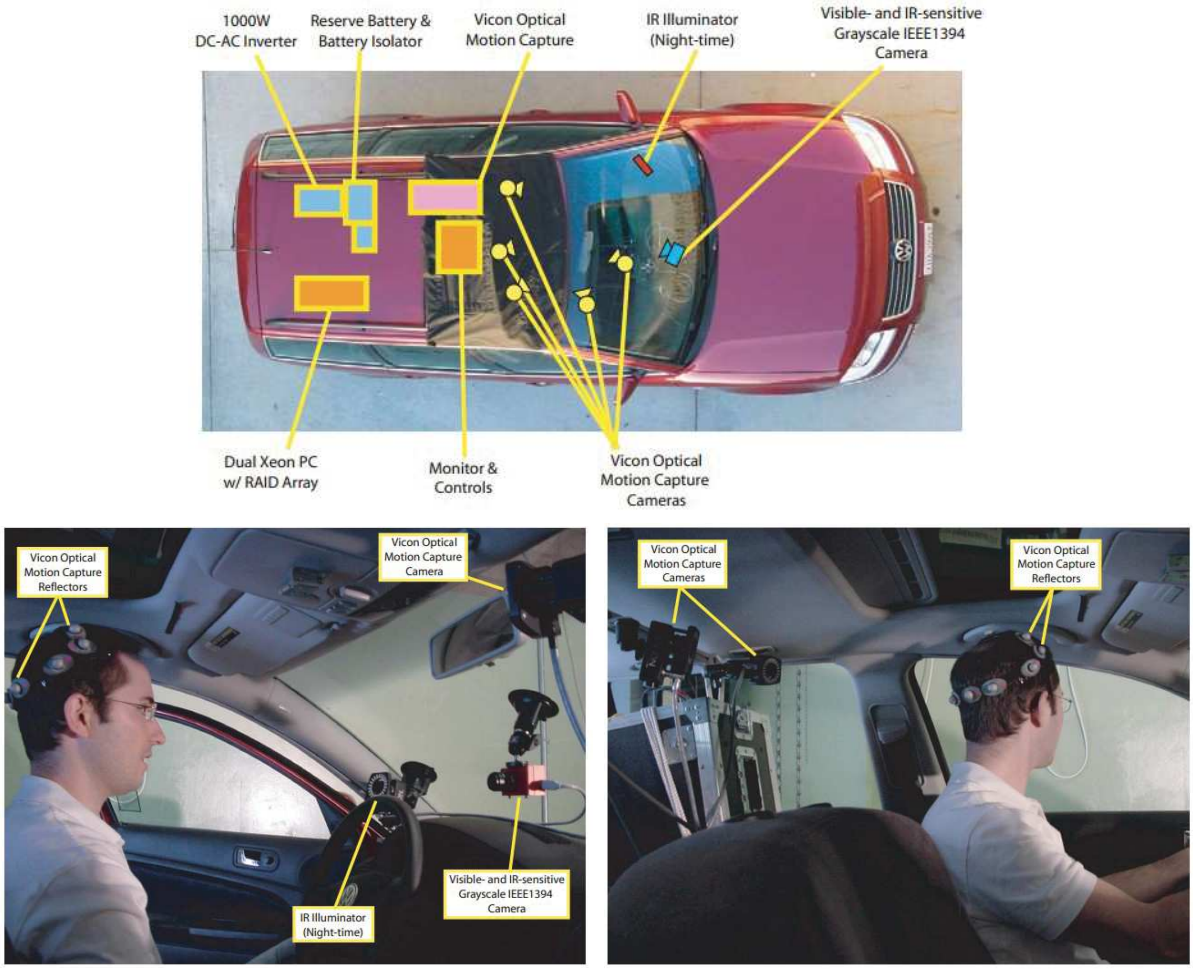


Figure 4.8 – (En haut) Le testbed LISA-P utilisé par Murphy-Chutorian *et al.* (2007) pour la collecte et l'évaluation de leur estimateur de la pose de la tête ; (En bas) Vue du conducteur par le testbed LISA-P.

de tests ont été conduites, l'une dans une salle et l'autre dans le testbed LISA-P. Les résultats obtenus en termes de MAE sont les suivants :

- Test laboratoire (dix sujets), pitch=5.58°, yaw=6.40
- Test véhicule LISA-P (six sujets), pendant la journée : pitch= 3.99°, yaw=9.28° ; pendant le soir : pitch=5.18°, yaw=7.74°

Dans (Murphy-Chutorian et Trivedi, 2008, 2010), l'estimation de la pose de la tête du conducteur est traitée par une combinaison de plusieurs techniques : régression, template d'apparence, modèle de la tête et suivi. L'orientation initiale de la tête est estimée par l'approche présentée ci-dessus (Murphy-Chutorian *et al.*, 2007), en ajoutant un troisième SVR pour l'estimation du roll. Cet estimateur constitue une étape d'initialisation (ou de réinitialisation) d'un algorithme de suivi de la pose par filtrage particulaire combiné à un modèle de texture 3D de la tête. Cet algorithme met à jour un modèle anthropométrique rigide de la tête en utilisant un ensemble de comparaisons basées sur l'apparence pour estimer le mouvement qui minimise la différence entre la projection virtuelle du modèle et la frame suivante. La figure 4.9 affiche un exemple de l'adaptation de la tête au modèle rigide par l'algorithme de suivi. L'implémentation sur un

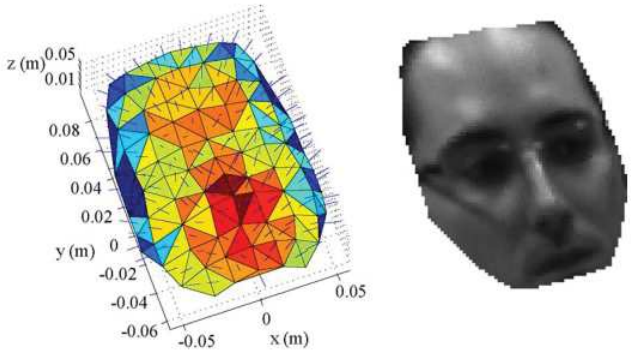


Figure 4.9 – (Gauche) Modèle anthropométrique 3D utilisé pour le suivi. (Droite) Exemple du modèle retourné par l'algorithme du suivi (Murphy-Chutorian et Trivedi, 2010)

processeur graphique « Graphics Processing Unit » (GPU) [3] permet le traitement en temps réel. Le même dispositif présenté dans la figure 4.8 est utilisé pour construire un ensemble de données représentant 14 sujets en situation de conduite. Les erreurs obtenues sont données en termes de MAE et correspondent aux résultats suivants : pitch= 8.57°, yaw= 11.24° et roll= 8.29°.

Il existe aussi des produits commerciaux pour l'estimation de la pose de la tête du conducteur. L'un de ces produits est le Smart Eye AntiSleep (AntiSleep 4, 2013; Bretzner et Krantz, 2005) qui correspond à un système conçu pour déterminer la somnolence et l'inattention chez le conducteur. Ce système peut s'exécuter sur un ordinateur portable en utilisant l'unité d'acquisition constituée d'une caméra à spectre visible en plus de deux sources de lumières infrarouges, comme illustré par la figure 4.10. AntiSleep mesure la position et l'orientation 3D de la tête, la direction du regard et la fermeture des paupières. Pour l'estimation de la pose de la tête, les auteurs utilisent une approche de suivi et une méthode géométrique basée sur un modèle générique de la tête comme étape d'initialisation. Toutefois, ce système est limité à un usage dans un environnement contrôlé pour des fins de simulations et n'est pas utilisé en pratique, à cause de l'influence des lumières infrarouges pendant la conduite en plein jour. En effet, la présence d'une forte luminosité réduit considérablement l'efficacité des caméras infrarouges, puisque les images acquises dans ce cas sont de mauvaise qualité.

Figure 4.10 – L'unité d'acquisition Smart Eye AntiSleep composée d'une seule camera et de deux sources de lumières infrarouges (Bretzner et Krantz, 2005)

3. GPU circuit intégré présent sur une carte graphique disposant d'une structure hautement parallèle

4.3 Conclusion

Tableau 4.3 – Tableau récapitulatif des estimateurs de la pose de la tête. (1) écart type, (2) RMS, (3) MAE, (4) CCR. (*) estimateur dédié au conducteur.
MF (Modèle flexible) ; MG (Méthode géométrique) ; Cl (Classification) ; Rg (Régression) ; Sv (Suivi) ; TA (Template d'apparence)

Approche	**Cat.**	**Résultats**			**Données**
		Pitch	Yaw	Roll	
AAM + POSIT (Martins et Batista, 2008)	MF + Sv	(1)2.6°	(1)1.7°	(1)1.9°	Personnelle
Modèle cylindrique + filtrage Particulaire (Aggarwal *et al.*, 2005)	MF + Sv	(3)4.4°	(3)5.2°	(3)2.5°	UCSD, BU, Li
Modèle ellipsoïdal + Filtrage Particulaire (Choi et Kim, 2009)	MF + Sv	(3)3.9°	(3)4.1°	(3)2.8°	Boston
Géométrie faciale (Horprasert *et al.*, 1996)	MG	-	-	-	Personnelle
Symétrie faciale + EM (Wang et Sung, 2008)	MG	(2)1.7°	(2)2.6°	(2)3.6°	Personnelle
Reconstruction 3D + Suivi (Gurbuz *et al.*, 2012)	MG + Sv	(2)2.5°	(2)3.2°	(2)2.6°	Personnelle
Symétrie faciale + Arbres décisionnels (Dahmane *et al.*, 2012)	MG + Cl	(4) 81%			FacePix
Filtre Gabor + PCA + Suivi (Sherrah *et al.*, 2001)	TA + Sv	-	-	-	Personnelle
HOG + Filtrage particulaire (Ricci et Odobez, 2009)	TA + Sv	(3)10.5°	(3)9.1°	-	Pointing'04
Modèle d'apparence + Suivi (Morency *et al.*, 2010)	TA + Sv	(3)4.7°	(3)6.9°	(3)4.3°	MIT
LAAM (Gourier *et al.*, 2007)	Cl	(3)10.3°	(3)15.9°	-	Pointing'04
SVM + PCA + Caméras stéréos (Munoz-Salinas *et al.*, 2012)	Cl	(2)6.5°	(2)9.6°	-	PEIS-Home
SP + PCA + SVM (Jain et Crowley, 2013)	Cl	(3)8°	(3)6.9°	-	Pointing'04
HOG + PLS noyau RBF (Al-Haj *et al.*, 2012)	Rg	(3)6.6°	(3)6.5°	-	Pointing'04
SVM + SVR (Guo *et al.*, 2008)	Rg + Cl	(3)7.7°	(3)9.2°	-	Pointing'04
SIFT + SVM + SVR + RP (Ho et Chellappa, 2012)	Rg + Cl	(3)5.8°	(3)6.1°	-	Pointing'04
LGO + SVR (*) (Murphy-Chutorian *et al.*, 2007)	Rg	(3)4.6°	(3)8.5°	-	LISAP
LGO + SVR + Filtrage Particulaire (*) (Murphy-Chutorian et Trivedi, 2010)	Rg + Sv	(3)8.6°	(3)11.2°	(3)8.3°	LISAP-14

Nous avons présenté dans ce chapitre divers systèmes dédiés à l'estimation de la pose de la tête. Nous illustrons dans le tableau 4.3 un résumé de quelques estimateurs en incluant les approches sur lesquelles ils se basent et les catégories auxquelles ils appartiennent. Nous pouvons conclure de ce tableau que les méthodes basées sur la classification sont les plus étudiées et les plus prometteuses. Aussi, nous pouvons en déduire que plusieurs auteurs élaborent les tests sur leurs propres données qu'ils gardent confidentielles, mais que Pointing'04 reste la base de données

publique la plus utilisée.

Dans notre thèse, nous désirons estimer la pose de la tête du conducteur afin de détecter son inattention. De ce fait, nous n'avons pas besoin de relever les orientations précises de la tête, mais plutôt un ensemble de classes de poses dans lesquels le conducteur est attentif ou pas. Après des observations méticuleuses du comportement du conducteur (détaillées dans la section 5.4), nous avons opté pour la construction d'estimateurs portant sur deux angles de liberté, à savoir le pitch et le yaw. D'après ces mêmes observations, nous pouvons déduire qu'il est suffisant d'estimer trois poses selon le pitch (frontale, haute et basse) et trois poses selon le yaw (frontale, gauche et droite) afin de déterminer le niveau d'attention du conducteur.

En tenant compte de ces contraintes et après avoir étudié l'état de l'art et visualisé les résultats pour chaque type d'estimateur, nous avons conclu que nous sommes en présence d'un problème d'estimation discrète de la pose de la tête, qui pourra être résolu par des techniques basées sur la classification. Nous avons choisi ce type d'approches grâce à son efficacité, à son adaptation aux applications temps-réel, et à sa capacité à ignorer la variation de l'apparence qui ne correspond pas au changement de la pose quand il est associé à un descripteur adapté. Dans ce qui suit, nous proposons deux estimateurs basés sur la classification, que nous détaillons dans le chapitre 5 et le chapitre 6. Ces deux estimateurs prennent en considération l'hypothèse de référence qui stipule que « les filtres basés sur les orientations sont les plus adaptés pour déterminer les caractéristiques spécifiques à la pose » Sherrah *et al.* (2001). Le premier estimateur est basé sur un apprentissage robuste appliqué sur des templates d'apparence extraits en utilisant un outil puissant pour la transformation de l'image. Nous avons choisi d'utiliser une transformation multi-échelle et multi-orientation par les pyramides orientables (Simoncelli *et al.*, 1992) ainsi qu'un apprentissage probabiliste par la fonction de vraisemblance paramétrique (Toyama et Blake, 2002). Le second estimateur est basée sur la classification et la fusion de plusieurs descripteurs de l'image permettant de relever les orientations de la pose de la tête. Nous utilisons une classification par SVM multi-classes et une fusion de quatre descripteurs.

CHAPITRE

ESTIMATION DE LA POSE DE LA TÊTE BASÉE SUR LA PYRAMIDE ORIENTABLE ET L'APPRENTISSAGE PROBABILISTE

Sommaire

5.1 Introduction **77**
5.2 Modélisation de la pose de la tête par la pyramide orientable **78**
5.2.1 Théorie des filtres orientables 78
5.2.2 Décomposition en pyramide orientable 80
5.2.3 Construction des templates de référence 82
5.3 Estimation de la pose de la tête par la fonction de vraisemblance paramétrique **83**
5.4 Estimation de la pose de la tête du conducteur par la pyramide orientable et la fonction de vraisemblance paramétrique **84**
5.4.1 Formulation du problème 85
5.4.2 Estimation de la pose de la tête du conducteur 85
5.5 Résultats expérimentaux **86**
5.5.1 La base de données Pointing'04 86
5.5.2 Optimisation des paramètres 87
5.5.3 Comparaison 90
5.5.4 Estimation de la pose de la tête appliquée à la séquence du conducteur 91
5.6 Conclusion **92**

5.1 Introduction

Nous proposons dans ce chapitre, un estimateur discret de la pose de la tête basé sur les templates d'apparence et un apprentissage probabiliste. En effet, nous développons une méthode hybride qui combine les templates d'apparence à une technique de classification assez originale, comme nous les avons détaillés dans le sous-section 4.1.3. Notre approche consiste à modéliser chaque pose discrète par un template de référence, selon la procédure que nous présentons dans la section 5.2. Ces templates sont élaborés à partir d'une transformation de l'image par une pyramide orientable « Steerable Pyramid » (SP) (Simoncelli *et al.*, 1992), qui correspond à une décomposition multi-échelle par des filtres orientables « Steerable Filters » (SF) (Freeman et Adelson, 1991). Notre choix a porté sur ces filtres puisqu'ils permettent une robustesse face aux déformations géométriques et aux changements des points de vue, ce qui rend les templates d'apparence plus performants. Un autre avantage fourni par ce filtrage à orientations sélectives

est lié à sa capacité de générer des images filtrées à n'importe quelle orientation par une combinaison linéaire d'un ensemble de filtres de base, ce qui permet de réduire considérablement le temps de calcul. Après avoir défini le template de référence pour chaque pose, nous effectuons un apprentissage des paramètres de la fonction de vraisemblance paramétrique « Likelihood Parametrized Function » (LPF) en utilisant une technique probabiliste. Nous utilisons cette fonction pour estimer la congruence entre une nouvelle image de la tête et les templates de référence. Nous détaillons dans la section 5.3, les différentes étapes de l'estimation de la pose de la tête par la fonction de vraisemblance paramétrique « Likelihood Parametrized Function » (LPF). Ensuite, nous expliquons dans la section 5.4, la procédure proposée pour estimer la pose de la tête du conducteur afin de détecter son inattention. Enfin, dans la section 5.5, nous présentons des résultats expérimentaux pour valider notre approche.

5.2 Modélisation de la pose de la tête par la pyramide orientable

La modélisation de la pose de la tête est une étape essentielle du processus de l'estimation de la pose. En effet, elle permet de construire une représentation de l'apparence de la tête en prenant en considération les variations produites par les changements de l'orientation. Puisque nous désirons analyser les structures orientées caractérisant la pose de la tête, nous avons choisi d'utiliser la décomposition en SP. Le concept fondamental de cette transformation de l'image réside dans la théorie des SF que nous détaillons dans la sous-section 5.2.1. Ensuite, nous définissons le principe de la décomposition en SP dans la sous-section 5.2.2. Enfin, nous expliquons dans la sous-section 5.2.3, la procédure proposée pour construire le template de référence pour chaque pose discrète à estimer.

5.2.1 Théorie des filtres orientables

Les SF constituent le concept principal de la décomposition en SP. Une fonction $f(x, y)$ est dite orientable si ses versions orientées $f^{\theta}(x, y)$ par un angle $theta$ peuvent être exprimées par une combinaison linéaire de M fonctions de base $f^{\theta_j}(x, y)$, comme exprimé par l'équation 5.1 où $k_j(\theta)$ représentent les fonctions d'interpolation correspondantes ($j = 1\ cdots M$).

$$f^{\theta}(x, y) = \sum_{j=1}^{M} k_j(\theta) f^{\theta_j}(x, y) \tag{5.1}$$

Nous désirons déterminer quelle fonction $f(x, y)$ peut satisfaire l'équation 5.1, combien de termes M sont nécessaires et que pouvons-nous choisir comme fonctions d'interpolations $k_j(\theta)$. Nous considérons par la suite la représentation polaire des coordonnées ($r = \sqrt{x^2 + y^2}$, $\phi = arg(x, y)$). Soit f une fonction qui peut être étendue par des séries de Fourier avec l'angle ϕ, exprimée par l'équation 5.2.

$$f^{\theta}(r, \phi) = \sum_{n=-N}^{N} a_n(r) e^{in\phi} \tag{5.2}$$

La condition de l'orientabilité présentée par l'équation 5.1, est valable pour les fonctions sous la forme de l'équation 5.2 si et seulement si les fonctions d'interpolation $k_j(\theta)$ sont des solutions

de l'équation 5.3.

$$\begin{bmatrix} 1 \\ exp(i\theta) \\ \vdots \\ exp(iN\theta) \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ exp(i\theta_1) & exp(i\theta_2) & \cdots & exp(i\theta_M) \\ \vdots & \vdots & \vdots & \vdots \\ exp(iN\theta_1) & exp(iN\theta_2) & \cdots & exp(iN\theta_M) \end{bmatrix} \begin{bmatrix} k_1(\theta) \\ k_2(\theta) \\ \vdots \\ k_M(\theta) \end{bmatrix} \tag{5.3}$$

De l'équation 5.1 et l'équation 5.2, nous pouvons déduire l'équation 5.4, où $g_j(r, \phi)$ peut correspondre à un ensemble quelconque de fonctions.

$$f^\theta(r, \phi) = \sum_{j=1}^{M} k_j(\theta) g_j(r, \phi) \tag{5.4}$$

Le nombre minimal de fonctions de base T correspond au nombre des $a_n(r) \neq 0$ dans l'équation 5.2, ce qui implique que $M \geq T$. Il est possible de choisir des versions orientées de f comme fonctions de base. Dans ce cas, nous calculons les T orientations θ_j des fonctions de base, réparties entre 0 and π, et données par l'équation 5.5.

$$\theta_j = \frac{j\pi}{T}, (j = 0 \cdots T - 1) \tag{5.5}$$

Fonction orientable choisie :

Pour déterminer les templates de référence de la pose de la tête, nous choisissons d'utiliser une fonction orientable simple introduite par Freeman et Adelson (1991), dont nous présentons la définition dans ce qui suit.

Soit $f(x, y)$ une fonction gaussienne 2D circulairement symétrique, exprimée par l'équation 5.6.

$$f(x, y) = e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{5.6}$$

Dans (Freeman et Adelson, 1991), l'orientabilité des dérivées directionnelles a été démontrée. Si nous notons f_1 la première dérivée de f, alors f_1 est une fonction orientable. Les filtres de bases $f_1^{0^\circ}$ et $f_1^{90^\circ}$ de cette fonction sont donnés par l'équation 5.7, et correspondent respectivement aux dérivées premières selon les directions x et y.

$$\begin{aligned} \frac{\partial}{\partial x} f(x, y) &= f_1^{0^\circ} = -\tfrac{1}{\sigma^2} x e^{-\frac{(x^2+y^2)}{2\sigma^2}} \\ \frac{\partial}{\partial y} f(x, y) &= f_1^{90^\circ} = -\tfrac{1}{\sigma^2} y e^{-\frac{(x^2+y^2)}{2\sigma^2}} \end{aligned} \tag{5.7}$$

D'après l'équation 5.1, le filtre f_1 à une orientation quelconque θ peut être synthétisé par une combinaison linéaire des filtres de bases $f_1^{0^\circ}$ and $f_1^{90^\circ}$, comme le montre l'équation 5.8.

$$f_1^\theta = cos(\theta) f_1^{0^\circ} + sin(\theta) f_1^{90^\circ} \tag{5.8}$$

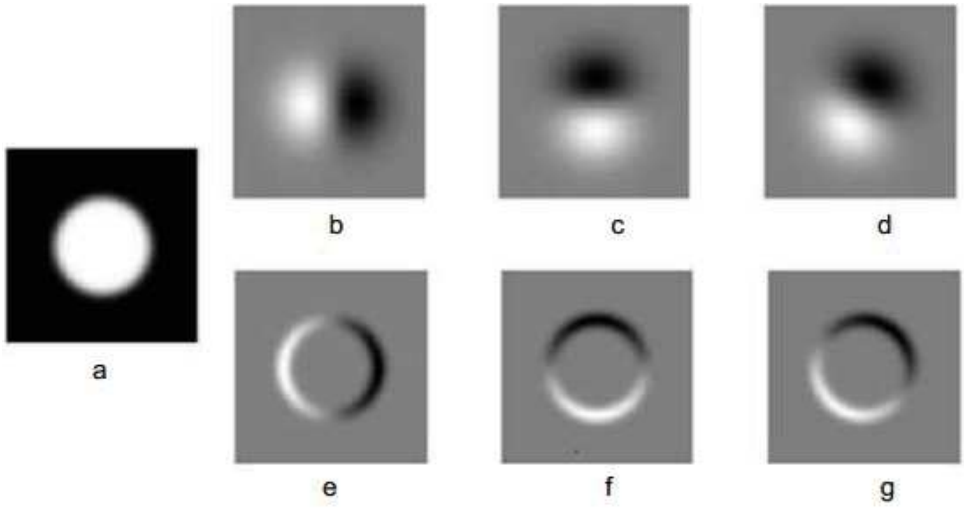


Figure 5.1 – Application des SF choisis sur l'image d'un disque (Freeman et Adelson, 1991). (a) Image du disque ; (b) $f_1^{0^\circ}$; (c) $f_1^{90^\circ}$; (d) $f_1^{30^\circ}$; (e) $R_1^{0^\circ}$; (f) $R_1^{90^\circ}$; (g) R_1^{θ} avec $\theta = 30$.

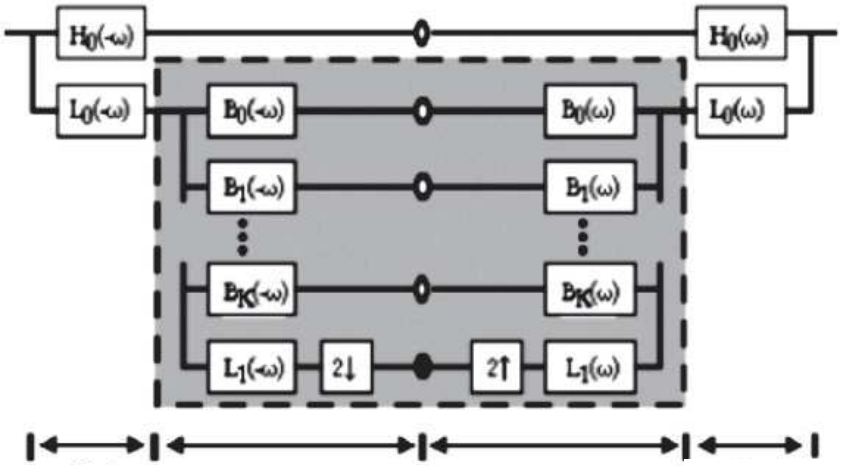
Puisque la convolution (notée $*$) est une opération linéaire, une image I filtrée à n'importe quelle orientation θ peut être synthétisée par la combinaison linéaire des images $R_1^{0^\circ}$ et $R_1^{90^\circ}$, correspondant au filtrage de I par les filtres de base $f_1^{0^\circ}$ et $f_1^{90^\circ}$. Le processus de filtrage de l'image I par le filtre orientable choisi f_1^{θ} est exprimé par l'équation 5.9

$$\begin{aligned} R_1^{0^\circ} &= f_1^{0^\circ} * I \\ R_1^{90^\circ} &= f_1^{90^\circ} * I \\ R_1^{\theta} &= cos(\theta) R_1^{0^\circ} + sin(\theta) R_1^{90^\circ} \end{aligned} \tag{5.9}$$

La figure 5.1 illustre un exemple de l'application des SF choisis sur l'image d'un disque (voir la figure 5.1-a). La figure 5.1-b et la figure 5.1-c correspondent respectivement aux filtres de base à 0° et 90°, tandis que la figure 5.1-e et figure 5.1-f représentent le résultat du filtrage de l'image par ces deux filtres. Si nous considèrons $\theta = 30^\circ$, le SF à cette orientation est représenté par la figure 5.1-d et le résultat du filtrage correspond à la figure 5.1-g.

5.2.2 Décomposition en pyramide orientable

L'idée principale derrière l'utilisation d'une décomposition pyramidale est que chaque ensemble de caractéristiques existe dans une certaine échelle de l'image. La transformation en ondelettes (Chui, 1992) est l'une des décompositions pyramidales les plus connues. Toutefois, cette transformation souffre de limites importantes qui correspondent à l'aliasing et la non représentation des orientations obliques. Ainsi, la décomposition en SP a été proposée par Simoncelli *et al.* (1992) pour résoudre ces problèmes et permettre une décomposition multi-orientation et multi-échelle de l'image. En effet, cette décomposition est invariante par rapport à la translation (les sous-bandes ne présentent pas d'aliasing, elles sont équivariantes par rapport à la translation) et aussi par rapport à la rotation (les sous-bandes sont orientables, elles sont équivariantes par


H0(-ω)
L0(-ω)
B0(-ω)
B1(-ω)
BK(-ω)
L1(-ω)
2↓
2↑
L1(ω)
BK(ω)
B1(ω)
B0(ω)
L0(ω)
H0(ω)

– Quand l'image est filtrée par L_1 et les filtres B_k, la somme des sorties de ces filtres doit produire l'image originale, qui est dans ce cas l'image filtrée par L_0. Cela peut s'écrire sous la forme de l'équation : $|L_1(\omega)|^2 + \sum_{(k=0)}^{n} |B_k(\omega)|^2 = 1$

Pour la construction des SF B_k, nous utilisons l'équation 5.8. Le nombre de niveaux de la pyramide ainsi que le nombre de SF pour chaque niveau seront déterminés par une étude expérimentale détaillée que nous présentons dans la section 5.5.

Dans (Castleman *et al.*, 1998), les auteurs proposent de construire les filtres H_0, L_0 et L_1 en utilisant le cosinus surélevé :

– Le filtre H_0 est donné par $H_0(u, v) = PH(f_2, f_N, s)$, avec PH une fonction de transfert à contours floutés donnée par le cosinus surélevé et exprimé par l'équation 5.11. a et b sont des paramètres établissant les limites des bandes.

$$PH(a, b, f) = \begin{cases} 0 & \text{pour} \quad f \leq a \\ \sqrt{\frac{1}{2}\left[1 - \cos\left[\pi\left(\frac{f-a}{b-a}\right)\right]\right]} & \text{pour} \quad a < f < b \\ 1 & \text{pour} \quad f \geq b \end{cases} \tag{5.11}$$

– Les filtres passe-bas sont donnés par : $L_0(u, v) = PB(f_2, f_N, s)$ et $L_1(u, v) = PB(f_1, f_N/2, s)$, avec PB une fonction de transfert à contours floutés donnée par le cosinus surélevé et exprimée par l'équation 5.12.

$$PB(a, b, f) = \begin{cases} 1 & \text{pour} \quad f \leq a \\ \sqrt{\frac{1}{2}\left[1 - \cos\left[\pi\left(\frac{f-a}{b-a}\right)\right]\right]} & \text{pour} \quad a < f < b \\ 0 & \text{pour} \quad f \geq b \end{cases} \tag{5.12}$$

5.2.3 Construction des templates de référence

Pour estimer K poses discrètes de la tête, nous aurons besoin de définir K templates de référence. Chaque pose doit être représentée par un nombre suffisant d'images. Pour chaque image i impliquée dans la construction du template pour la pose k, nous localisons le patch de la tête en utilisant une segmentation de l'image en pixels peau et non peau. Ensuite, nous appliquons sur ce patch une décomposition en SP en utilisant n niveaux et j SF (les valeurs optimales de ces paramètres seront déterminées dans la sous-section 5.5.2). Nous représentons le résultat de la décomposition dans un seul vecteur caractéristique v_i, qui remplacera l'image i. Après avoir traité toutes les images m dédiées à la construction du template pour la pose k, nous calculons la moyenne E_k des vecteurs caractéristiques associés à cette pose k, exprimée par l'équation 5.13.

$$E_k = mean(v_k^1, \cdots, v_k^m) \tag{5.13}$$

Chaque vecteur caractéristique moyen E_k représente le template de référence de la pose k. Nous notons ξ l'ensemble des vecteurs caractéristiques moyens, $\xi = (E_1, \cdots, E_K)$. Nous calculons également la matrice de covariance diagonale σ_k par l'équation 5.14 et nous notons Σ l'ensemble

des σ_k, $\Sigma = (\sigma_1, \cdots, \sigma_K)$.

$$\sigma_k = diag(cov(v_k^1, \cdots, v_k^m)) \tag{5.14}$$

Cet ensemble sera utilisé pour le calcul de la LPF.

5.3 Estimation de la pose de la tête par la fonction de vraisemblance paramétrique

Dans (Ricci et Odobez, 2009; Smith *et al.*, 2008), un apprentissage probabiliste, qui correspond à la LPF, est présenté comme une mesure de compatibilité entre une nouvelle entrée et le template de référence d'une certaine pose. Cette fonction est exprimée par un ensemble de paramètres appris afin de fournir une grande similarité si la pose de l'entrée et du template de référence sont proches. En effet, les images de test seront analysées en termes d'un modèle probabiliste déterminé à partir des images d'apprentissage. Dans (Toyama et Blake, 2002), une approche pour le suivi des objets utilisant des templates d'apparence avec un mécanisme probabiliste est introduite pour la première fois. Au lieu d'utiliser un apprentissage classique (SVM par exemple), les auteurs emploient une approche de mixture de métriques basée sur la LPF. Notre approche est inspirée des travaux de Ricci et Odobez (2009) et de Toyama et Blake (2002) qui utilisent la LPF pour traiter le problème du suivi. Dans notre cas, nous définissons cette fonction pour estimer la pose de la tête. Dans ce qui suit, nous présentons les différentes étapes de l'apprentissage de la LPF pour l'estimation de la pose de la tête à partir des données utilisées pour la construction des templates de référence.

La LPF d'une image représentée par son vecteur caractéristique v, étant donnée une pose de la tête k, est exprimée par l'équation 5.15

$$p(v|k) = \frac{1}{Z_k} e^{(-\lambda_k \rho_k(v, E_k))} \tag{5.15}$$

Où :

E_k : le template de référence de la pose k, donné par l'équation 5.13

σ_k : la matrice de covariance diagonale exprimée par l'équation 5.14

Z_k : une fonction de partition, aussi appelée constante de normalisation ;

λ_k : le paramètre de l'exponentiel.

ρ_k : la distance de Mahalanobis tronquée et normalisée, donnée par l'équation 5.16 (Smith *et al.*, 2008).

$$\rho_k(v, E_k) = \frac{1}{n} \sum_{i=1}^{n} max \left\{ \left(\frac{(v(i) - E_k(i))}{\sigma_k} \right)^2, T^2 \right\} \tag{5.16}$$

Avec :

T : un seuil permettant d'exclure les valeurs inappropriées pour la distance (fixé à 3).

Les paramètres de la LPF qui devront être appris correspondent à λ_k et Z_k. Cependant, pour apprendre la valeur d'un paramètre de l'exponentiel λ_k à partir des données d'entrainement, il est nécessaire d'avoir des connaissances préalables de la fonction de partition Z_k. Généralement, cela est difficile, mais faisable quand la distance utilisée peut être approchée par une distribution

gaussienne, comme c'est le cas pour la distance ρ_k choisie.

Dans ce cas précis, les paramètres de la LPF peuvent être exprimés par l'équation 5.17

$$\begin{aligned} \lambda_k &= \frac{1}{2\delta_k^2} \\ Z_k &= \delta_k^{d_k} \end{aligned} \tag{5.17}$$

D'après la forme de l'équation 5.17, Toyama et Blake (2002) ont démontré que la distance ρ_k peut être approchée par une variable aléatoire $\delta_k^2\chi_{d_k}^2$ suivant une loi χ^2, avec δ_k son écart-type et d_k sa dimension. Cette approximation permet l'apprentissage des paramètres δ_k et d_k à partir des données d'entrainement. Pour ceci, nous construisons un ensemble F_v à partir de ces données. Pour chaque $f_v \epsilon F_v$, nous déterminons la pose k permettant de minimiser la distance ρ entre f_v et tous les templates de référence E_p, comme exprimé par l'équation 5.18.

$$k = arg\min_p \rho_p(f_v, E_p) \tag{5.18}$$

Nous notons $\rho_k(f_v) = \rho_k(f_v, E_k)$ pour simplifier l'écriture. Rappelons que $\rho_k(f_v)$ peut être approchée par $\delta_k^2\chi_{d_k}^2$. Nous pouvons ainsi approcher les paramètres de la loi χ^2 par les moments simples, comme exprimé par l'équation 5.19.

$$\begin{aligned} \overline{\rho}_k &= \frac{1}{N_k}\sum_{f_v \epsilon k} \rho_k(f_v) \\ \overline{\rho^2}_k &= \frac{1}{N_k}\sum_{f_v \epsilon k} \rho_k^2(f_v) \end{aligned} \tag{5.19}$$

À partir de la forme de la moyenne et de l'écart-type de la loi χ^2, δ_k et d_k peuvent être estimés par l'équation 5.20.

$$\begin{aligned} d_k &= 2\frac{\overline{\rho}_k^2}{\overline{\rho^2}_k - \overline{\rho}_k^2} \\ \delta_k &= \sqrt{\frac{\overline{\rho}_k}{d_k}} \end{aligned} \tag{5.20}$$

En replaçant l'équation 5.20 dans l'équation 5.17, nous obtenons les paramètres de la LPF, exprimée par l'équation 5.15.

5.4 Estimation de la pose de la tête du conducteur par la pyramide orientable et la fonction de vraisemblance paramétrique

Dans la section 5.3, nous avons proposé une approche d'apprentissage probabiliste des paramètres de la LPF pouvant être utilisée pour n'importe quelle application qui requière une estimation discrète de la pose de la tête. Dans la sous-section 5.4.1, nous présentons notre formulation du problème de la détection de l'inattention chez le conducteur basée sur l'estimation de la pose de la tête, et dans la sous-section 5.4.2, nous expliquons la procédure proposée pour résoudre ce problème. Par la suite, nous désignons notre approche par l'abréviation SP-LPF.

5.4.1 Formulation du problème

Dans (Miyaji *et al.*, 2009), les auteurs ont étudié la distraction cognitive chez le conducteur et ont conclu que le mouvement de la tête est un facteur révélateur de l'inattention. L'observation des orientations de la tête permet de conclure que le conducteur est attentif à la route qu'en cas de position frontale. Cependant, en conduisant, il est aussi nécessaire de vérifier les rétroviseurs et le tableau de bord ou bien faire une marche arrière, ce qui implique que nous devons tourner la tête vers la gauche, la droite, le haut et le bas pour une brève durée. Ces positions peuvent être synthétisées par le pitch (mouvement de haut en bas) et le yaw (mouvement de gauche à droite) uniquement, sans avoir recours au roll (mouvement de la tête en direction des épaules). Martin *et al.* (2012) ont prouvé que pendant une conduite typique, le conducteur passe 95% du temps en regardant en face de lui et que les 5% du temps restant correspondent à des positions non frontales qui peuvent révéler des états d'inattention. Nous pouvons donc suggérer que les poses non frontales ne peuvent être maintenues que pour quelques secondes, ce qui est suffisant pour vérifier les rétroviseurs par exemple. En cas de prolongement de la durée des poses non frontales, nous considérons que nous sommes en présence d'un état d'inattention du conducteur. Ainsi, nous pouvons déduire de cette analyse qu'il est suffisant d'estimer trois poses selon le pitch (frontale, haute et basse) et trois poses selon le yaw (frontale, gauche et droite) afin de déterminer le niveau d'attention du conducteur.

5.4.2 Estimation de la pose de la tête du conducteur

Nous proposons d'effectuer l'apprentissage de deux LPF, la première pour le pitch et la seconde pour le yaw, puisque nous supposons que ces deux angles sont indépendants, comme proposé par Munoz-Salinas *et al.* (2012). Après voir déterminé les templates de référence (voir la section 5.2) pour chaque pose et pour chaque angle, nous pouvons effectuer l'apprentissage des paramètres des deux LPF selon la procédure proposée dans la section 5.3. Pour chaque frame de la séquence vidéo du conducteur, nous appliquons le processus suivant :

- Localiser le patch de la tête en utilisant la même technique de segmentation de l'image en pixels peau et non peau adoptée pour la construction des templates de référence.
- Appliquer la décomposition en SP pour construire le vecteur caractéristique v (voir la section 5.2). Les paramètres optimaux de la SP seront déterminés par l'étude présentée dans la sous-section 5.5.2.
- Calculer la LPF définie pour le pitch entre le vecteur v et tous les templates de référence, décrite par l'équation 5.15. La pose de la tête estimée selon le pitch k^*_{pitch} correspond à la pose ayant fourni la valeur maximal pour la LPF, comme indiqué par l'équation 5.21

$$k^*_{angle} = arg \max_k p(v|k) \tag{5.21}$$

- Si la tête est baissée ou levée, nous observons la durée pour laquelle une seule position est fixée et nous émettons une alarme d'inattention quand cette durée est importante.
- Si la pose selon le pitch est frontale, nous calculons la LPF définie pour le yaw entre le vecteur v et tous les templates de référence. La pose de la tête estimée selon le yaw k^*_{yaw}

est aussi déduite à partir de l'équation 5.21.

– Si la tête est tournée à gauche ou à droite, nous observons la durée pour laquelle une seule position est fixée et nous émettons une alarme d'inattention quand cette durée est importante.

5.5 Résultats expérimentaux

Le premier test que nous avons effectué correspond à une étude expérimentale pour déterminer les paramètres optimaux de la SP afin d'estimer la pose de la tête. Nous avons à maintes reprises contacté l'équipe LISA (voir section 4.2), ayant développé leur base de données en utilisant le testbed LISA-P, pour l'obtention de quelques frames afin de tester notre approche mais aucune demande n'a abouti. Puisqu'il n'existe pas de données publiques qui permettent de valider les approches de l'estimation de la pose de la tête du conducteur, nous avons acquis et annoté nos propres séquences vidéo correspondant à des conducteurs dans des situations simulées d'attention et d'inattention. Toutefois, pour ce premier test présenté dans la sous-section 5.5.2, mais aussi pour effectuer une comparaison avec des techniques existantes dans la sous-section 5.5.3, nous avons opté pour l'utilisation de la base de données publique Pointing'04, qui correspond à la base la plus exploitée dans la littérature pour estimer la pose de la tête (Murphy-Chutorian et Trivedi, 2009). Nous présentons brièvement cette base de données dans la sous-section 5.5.1. Une fois les paramètres optimaux de notre estimateur sont déterminés et des comparaisons avec des techniques définies dans le chapitre 4 sont présentées, nous exposons dans la sous-section 5.5.4, les résultats de l'estimation de la pose de la tête pour une séquence réelle d'un conducteur simulant l'inattention.

5.5.1 La base de données Pointing'04

Figure 5.3 – Poses frontales de la base Pointing'04

La base de données Pointing'04 représente 15 sujets différents sous 93 poses discrètes de la tête. Pour chaque sujet, deux séries d'images sont acquises pour toutes ces poses. Les orientations de la tête sont décrites par neuf poses selon le pitch ($\{0; \pm 90; \pm 60; \pm 30; \pm 15\}$) et treize poses selon

le yaw ($[-90°; +90°]$ avec un pas de 15°). La figure 5.3 affiche la pose frontale ($pitch = yaw = 0°$) pour chacune des deux séquences de test des 15 sujets.

Pour les tests où nous utilisons la base de données Pointing'04, nous considérons 80% de celle-ci pour l'apprentissage (2232 images, soit les 24 premières séries) et 20% pour le test (558 images, soit les 6 dernières séries).

5.5.2 Optimisation des paramètres

Nous rappelons que nous utilisons pour ce test la base publique Pointing'04. Comme nous l'avons expliqué dans la section 5.4, nous avons besoin d'identifier uniquement trois poses pour le pitch et trois poses pour le yaw afin de déterminer le niveau d'attention du conducteur. Cependant, la majorité des travaux utilisant Pointing'04 manipulent sa représentation standard (neuf poses pour le pitch et treize pour le yaw). Par conséquent, nous décidons d'optimiser les paramètres de la SP en utilisant la représentation standard de la base de données afin de permettre une comparaison équitable avec les techniques existantes. Les résultats sont présentés en termes de MAE entre la vérité terrain et la pose estimée pour le pitch et le yaw séparément. Pour obtenir des mesures précises du temps d'exécution, nous contrôlons notre machine, équipée d'un processeur multi-cœur Intel i3, afin de permettre à Matlab d'être l'unique processus à s'exécuter sur l'un des cœurs. Les temps d'exécution présentés par la suite sont exprimés pour une image de test en millisecondes (ms) et correspondent à la durée moyenne de plusieurs exécutions pour un seul degré de liberté.

- **Optimisation des paramètres des filtres orientables**

Pour cette première expérimentation, nous effectuons plusieurs tests afin de déterminer le nombre optimal de SF (nb_{filt}) à considérer pendant la construction de la SP. En général, les orientations des SF commencent par $\theta = 0°$ et sont ensuite incrémentées en utilisant un pas fixe noté *step*. Par exemple, lorsque trois SF sont utilisés avec $step = 60°$, leurs orientations correspondent à $\Theta = \{0°, 60°, 120°\}$. Pour ce test, nous varions le nombre de SF de 2 à 10 et le pas d'orientation 20° à 90°. Le tableau 5.1 et le tableau 5.2 affichent les résultats correspondant respectivement au pitch-MAE et au yaw-MAE. Nous observons que certaines valeurs sont redondantes puisque certaines informations produites par des orientations différentes sont similaires en raison de la modularité angulaire.

Dans ces deux tableaux, la MAE minimale pour chaque pas d'orientation *step* est exprimée en caractères gras. Nous pouvons conclure que les trois plus petites MAE (valeurs soulignées) sont observées pour des grands pas d'orientation et correspondent à $step = \{90°, 80°\}$ avec $nb_{filt} = 3$ et $step = 60°$ avec $nb_{filt} = 4$. Puisque le pas d'orientation n'influence pas le temps d'exécution, nous affichons dans la figure 5.4 le temps d'exécution moyen d'une image de test en variant uniquement le nombre de SF. Le traitement d'une image de test avec un code Matlab non optimisé prend environ 65 ms pour $nb_{filt} = 3$ et 74 ms pour $nb_{filt} = 4$.

Tableau 5.1 – Pitch-MAE en variant nb_{filt} et *step*

	nb_{filt}								
step	2	3	4	5	6	7	8	9	10
20	13.01	13.38	13.06	**12.66**	12.66	13.27	12.95	12.66	12.68
30	13.19	13.11	**12.41**	13.14	12.66	12.47	12.63	12.44	12.66
40	13.17	12.39	12.93	12.60	**12.30**	12.47	12.84	12.66	12.68
50	12.90	12.71	12.60	**12.33**	12.71	12.71	12.63	12.58	12.68
60	13.06	12.80	**12.18**	12.55	12.80	12.55	12.66	12.80	12.80
70	12.58	**12.20**	12.33	12.74	12.39	12.60	12.79	12.60	12.71
80	12.66	**11.98**	12.71	12.55	12.58	12.79	12.63	12.66	12.68
90	12.74	**11.85**	12.74	12.44	12.74	12.50	12.74	12.66	12.75

Tableau 5.2 – Yaw-MAE en variant nb_{filt} et *step*

	nb_{filt}								
step	2	3	4	5	6	7	8	9	10
20	**8.95**	9.05	9.11	9.73	10.37	11.31	10.96	10.08	9.70
30	**8.87**	9.11	10.01	11.07	10.08	9.43	9.16	9.48	10.01
40	**8.73**	9.65	10.51	9.73	9.11	9.48	10.26	10.08	9.70
50	**8.73**	10.40	9.73	9.08	9.89	10.13	9.73	9.48	10.08
60	9.08	10.19	**8.54**	9.48	10.16	9.46	9.78	10.16	9.65
70	9.40	9.08	**9.05**	10.13	9.43	9.83	10.08	9.65	10.13
80	9.97	**8.49**	9.73	9.73	9.48	10.02	9.65	10.08	9.70
90	10.13	**8.44**	10.13	9.13	10.13	9.48	10.13	9.67	10.13

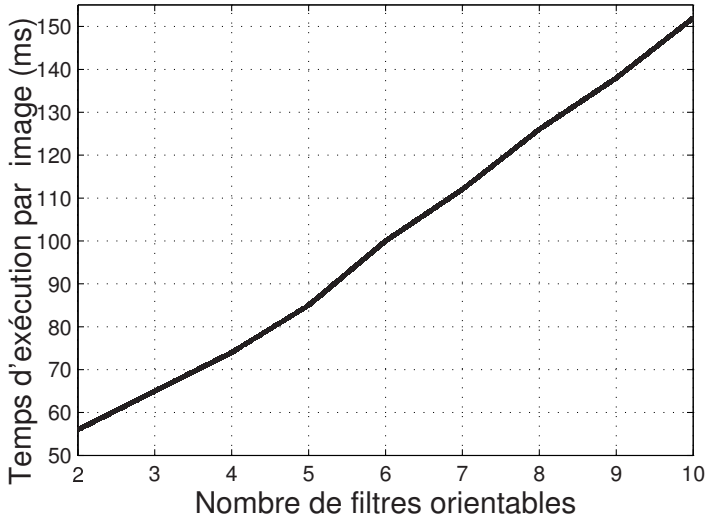


Figure 5.4 – Temps d'exécution pour une image de test en variant nb_{filt}

- **Optimisation des paramètres de la pyramide orientable**

Pour étendre l'étude portant sur l'optimisation des paramètres de notre estimateur, nous considérons les trois meilleurs valeurs des paramètres des SF (nb_{filt} et *step*), soulignées dans le tableau 5.1 et le tableau 5.2. Pour chaque combinaison de ces paramètres, nous varions le nombre de niveaux de la SP, noté *level*. En plus du nombre considéré de SF, nous testons

Tableau 5.3 – Pitch-MAE en variant *level* et en considérant les trois meilleurs valeurs pour nb_{filt} et *step*

		level			
step	nb_{filt}	**1**	**2**	**3**	**4**
60	3	12.80	**<u>10.86</u>**	11.42	12.45
60	4	12.18	**11.67**	12.66	15.75
80	2	12.66	**11.40**	11.61	13.60
80	3	11.98	**11.77**	12.55	15.73
90	2	12.74	**10.94**	11.56	12.47
90	3	11.85	**12.10**	13.23	16.85

Tableau 5.4 – Yaw-MAE en variant *level* et en considérant les trois meilleurs valeurs pour nb_{filt} et *Step*

		level			
step	nb_{filt}	**1**	**2**	**3**	**4**
60	3	10.19	**7.52**	7.93	10.2
60	4	8.54	**7.20**	8.22	12.28
80	2	9.97	**7.50**	8.09	10.73
80	3	8.49	**7.12**	7.98	12.45
90	2	10.13	**7.55**	7.93	10.22
90	3	8.44	**<u>7.09</u>**	8.71	13.28

notre estimateur sur moins de filtres afin d'évaluer la possibilité de réduire leur nombre, afin de minimiser le temps d'exécution. Les paramètres pris en compte dans ce test sont : $nb_{filt} = \{2, 3\}$ avec $step = \{90°, 80°\}$ et $nb_{filt} = \{3, 4\}$ avec $step = 60°$.

Les résultats sont présentés par le tableau 5.3 et le tableau 5.4 qui correspondent respectivement au pitch-MAE et au yaw-MAE.

Nous remarquons que l'utilisation de deux niveaux pour la SP (valeurs en gras) fournit la meilleure estimation de la pose pour tous les paramètres des SF. Cette observation pourrait s'expliquer par la présence dans ces deux niveaux de suffisamment d'informations pour représenter la pose de la tête. Toutefois, ces informations sont perturbées par l'ajout d'images de très petites tailles résultant de la décomposition en trois niveaux ou plus. D'après le tableau 5.3, nous observons que le pitch-MAE minimal (souligné) est obtenu par $level = 2$, $nb_{filt} = 3$ et $step = 60°$, tandis que pour le tableau 5.4, le yaw-MAE minimal (souligné) est donné par la même configuration des paramètres *level* et nb_{filt} mais avec $step = 90°$. Dans ce qui suit, nous considérons que les paramètres optimaux de la pyramide sont : $level = 2$, $nb_{filt} = 3$ et $step = 60°$. Nous avons choisi le pas d'orientation optimal du pitch ($step = 60°$) et non pas celui du yaw ($step = 90°$) puisque les valeurs $MAE_{yaw}(60°) = 7.52$ et $MAE_{yaw}(90°) = 7.09$ sont proches, alors que la valeur $MAE_{pitch}(90°)) = 12.10$ est plus importante que $MAE_{pitch}(60°)) = 10.86$.

La figure 5.5 affiche le temps d'exécution pour une image de test en utilisant les paramètres optimaux. À partir de cette figure, nous remarquons que le temps d'exécution est presque identique pour $level = \{2, 3, 4\}$. Cet effet peut s'expliquer par les variations minimes de la taille du vecteur caractéristique quand nous utilisons plusieurs niveaux pour la décomposition pyrami-

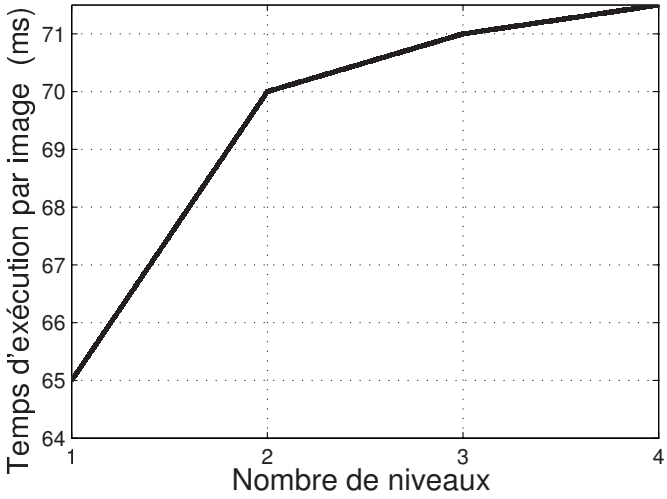


Figure 5.5 – Temps d'exécution pour une image de test en variant *level* et en considérant ($nb_{filt} = 3$, $step = 60°$)

dale. Par exemple, si nous considérons $nb_{filt} = 3$, $step = 60°$ et une taille de l'image égale à $size = 64^2$, nous obtenons les tailles suivantes du vecteur caractéristique en fonction des niveaux de la pyramide :

- $level = 1 : size = 3 * 64^2 = 12288$
- $level = 2 : size = 3 * (64^2 + 32^2) = 15360$
- $level = 3 : size = 3 * (64^2 + 32^2 + 16^2) = 16128$
- $level = 4 : size = 3 * (64^2 + 32^2 + 16^2 + 8^2) = 16320$

Le temps d'exécution moyen correspondant à la configuration optimale retenue pour la transformation en SP, pour un seul angle de liberté, est de 70 ms. Ainsi, nous pouvons conclure que la décomposition en SP ne consomme pas plus de temps que l'utilisation des SF (65 ms pour le même nombre de filtres).

5.5.3 Comparaison

Tableau 5.5 – Comparaison de l'approche SP-LPF avec la littérature en termes de pitch-MAE et yaw-MAE

Ligne	**Approche**	**Pitch-MAE**	**Yaw-MAE**
1	SP-LPF	10.86°	7.52°
2	SF-LPF	11.85°	8.44°
3	Performance humaine (Gourier *et al.*, 2007)	11°	11.9°
4	LAAM (Gourier *et al.*, 2007)	15.9°	10.03°
5	SVM + SVR (LARR) (Guo *et al.*, 2008)	7.69°	9.23°
6	SIFT + SVM + SVR (Ho et Chellappa, 2012)	5.84°	6.05°
7	PLS noyau RBF (Al-Haj *et al.*, 2012)	6.61°	6.56°
8	SP + SVM (Jain et Crowley, 2013)	8°	6.9°

Dans le tableau 5.5, nous exposons une comparaison entre notre estimateur de la pose de la

tête et quelques travaux de la littérature que nous avons détaillés dans le chapitre 4 et qui utilisent principalement la base de données Pointing'04. Les lignes 1 et 2 de ce tableau correspondent aux résultats obtenus par notre estimateur en considérant les paramètres optimaux pour la SP et les SF respectivement. Dans la ligne 3, nous rapportons le résultat de l'étude conduite par Gourier *et al.* (2007) pour déterminer la capacité de l'être humain à estimer la pose de la tête. Les estimateurs rapportés dans les lignes 3 et 4 correspondent à des travaux de référence fréquemment cités dans la littérature. Les résultats obtenus par SP-LPF sont meilleurs que ceux des travaux de référence. Dans la ligne 5, un estimateur combinant les SVM et SVR est utilisé sur des patchs de la tête extraits manuellement. Notre estimateur est plus performant pour le yaw que ce dernier. Malheureusement, le SP-LPF ne permet pas d'obtenir de meilleurs résultats que les estimateurs les plus récents correspondant aux lignes 6, 7 et 8 du tableau. Cependant, les résultats restent très acceptables en termes de MAE et temps d'exécution. Le temps de traitement d'une image de la base de données Pointing'04 n'est pas donné par toutes les approches. Toutefois, dans Jain et Crowley (2013), ce temps est estimé à 108 ms alors que notre approche permet un traitement en 70 ms, pour un seul angle de liberté.

5.5.4 Estimation de la pose de la tête appliquée à la séquence du conducteur

Après avoir validé notre estimateur de la pose de la tête sur la base de données Pointing'04, nous effectuons un test sur la séquence vidéo du conducteur acquise dans une voiture par un téléphone portable avec une caméra de résolution 1024 *times*768 pixels. La vidéo est composée de 1416 frames, nous utilisons 946 pour l'apprentissage et 470 pour le test. Puisque nous estimons la pose de la tête du conducteur, nous avons annoté trois classes pour le pitch et trois classes pour le yaw, comme indiqué dans la sous-section 5.4.1. Dans ce test, nous exprimons les résultats par le CCR au lieu de la MAE, puisque nous ne disposons pas de mesures angulaires précises. La figure 5.6 et la figure 5.7 affichent respectivement l'étape d'acquisition et les échantillons des différentes poses de la tête du conducteur.

Figure 5.7 – Frames du conducteur. (a) tête frontale (Pitch et Yaw) ; (b) Profil gauche (Yaw) ; (c) Profil droit (Yaw) ; (d) Tête haute (Pitch) ; (e) Tête basse (Pitch)

des mauvaises classifications.

À partir de ces deux tableaux et des formules présentées dans la sous-section 3.6.1, nous obtenons :

- Pitch : $CCR = 0.88$; $\kappa = 0.72$
- Yaw : $CCR = 0.86$; $\kappa = 0.74$

Ces résultats sont acceptables et prouvent que SP-LPF est adapté à l'estimation de la pose de la tête du conducteur, et donc à la détection de son inattention même pour une séquence soumise à des contraintes réelles.

Tableau 5.6 – Matrice de confusion de SP-LPF pour la séquence du conducteur selon le pitch

Réelle/Estimée	**Frontale**	**Haute**	**Basse**
Frontale	318	16	12
Haute	8	38	4
Basse	13	2	59

Tableau 5.7 – Matrice de confusion de SP-LPF pour la séquence du conducteur selon le yaw

Réelle/Estimée	**Frontale**	**Gauche**	**Droite**
Frontale	252	21	27
Gauche	12	86	2
Droite	9	3	67

5.6 Conclusion

Dans ce chapitre, nous avons défini un estimateur de la pose de la tête basé sur une transformation de l'image par une SP afin d'extraire des templates d'apparence pour chaque orientation à estimer. Ensuite, nous avons défini un apprentissage probabiliste qui permet d'apprendre les paramètres de la LPF afin d'établir une correspondance entre les templates et les nouvelles entrées. Nous avons effectué des séries de tests sur la base publique Pointing'04 pour détermi-

ner les paramètres optimaux de la SP et comparer SP-LPF à des travaux existants utilisant la même base. Notre estimateur est plus performant que certaines approches de référence, mais il existe d'autres techniques qui donnent de meilleurs résultats. En analysant les résultats de la comparaison, nous avons pensé à la conception d'une autre approche qui consiste à fusionner plusieurs descripteurs connus pour leur performance à discriminer la pose de la tête. La fusion de ces descripteurs nous permettra de construire des vecteurs caractéristiques plus robustes. Par la suite, nous utilisons un apprentissage par deux SVM multi-classes afin de déterminer la classe d'appartenance des nouvelles entrées. Dans le chapitre 6, nous présentons cette approche en détail.

Chapitre

Estimation de la pose de la tête basée sur la classification et la fusion de descripteurs

Sommaire

6.1 Introduction **95**
6.2 Vecteur caractéristique basé sur la fusion de descripteurs **96**
6.2.1 Descripteurs utilisés 96
6.2.1.1 Filtres orientables 96
6.2.1.2 Histogramme des Gradients Orientés (HOG) 96
6.2.1.3 Caractéristiques de Haar 97
6.2.1.4 Speeded-Up Robust Features (SURF) 97
6.2.2 Sélection des variables 98
6.2.2.1 Méthodes de recherche 99
6.2.2.2 Méthodes d'évaluation des attributs 99
6.3 Estimation de la pose de la tête du conducteur par des SVM multi-classes . . **100**
6.4 Résultats expérimentaux **101**
6.4.1 Optimisation des paramètres 101
6.4.2 Comparaison 104
6.4.3 Estimation de la pose de la tête appliquée à la séquence du conducteur 105
6.5 Conclusion **105**

6.1 Introduction

Dans ce chapitre, nous proposons un autre estimateur discret de la pose de la tête adapté à l'étude de l'inattention chez le conducteur. En analysant les résultats présentés par le tableau comparatif 5.5, nous avons remarqué que certains descripteurs comme SIFT ou les SF ainsi que les SVM, permettent d'obtenir de bons résultats. Dans la section 6.2, nous proposons de fusionner plusieurs descripteurs performants pour construire un vecteur caractéristique plus robuste pour discriminer la pose de la tête. Ensuite, pour déterminer la pose d'une nouvelle entrée, nous utilisons des SVM multi-classes, décrits par la section 6.3. Enfin, dans la section 6.4, nous présentons des résultats expérimentaux pour valider cette approche pour l'estimation de la pose de la tête.

6.2 Vecteur caractéristique basé sur la fusion de descripteurs

Nous présentons dans la sous-section 6.2.1 les quatre descripteurs de l'image que nous estimons être les plus représentatifs des variations de la pose de la tête. Ces descripteurs correspondent aux SF, HOG, caractéristiques de Haar et Speeded-Up Robust Features (SURF). Dans le chapitre 5, nous avons prouvé l'efficacité des SP pour l'estimation de la pose de la tête. Le HOG et les caractéristiques de Haar sont utilisés par plusieurs estimateurs de la pose de la tête (Murphy-Chutorian et Trivedi, 2009), mais aussi pour d'autres applications telles que la détection des piétons. SURF est choisi pour sa rapidité et sa robustesse pour la détection des objets. L'avantage de ces descripteurs est qu'ils sont tous invariants aux transformations de l'image qui correspondent à la rotation, le changement de l'échelle et la variation de l'éclairage. Ces propriétés les rendent parfaitement adaptés à la construction des vecteurs caractéristiques représentant la pose de la tête. Puisque les vecteurs résultants de la combinaison des descripteurs sont volumineux, nous définissons dans la sous-section 6.2.2 des techniques de sélection des variables qui permettent de réduire la dimension de ces vecteurs en choisissant les attributs les plus pertinents.

6.2.1 Descripteurs utilisés

6.2.1.1 Filtres orientables

Les SF ont été présentés en détail dans la sous-section 5.2.1. Nous les utilisons dans cette seconde approche puisque nous avons prouvé leur robustesse pour l'estimation de la pose de la tête dans le chapitre 5. Nous avons décidé d'utiliser les SF au lieu de la SP afin de réduire la taille des vecteurs caractéristiques de la pose. Nous rappelons que la performance des SF dépend du nombre de filtres utilisés nb_{filt} et du pas de l'orientation *step* (voir la sous-section 5.5.2).

6.2.1.2 Histogramme des Gradients Orientés (HOG)

Le concept du HOG a été introduit par Dalal et Triggs (2005). Le HOG considère que l'apparence de l'objet et sa forme peuvent être représentées par la distribution des gradients locaux de l'intensité ou par des directions du contour. Ce concept peut être implémenté en divisant l'image en petites régions (cellule) avec une taille prédéfinie, adaptée à la taille et la résolution de l'objet à représenter. Pour chaque cellule, les occurrences de l'orientation du gradient pour tous les pixels sont cumulées dans un histogramme local. Pour calculer le gradient, l'image est convoluée aux filtres $G_x = (-1, 0, 1)$ et $G_y = (-1, 0, 1)^T$. L'équation 6.1 est utilisée pour calculer l'orientation Or_G et la magnitude mg_G du gradient de chaque pixel. $I_x(x, y)$ et $I_y(x, y)$ correspondent au résultat du filtrage de l'image par G_x et G_y.

$$\begin{aligned} Or_G &= arctan(\frac{I_x(x,y)}{I_y(x,y)}) \\ mg_G &= \sqrt{I_x^2(x,y) + I_y^2(x,y)} \end{aligned} \tag{6.1}$$

Chaque histogramme d'orientation divise les angles du gradient en un nombre fixe d'intervalles (bins). Chaque pixel de la cellule participe à un vote. Ce vote est pondéré par la magnitude du gradient mg_G à l'emplacement du pixel, ce qui permet de fournir plus d'importance aux votes des pixels du contour. La représentation de l'image est formée par les histogrammes combinés et peut être améliorée par la normalisation du contraste des réponses locales afin de réduire les effets de l'éclairage. La normalisation est effectuée par l'accumulation d'une mesure de l'histogramme local (l'énergie) sur des groupes de cellules (blocs) et les résultats sont utilisés pour normaliser les cellules du bloc. Chaque cellule est présente dans plusieurs blocs, mais ses normalisations sont différentes car elles dépendent du bloc. Par conséquent, une cellule apparaît plusieurs fois dans le vecteur final avec différentes normalisations. Cette propriété semble introduire une redondance, mais en réalité elle contribue à l'amélioration des performances. Les blocs normalisés correspondent au descripteur HOG.

La performance du HOG dépend du nombre de cellules à considérer par ligne et par colonne, notés respectivement n_x et n_y. Le nombre des bins utilisés pour construire les intervalles des angles du gradient, noté $bins$, est aussi un paramètre important. En plus de l'invariance du HOG aux changements de l'orientation et de l'échelle, il est pratique pour représenter la pose de la tête grâce à sa rapidité de calcul.

6.2.1.3 Caractéristiques de Haar

Les caractéristiques de Haar ont été proposées par Papageorgiou et Poggio (2000) et correspondent à une représentation dense basée sur les ondelettes. La décomposition de Haar à deux dimensions d'une image de taille n^2 consiste en n^2 coefficients d'ondelettes de Haar distinctes. La première ondelette est l'intensité moyenne de tous les pixels de l'image. Les autres ondelettes sont calculées par la différence des intensités moyennes des carrées adjacents selon l'axe horizontal, vertical ou diagonal. Les variations de contraste entre les pixels des groupes adjacents sont utilisés pour déterminer les zones sombres et éclairées. La performance des caractéristiques de Haar dépend du nombre de coefficient n^2, qui n'est autre que la taille de l'image.

6.2.1.4 Speeded-Up Robust Features (SURF)

SURF est un algorithme de représentation et de comparaison des caractéristiques d'une image, proposé par Bay *et al.* (2008) et considéré comme une amélioration de l'algorithme SIFT. SURF est structuré en trois étapes : la détection des points d'intérêt, la construction du descripteur pour chaque point et la correspondance des descripteurs. L'originalité du SURF consiste à accélérer les opérations en utilisant des images intégrales pour obtenir une implémentation rapide de la convolution par les filtres. L'image intégrale $In(x, y)$ à la position (x, y) correspond à la somme des intensités des pixels à partir de l'origine de l'image jusqu'à cette position. L'image intégrale peut être obtenue en un seul parcours en utilisant la relation de récurrence présentée par l'équation 6.2.

$$\begin{aligned} s(x,y) &= s(x,y-1) + I(x,y) \\ In(x,y) &= I(x-1,y) + s(x,y) \end{aligned} \tag{6.2}$$

Pour l'étape de détection, les candidats pour les points d'intérêt sont extraits par les maxima locaux de l'opérateur Hessien. Un candidat est considéré comme un point d'intérêt si sa réponse dépasse un seuil fixe. Pour un pixel $X = (x, y)$ de l'image I, la matrice Hessienne $H(X, \sigma)$ pour le point X et l'échelle σ est donnée par l'équation 6.3

$$H(X, \sigma) = \begin{bmatrix} L_{xx}(X, \sigma) & L_{xy}(X, \sigma) \\ L_{xy}(X, \sigma) & L_{yy}(X, \sigma) \end{bmatrix} \tag{6.3}$$

avec $L_{xx}(X, \sigma)$: la convolution de la dérivée de second ordre de la gaussienne $\frac{\delta^2}{\delta x^2} g(\sigma)$, avec l'image I dans le point X. Les valeurs $L_{xy}(X, \sigma)$ et $L_{yy}(X, \sigma)$ sont obtenues de la même façon.

Les gaussiennes sont optimales pour l'analyse espace-échelle. Cependant, en pratique, elles doivent être discrétisées (voir la figure 6.1), ce qui produit une perte d'information pour les orientations de l'image modulo $\frac{\pi}{4}$. Ainsi, pour résoudre ce problème, une approximation des dérivées de second ordre de la gaussienne par des filtres carrés est utilisée (voir la figure 6.2).

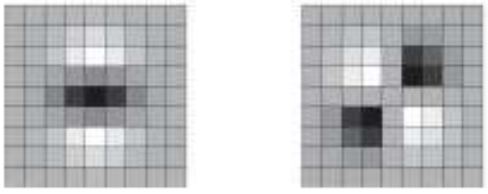


Figure 6.1 – Dérivées partielles discrétisées $L_{yy}(X, \sigma)$ et $L_{xy}(X, \sigma)$

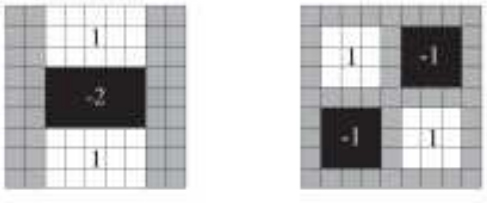


Figure 6.2 – Approximation de $L_{yy}(X, \sigma)$ et $L_{xy}(X, \sigma)$ par des filtres carrés

La seconde étape consiste à construire un descripteur pour le voisinage local de chaque point d'intérêt. Un descripteur de 64 éléments est déterminé en utilisant une grille de localisation spatiale correspondant à un histogramme local des ondelettes de Haar.

La troisième étape permet une correspondance entre les descripteurs des deux images à comparer en utilisant une mise en correspondance exhaustive.

Pour construire notre estimateur de la pose de la tête, nous n'utilisons pas la version classique du SURF, mais nous y apportons les modifications suivantes afin de l'adapter à notre problème. Après l'extraction des descripteurs des points d'intérêt, nous les trions selon leur orientation. Ensuite, nous fixons le nombre final de descripteurs N que nous désirons obtenir et nous divisons les descripteurs triés en N groupes. Chaque caractéristique SURF que nous utilisons pour la formulation de notre problème est composée de 64 éléments. Elle correspond à la moyenne des descripteurs d'un groupe. Ainsi, il est nécessaire de choisir judicieusement le nombre N afin de conserver un bon compromis entre la performance et le temps de calcul.

6.2.2 Sélection des variables

Nous avons choisi de représenter des caractéristiques aussi diverses et riches que possible afin de prendre avantage de leur complémentarité, mais nous n'ignorons pas la possibilité de redondance de l'information. Le but de l'étape de sélection des variables est de déterminer un

ensemble d'attributs qui soit compacte, significatif et consistant afin de faciliter la classification. En utilisant les données d'apprentissage, les techniques de sélection des variables recherchent le sous-ensemble qui permet la meilleure prédiction parmi toutes les combinaisons possibles des attributs. Ainsi, la sélection des variables consiste à effectuer deux tâches en appliquant :

- Une méthode de recherche qui génère les sous-ensembles des variables et tente de trouver un sous-ensemble optimal. Nous définissons dans ce qui suit quelques méthodes de recherche.
- Une méthode d'évaluation des attributs qui détermine si un sous-ensemble est optimal et retourne quelques mesures d'efficacité à la méthode de recherche. Nous présentons dans ce qui suit les techniques d'évaluation des attributs les plus connues.

6.2.2.1 Méthodes de recherche

- **La méthode BestFirst (BF)** débute par un ensemble vide de variables et génère tous les singletons possibles. Le sous-ensemble possédant la plus grande évaluation est choisi et est étendu de la même façon par l'ajout de tous les singletons possibles. Si l'extension d'un sous-ensemble ne produit pas d'amélioration, la recherche continue du second meilleur sous-ensemble non étendu. L'algorithme retourne un ensemble de variables ordonnées selon leur pertinence et l'utilisateur choisi le nombre d'éléments qui lui convient.
- **La méthode GreedyStepwise (GS)** effectue une recherche gloutonne en avant ou en arrière parmi l'espace des sous-ensembles des variables. L'algorithme peut débuter par aucune/toutes les variables et s'arrête quand l'ajout/suppression d'un attribut ne produit pas d'amélioration. L'algorithme peut aussi produire une liste ordonnée des variables en traversant tout l'espace de recherche et en conservant l'ordre de sélection des variables.
- **La méthode Ranker (Rk)** ordonne les variables selon leur score d'évaluation individuel fourni par la méthode d'évaluation des attributs choisis. L'utilisateur doit spécifier le nombre de variables dont il a besoin.

6.2.2.2 Méthodes d'évaluation des attributs

- **La méthode CorrelationFeatureSelection (CFS)** évalue la performance d'un sous-ensemble d'attributs en considérant la capacité de prédiction individuelle de chaque élément ainsi que le degré de redondance entre ces éléments. Les sous-ensembles fortement corrélés à la classe et disposant d'une faible inter-corrélation possèdent les meilleurs évaluations. Cet évaluateur peut être associé aux méthodes de recherche BF ou GS.
- **La méthode GainRatio (GR)** évalue la performance d'un attribut en mesurant le rapport de gain par rapport à la classe. Cet évaluateur est associé à la méthode de recherche Rk. L'évaluation des attributs par la méthode **InformationsGain (IG)** est similaire à celle proposée par GR, mais au lieu de mesurer le rapport du gain, elle mesure le gain d'information par rapport à la classe.
- **La méthode OneRule (OneR)** évalue la performance d'un attribut en utilisant le classifieur simple OneR qui génère une règle pour chaque variable. La méthode de Rk est associée à cet évaluateur.
- **La méthode ReliefF (RF)** évalue la performance d'un attribut selon sa capacité à

distinguer entre les instances voisines. Cet évaluateur est associé à la méthode de recherche Rk.

Dans la section 6.4, nous évaluerons ces techniques de sélection des variables et la meilleure sera retenue pour la construction du vecteur caractéristique de la pose de la tête. Ces vecteurs serviront à l'apprentissage des SVM multi-classes afin d'estimer la pose de la tête du conducteur (voir la section 6.3).

6.3 Estimation de la pose de la tête du conducteur par des SVM multi-classes

Dans la sous-section 5.4.1, nous avons formulé le problème de l'estimation de la pose de la tête pour le conducteur et nous avons conclu que l'utilisation de trois poses pour le pitch (frontale, haute et basse) et trois poses pour le yaw (frontale, gauche et droite) est suffisante pour déterminer le niveau d'attention. Dans la sous-section 5.4.2, nous avons effectué l'apprentissage de deux fonctions de vraisemblance paramétriques l'une pour le pitch et l'autre pour le yaw, puisque nous supposons que ces deux angles sont indépendants. Grâce à l'indépendance des deux angles, nous pouvons effectuer l'apprentissage de deux SVM multi-classes, que nous notons pitch-SVM et yaw-SVM, chacun dédié à la classification de trois poses.

Le principe des SVM a été présenté dans la sous-section 2.2.1. Nous rappelons que le SVM binaire permet d'optimiser un hyperplan séparateur entre les exemples d'apprentissage positifs et négatifs. Pour les SVM multi-classes, le problème original doit être décomposé en séries de problèmes d'apprentissage binaires. Une solution standard est proposée par l'approche « tous contre un » (one-against-all) qui consiste à construire un classifieur binaire pour chaque classe. Une approche plus rapide et plus robuste pour un nombre réduit de classes est proposée par la classification par pair (pairwise classification). La classification par pair transforme un problème à c classes en $\frac{c(c-1)}{2}$ problèmes binaires, chacun dédié à la classification d'une paire de classes. Ainsi, en utilisant cette approche, chacun de nos deux problèmes à trois classes est décomposé en trois sous-problèmes binaires.

Après l'apprentissage des deux SVM multi-classes en utilisant les vecteurs caractéristiques définis par la section 6.2, nous effectuons les étapes suivantes lors de la présentation d'une frame du conducteur à notre estimateur de la pose de la tête :

- Localiser le patch de la tête par une technique de segmentation de l'image en pixels peau et non peau.
- Construire le vecteur caractéristique à partir de la fusion des descripteurs et la sélection des variables pertinentes (voir la section 6.2). Les paramètres des descripteurs et la technique de sélection des variables choisie seront déterminés par l'étude présentée dans la section 6.4.
- Estimer la pose de la tête par le classifieur pitch-SVM. Si la tête est baissée ou levée, nous observons la durée pour laquelle une seule position est fixée et nous émettons une alarme d'inattention quand cette durée est importante.
- Si la pose selon le pitch est frontale, nous estimons la pose de la tête en utilisant le classifieur yaw-SVM. Si la tête est tournée à gauche ou à droite, nous observons la durée pour laquelle une seule position est fixée et nous émettons une alarme d'inattention quand

cette durée est importante.

6.4 Résultats expérimentaux

Comme nous l'avons précisé dans la section 5.5 de notre approche SP-LPF, nous avons besoin d'utiliser la base de données Pointing'04 pour déterminer les paramètres optimaux du système actuel (voir la sous-section 6.4.1) et comparer avec les techniques présentées dans la littérature (voir la sous-section 6.4.2). Pour les tests effectués sur la base de données Pointing'04, nous utilisons les mêmes ensembles que ceux définis pour l'approche précédente, à savoir 80% de la base pour l'apprentissage (2232 images) et 20% pour le test (558 images). Ensuite, dans la sous-section 6.4.3, nous testons l'approche actuelle, notée Descriptors Fusion-SVM (DF-SVM), sur la même séquence vidéo utilisée pour valider SP-LPF (voir la sous-section 5.5.4).

6.4.1 Optimisation des paramètres

Puisque le nombre de classes influence considérablement les résultats des SVM, nous avons décidé d'optimiser les paramètres de DF-SVM en considérant la base de données Pointing'04 avec la représentation adaptée à l'estimation de la pose du conducteur, à savoir trois poses pour le pitch et trois poses pour le yaw. Les résultats seront présentés en termes de CCR, de coefficient κ et de temps d'exécution moyen pour le traitement d'un degré de liberté d'une frame en ms, noté TE. Nous définissons les classes adaptées à l'estimation de la pose de la tête du conducteur à partir de la base de données Pointing'04 comme suit :

- Pitch :
  - tête baissée : $\{-90; -60; -30\}$
  - tête frontale : $\{-15; 0; +15\}$
  - tête haute : $\{+30; +60; +90\}$
- Yaw :
  - profil gauche : $\{-90; -75; -60; -45; -30\}$
  - tête frontale : $\{-15; 0; +15\}$
  - profil droit : $\{+30; +45; +60; +75; +90\}$

- **Optimisation des paramètres des descripteurs**

Nous avons effectué plusieurs séries de tests pour déterminer les paramètres optimaux pour chaque descripteur en utilisant la procédure définie dans la section 6.3 et nous avons obtenu les meilleurs résultats par les valeurs suivantes :

- SF : [taille_SF = 450 $(15 \times 15 \times 2)$]
  - Nombre de filtres = 2
  - Patch de l'image = 15×15
  - Pas d'orientation = $50°$
- HOG : [taille_HOG = 90 $(3 \times 3 \times 10)$]
  - Nombre de cellule par ligne = 3
  - Nombre de cellule par colonne = 3

  - Nombre de bins = 10
- SURF : [taille_SURF = 256 (64×4)]
  - Dimension du descripteur= 64
  - Nombre de descripteurs = 4
- Haar : [taille_Haar = 1024 (32×32)]
  - Nombre d'ondelettes = 32

Nous avons aussi testé plusieurs noyaux pour les SVM et nous avons choisi le noyau RBF donné par l'équation 6.4 avec le paramètre $\Gamma = 0.15$. Le terme $\|x - y\|_2^2$ correspond à la distance euclidienne carrée.

$$K(x, y) = e^{-(\Gamma * \|x-y\|_2^2)} \tag{6.4}$$

Tableau 6.1 – Évaluation des descripteurs sans utiliser la sélection des variables

	3-classes pitch-SVM			**3-classes yaw-SVM**		
Descripteur	**CCR**	κ	**TE**	**CCR**	κ	**TE**
SF	**87.2**	**0.80**	**42**	**94.3**	**0.91**	**30**
HOG	85.6	0.77	10	94	0.90	10
SURF	83.8	0.75	40	93.7	0.90	40
Haar	85.9	0.78	200	93	0.89	200
(SF,HOG)	**89.3**	**0.8**	**70**	**95.5**	**0.93**	**60**
(SF,SURF)	89.0	0.83	130	95.3	0.92	120
(SF,Haar)	86.7	0.79	500	94.7	0.91	470
(HOG,Haar)	88.9	0.82	240	94.5	0.91	210
(HOG,SURF)	87.2	0.80	60	95	0.92	60
(SURF,Haar)	87.3	0.80	350	94.6	0.91	320
(SF,HOG,SURF)	**89.1**	**0.83**	**150**	**95.5**	**0.93**	**110**
(SF,HOG,Haar)	85.6	0.77	280	95.1	0.92	290
(SF,SURF,Haar)	77.9	0.64	530	92.3	0.88	480
(HOG,SURF,Haar)	87.8	0.81	190	95	0.92	170
(SF,HOG,SURF,Haar)	**87.5**	**0.80**	**530**	**94.9**	**0.91**	**520**

Dans le tableau 6.1, nous présentons dans un premier temps les résultats des descripteurs évalués séparément, ensuite leurs combinaisons par paire et par trio et enfin l'association des quatre descripteurs sans sélection de variables. Nous remarquons que les SF fournissent le meilleur résultat lors de l'évaluation individuelle des descripteurs. La meilleure combinaison par pair et par trio correspondent respectivement aux vecteurs caractéristiques (SF,HOG) et (SF,HOG,SURF). En ce qui concerne les temps de traitement d'une frame (TE), il est évident qu'il augmente en fonction du nombre de descripteurs utilisés. Toutefois, les caractéristiques de Haar sont les plus couteuses en termes de temps de calcul à cause de la taille importante de leur vecteur caractéristique. Quand nous combinons les quatre descripteurs, les résultats sont moins avantageux que ceux de la meilleure combinaison par paire ou par trio. Ceci peut être expliqué par une interaction entre les attributs du vecteur caractéristique global, qui produit des contradictions au niveau de la prise de décision par les SVM multi-classes. Ce problème peut être résolu par l'étape de sélection des variables qui nous permettra de conserver les attributs les plus pertinents et de réduire ainsi le temps de traitement.

- **Optimisation des paramètres de la sélection des variables**

Tableau 6.2 – Performance de DF-SVM en variant les techniques de sélection des variables

		3-classes pitch-SVM			**3-classes yaw-SVM**		
Descripteur	**Selec. var.**	**CCR**	κ	**TE**	**CCR**	κ	**TE**
(SF,HOG,SURF)	(CFS,BF,$^{400}/_{796}$)	89.2	0.83	80	95.5	0.93	70
(SF,HOG,SURF)	(GR,Rk,$^{400}/_{796}$)	87.0	0.79	60	94.5	0.91	50
(SF,HOG,SURF)	(IG,Rk,$^{400}/_{796}$)	86.5	0.79	50	94.5	0.91	50
(SF,HOG,SURF)	(OneR,Rk,$^{400}/_{796}$)	86.4	0.79	40	94.6	0.91	40
(SF,HOG,SURF)	(RF,Rk,$^{400}/_{796}$)	90.1	0.84	40	95.4	0.93	30
(SF,HOG,SURF,Haar)	(RF,Rk,$^{600}/_{1820}$)	90.5	0.85	220	96.7	0.94	210
(SF,HOG,SURF,Haar)	(RF,Rk,$^{400}/_{1820}$)	90.5	0.85	88	96.6	0.94	80
(SF,HOG,SURF,Haar)	(RF,Rk,$^{200}/_{1820}$)	88.1	0.80	58	94.2	0.91	53
(SF,HOG,SURF,Haar)	(RF,Rk,$^{400}/_{1820}$)	91.9	0.87	CV	96.4	0.94	CV

Dans le tableau 6.2, nous évaluons les techniques de sélection des variables présentées dans la sous-section 6.2.2 sur les combinaisons (SF,HOG,SURF) et (SF,HOG,SURF,Haar), qui correspondent à la meilleure combinaison par trio et à tous les descripteurs. En premier lieu, nous présentons les résultats de l'application des techniques de sélection de variables (Selec. var.) sur la combinaison (SF,HOG,SURF), en choisissant les 400 variables les plus pertinentes sur un total de 796, ce qui correspond à environ la moitié des attributs. D'après le tableau, le meilleur résultat de ce test est donnée lorsque nous appliquons la méthode d'évaluation des attributs RF associée à la méthode de recherche Rk. En second lieu, nous appliquons la méthode RF associée à la méthode Rk sur la combinaison de tous les descripteurs en variant le nombre de variables pertinentes sélectionnées. La variation de ce paramètre nous permettra de déterminer le nombre qui fournit un bon compromis entre le temps de traitement (TE), le CCR et le coefficient κ. Nous choisissons d'utiliser 400 variables puisque ce nombre de caractéristiques pertinentes permet de fournir un bon compromis entre l'efficacité et le temps d'exécution (80 ms au lieu de 430 ms pour 600 variables). Dans la dernière colonne du tableau contenant la mention « CV », nous affichons le résultat de la sélection de 400 variables pertinentes à partir de la combinaison de tous les descripteurs en utilisant la validation croisée « Cross Validation » (CV) k-fold avec $k = 10$. La validation croisée réordonne la base de données et la divise en 10 parties égales. Pour chaque itération, une partie est utilisée pour le test et les 9 autres pour l'apprentissage du classifieur. Tous les résultats sont collectés et moyennés à la fin de la validation croisée. L'utilisation de cette procédure de validation permet d'estimer la fiabilité du système en variant les échantillons. Nous remarquons que le résultat obtenu par la validation croisée améliore le test classique, ce qui prouve que l'approche DF-SVM permet une bonne classification des poses même en variant les échantillons. Par la suite, nous nommons « vecteur caractéristique final », le vecteur composé de la combinaison des quatre descripteurs à laquelle nous appliquons la sélection de 400 variables par la méthode RF associée à la méthode Rk.

Afin de visualiser les descripteurs les plus pertinents dans le vecteur caractéristique final, nous présentons les éléments suivants dans le tableau 6.3 pour les angles pitch et yaw :

– Nb_Att : nombre total des attributs du descripteur

– Nb_Slc : nombre des attributs sélectionnés à partir du descripteur
– Pct_Slc : pourcentage des attributs sélectionnés (Nb_Slc) par rapport à tous les éléments du descripteur
– Pct_Dsc : pourcentage des attributs sélectionnés du descripteur par rapport à tous les éléments du vecteur caractéristique final

Tableau 6.3 – Visualisation de la participation de chaque descripteur dans le vecteur caractéristique final

		3-classes pitch-SVM			**3-classes yaw-SVM**		
Descripteur	**Nb_Att**	**Nb_Slc**	**Pct_Slc**	**Pct_Dsc**	**Nb_Slc**	**Pct_Slc**	**Pct_Dsc**
SF	450	263	58%	66%	275	61%	69%
HOG	90	62	68%	16%	70	78%	18%
SURF	256	0	0%	0%	11	4%	2%
Haar	1024	75	7%	18%	44	4%	11%

Si nous analysons la colonne Pct_Slc, nous remarquons que plus de 50% des attributs des SF et des HOG sont sélectionnés alors que moins de 10% des attributs Haar et SURF sont choisis. De plus, l'analyse de la colonne Pct_Dsc permet de déduire que les attributs des SF sont les plus présents dans le vecteur caractéristique final avec un pourcentage supérieur à 65%, alors que les attributs de SURF sont inexistants pour le vecteur caractéristique du pitch et ne dépassent pas les 2% pour le vecteur du yaw.

6.4.2 Comparaison

Tableau 6.4 – Comparaison de l'approche DF-SVM avec la littérature en termes de pitch-MAE et yaw-MAE

Ligne	**Approche**	**Pitch-MAE**	**Yaw-MAE**
1	DF-SVM	**4.6**°	**6.1**°
2	SP-LPF	10.86°	7.52°
3	SIFT + SVM + SVR (Ho et Chellappa, 2012)	5.84°	6.05°
4	PLS noyau RBF (Al-Haj *et al.*, 2012)	6.61°	6.56°
5	SP + SVM (Jain et Crowley, 2013)	8°	6.9°
6	SVM + SVR (LARR) (Guo *et al.*, 2008)	7.69°	9.23°
7	Performance humaine (Gourier *et al.*, 2007)	11°	11.9°
8	LAAM (Gourier *et al.*, 2007)	15.9°	10.3°

La majorité des approches testées sur la base de données Pointing'04 adoptent sa représentation standard des poses, à savoir 9 poses pour le pitch et 13 poses pour le yaw. Ainsi, nous considérons cette représentation pour l'estimation de la pose de la tête par DF-SVM afin d'effectuer une comparaison équitable. De plus, nous présentons les résultats dans cette sous-section en termes de MAE, puisqu'elle correspond à la mesure la plus utilisée dans la littérature.

Dans le tableau 6.4, nous reprenons les résultats des techniques présentées dans le tableau 5.5. Contrairement à l'approche SP-LPF, nous pouvons déduire de ce tableau que l'approche DF-SVM présentée dans la première ligne dépasse tous les autres estimateurs présentés dans la

littérature.

Le temps de traitement d'une image de la base de données Pointing'04 n'est fourni que par Jain et Crowley (2013). Ce temps est estimé à 108 ms alors que notre approche permet un traitement en 88 ms avec beaucoup plus de précision, pour un seul angle de liberté.

6.4.3 Estimation de la pose de la tête appliquée à la séquence du conducteur

Dans cette sous-section, nous validons l'approche DF-SVM sur la même séquence vidéo que nous avons acquise pour valider l'approche SP-LPF (voir la sous-section 5.5.4). Nous rappelons que cette séquence est composée de 1416 frames (946 pour l'apprentissage et 470 pour le test). Puisque nous étudions la pose de la tête du conducteur, nous avons annotée cette séquences par trois classes pour le pitch et trois classes pour le yaw. Nous présentons dans le tableau 6.5 et le tableau 6.6 la matrice de confusion pour le pitch et le yaw respectivement.

Tableau 6.5 – Matrice de confusion de l'estimation de la pose de la tête selon le pitch en utilisant l'approche DF-SVM pour la séquence du conducteur

Réelle/Estimée	**Frontale**	**Haute**	**Basse**
Frontale	341	3	2
Haute	2	47	1
Basse	3	1	70

Tableau 6.6 – Matrice de confusion de l'estimation de la pose de la tête selon le yaw en utilisant l'approche DF-SVM pour la séquence du conducteur

Réelle/Estimée	**Frontale**	**Gauche**	**Droite**
Frontale	289	2	1
Gauche	2	97	1
Droite	0	3	76

À partir de ces deux tableaux et des formules présentées dans la sous-section 3.6.1, nous obtenons :

- Pitch : $CCR = 0.97$; $\kappa = 0.93$
- Yaw : $CCR = 0.98$; $\kappa = 0.96$

Ces résultats sont très satisfaisants comparés aux résultats fournis par le tableau 5.6 et le tableau 5.7 et prouvent que l'approche DF-SVM est plus adaptée à l'estimation de la pose de la tête du conducteur que l'approche SP-LPF.

6.5 Conclusion

Dans ce chapitre, nous avons défini un estimateur de la pose de la tête basé sur une combinaison de quatre descripteurs (SF,HOG,SURF,Haar) à laquelle nous avons appliqué une technique de sélection des variables pertinentes. Ensuite, nous avons effectué l'apprentissage de deux classifieurs SVM multi-classes, chacun dédié à estimer la pose selon un degré de liberté précis (pitch ou yaw). Nous avons réalisé des séries de tests sur la base de données publique Pointing'04 pour

déterminer les paramètres optimaux de chaque descripteur, des SVM et des techniques de sélection des variables. Par la suite, nous avons comparé notre approche avec des travaux existants utilisant la même base de données. DF-SVM dépasse toutes les approches de la littérature et aussi l'approche précédente SP-LPF. De plus, le test effectué sur la séquence vidéo du conducteur a fourni un CCR supérieur à 96% et un coefficient κ supérieur à 0.93. Ces résultats prouvent que DF-SVM est très robuste pour l'estimation de la pose de la tête du conducteur même quand l'environnement de l'acquisition n'est pas contrôlé. Toutefois, nous tenons à préciser que l'approche SP-LPF possède un avantage qui s'illustre par sa rapidité de traitement. En effet, pour un seul angle de liberté, une même image est traitée par SP-LPF en 70 ms alors que DF-SVM a besoin de 88 ms, et que la technique proposée par Jain et Crowley (2013) nécessite 108 ms.

Conclusion et perspectives

Conclusion

L'hypovigilance chez le conducteur est la cause principale des accidents sur la route. Elle engendre plusieurs dégâts matériels et humains chaque année et partout dans le monde. Un grand nombre de ces accidents peuvent être évités si le conducteur est averti de sa baisse de vigilance. C'est dans cette optique que les systèmes de surveillance de l'état du conducteur ont été proposé. Nous avons vu, dans le chapitre 1.5, qu'il existe divers types de systèmes dédiés à la détection de l'hypovigilance, selon le type du signal analysé. Nous avons pu relever des systèmes basés sur les signaux physiologiques (exemple : activité cérébrale), puis ceux analysant le comportement du véhicule (exemple : la vitesse) et enfin ceux basés sur l'étude des signaux physiques (exemple : les yeux). Nous avons pu constater à partir de ce même chapitre, que plusieurs systèmes basés sur le comportement du véhicule sont déjà commercialisés sur des marques de haut de gamme. Cependant, l'inconvénient de ce genre de systèmes est leur dépendance au type du véhicule et aux conditions de la route. Il est primordial d'élargir l'intégration des systèmes de surveillance de l'état du conducteur dans l'industrie automobile et de les rendre accessibles à tous. L'étude des signaux physiologiques reste très intrusive par le fait qu'elle nécessite la mise en place de capteurs directement liés au corps du conducteur. Ainsi, il est plus judicieux de s'intéresser à l'étude des signaux physiques, qui se base sur l'analyse des changements affectant les caractéristiques faciales et ne nécessite que des caméras pour capter les informations sur le visage du conducteur.

L'objectif de cette thèse était de proposer des techniques basées sur le traitement de la vidéo du conducteur afin de déterminer son niveau de vigilance. Nous avons donc défini trois niveaux d'hypovigilance, à savoir la somnolence, la fatigue et l'inattention (voir chapitre 1). La somnolence est l'état le plus critique qui correspond à une incapacité à se maintenir éveillé, suivi par la fatigue qui est une baisse progressive des performance et l'inattention qui peut être considérée comme une distraction qui nous détourne de l'activité de conduite. Après une étude bibliographique approfondie, nous avons pu distinguer entre deux types d'approches pour la surveillance de l'état du conducteur selon les informations du visage à analyser. Ainsi, nous avons divisé notre travail en deux parties. La partie Partie I permet d'analyser les caractéristiques faciales à l'intérieur du visage, qui correspondent aux yeux et à la bouche afin d'étudier la somnolence et la fatigue. Dans la partie Partie II, l'estimation de la pose de la tête est effectuée pour détecter l'inattention.

Dans la partie Partie I, le chapitre 2 a été consacré à la présentation d'un état de l'art sur les

différentes mesures de l'état de l'œil et de la bouche présentes dans la littérature. Ensuite, nous avons proposé dans le chapitre 3 une approche pour la détection de la somnolence à partir de la détermination des périodes de micro-sommeil. Nous avons utilisé la CHT puisqu'elle permet de détecter les cercles présents dans une image afin de relever la présence de l'iris et d'en déduire l'ouverture/fermeture de l'œil. Du fait que la CHT opère sur des images du contour, nous avons proposé un détecteur de contour original adapté à la morphologie de l'œil pour permettre une meilleure performance de la CHT. Dans ce même chapitre, nous avons présenté une approche dédiée à la détection de la fatigue à partir du bâillement en utilisant le même concept que celui proposé pour la détection de la somnolence. En effet, puisque le bâillement correspond à une grande ouverture de la bouche qui dure plus de deux secondes, nous appliquons la CHT à un détecteur de contour original que nous avons construit pour relever la grande ouverture de la bouche. Les techniques que nous avons proposées dans ce chapitre sont simples mais néanmoins efficaces. En effet, nous avons effectué plusieurs tests sur 18 séquences vidéos réelles que nous avons acquises (voir sous-section 3.6.2), puisqu'aucune base de données n'est disponible pour tester les approches conçues pour la détection des micro-sommeils et du bâillement. Ces tests ont révélé des CCR moyens supérieures à 97% et un coefficient κ moyen supérieur à 0.94 pour l'analyse de l'état de l'œil et de la bouche, lorsque les séquences sont acquises sous la lumière ambiante du jour. Quand l'acquisition se fait sous un éclairage artificiel pendant la nuit, ces valeurs sont réduites à 88% et à 0.75 respectivement, mais restent tout de même acceptables. Il est toutefois nécessaire de préciser qu'il est impossible d'étudier ces deux caractéristiques si le visage n'est pas frontal à la caméra. Afin de surveiller l'état du conducteur même dans les cas de non visibilité des caractéristiques faciales, il est primordial d'estimer la pose de la tête.

Dans la partie Partie II, le chapitre 4 a été élaboré pour présenter un état de l'art sur les techniques de l'estimation de la pose de la tête en général, puis celles dédiées à détection de l'inattention chez le conducteur. Nous avons retenu de cette étude que les transformations de l'image basées sur les orientations sont les plus adaptées pour déterminer les caractéristiques spécifiques à la pose. Nous avons donc proposé deux estimateurs de la pose de la tête pouvant être appliqués pour détecter l'inattention chez le conducteur. Le premier estimateur (voir chapitre 5) est basé sur les templates d'apparence construits par une transformation multi-échelle et multi-orientation par SP avec comme technique de mise en correspondance, un apprentissage probabiliste par la LPF. La LPF a été utilisée précédemment pour le suivi d'objet, notamment la tête (Toyama et Blake, 2002; Ricci et Odobez, 2009), mais n'a jamais été exploitée pour l'estimation de la pose de la tête. Nous avons nommé ce premier estimateur SP-LPF. Le second estimateur (voir chapitre 6), nommé DF-SVM, est basé sur la classification et la fusion inédite de plusieurs descripteurs de l'image permettant de relever les orientations de la pose de la tête. Nous utilisons une classification par SVM multi-classes et une fusion de quatre descripteurs. Puisqu'il n'existe aussi aucune base de données pour la détection de l'inattention du conducteur, nous avons choisi d'effectuer des séries de tests sur la base publique Pointing'04 pour déterminer les paramètres optimaux des deux approches ainsi que pour les comparer avec des travaux existants utilisant la même base de données. Nous avons acquis une séquence de test pour valider ces approches pour l'estimation de la pose du conducteur et nous avons obtenus des CCR supérieurs à 86% et 97% et un coefficient κ supérieur à 0.93 et 0.72 pour SP-LPF et DF-SVM respectivement.

Toutefois, nous tenons à préciser que SP-LPF possède un avantage qui s'illustre par sa rapidité de traitement. En effet, pour un seul angle de liberté, une même image est traitée par cette approche en 70 ms alors que l'approche DF-SVM a besoin de 88 ms.

Perspectives

Après avoir analysé les méthodes proposées dans ce travail de thèse, nous présentons les perspectives envisageables de nos travaux de recherche. Les pistes à explorer et les applications sont nombreuses :

- Acquérir et d'annoter une base de données englobant les différents états du conducteur. Par la suite, nous pourrons tester l'ensemble du système sur cette base de données. L'acquisition d'une base de données que nous rendrons publique est très important, puisque cela permettra de fournir à la communauté de recherche des données prêtes à l'emploi et propices à de futures comparaisons entre différents systèmes.
- Implémenter le travail proposé sur un système embarqué afin de réduire le temps de calcul et avoir une meilleure visualisation de son utilisation réelle. Les équipements que nous utiliserons devront être à très faibles coûts pour garantir une accessibilité du système à tous.
- Intégrer un système d'éclairage infrarouge qui ne sera activé que pendant la nuit afin d'améliorer l'acquisition de la scène. Il est à noter que ce changement entrainera la nécessité d'adapter notre système à ce nouveau type d'éclairage.
- Enrichir le système en incluant une analyse de la direction du regard du conducteur pour contrôler plus précisément son centre d'intérêt et le comparer avec les obstacles présents sur la route.
- Lorsque nous conduisons en état d'hypovigilance, nous présentons un danger pour soit mais aussi pour les autres conducteurs et usagers de la route. Nous pourrons envisager d'exploiter les réseaux de capteurs sans fil pour tenir informé les conducteurs de l'état de vigilance de leurs voisins.

Pour conclure, il semble que le domaine de surveillance de l'état du conducteur par des caméras reste toujours actif malgré le progrès technique et la recherche avancée des systèmes de vison par ordinateur.

Liste des publications

Revues internationales

Nawal Alioua, Aouatif Amine and Mohammed Rziza, « Driver's Fatigue Detection Based on Yawning Extraction », ***International Journal of Vehicular Technology (IJVT)***, vol. 2014, Article ID 678786, 2014, Hindawi Publishing Corporation.

Nawal Alioua, Aouatif Amine, Mohammed Rziza, Abdelaziz Bensrhair, « Estimating driver head pose using steerable pyramid and probabilistic learning », ***International Journal of Computer Vision and Robotics (IJCVR)***, in press, Inderscience Publishers, ISSN online : 1752-914X.

Nawal alioua, Aouatif Amine, Alexandrina Rogozan, Abdelaziz Bensrhair, Mohammed Rziza, « Driver head pose estimation using efficient descriptor fusion », submitted to *EURASIP Journal on Image and Video Processing*, Springer

Conférences internationales

Nawal Alioua, Aouatif Amine, Mohammed Rziza, Driss Aboutajdine, « Eye State Analysis Using Iris Detection to Extract Micro-Sleep Periods », ***International Conference on Computer Vision Theory and Applications (VISAPP'11)***, Vilamoura, Protugal, 2011.

Nawal Alioua, Aouatif Amine, Mohammed Rziza, Driss Aboutajdine, « Eye State Analysis using Iris Detection based on Circular Hough Transform », ***International Conference on Multimedia Computing and Systems, (ICMCS'11)***, Ouarzazate, Morocco, 2011.

Nawal Alioua, Aouatif Amine, Mohammed Rziza, Driss Aboutajdine, « Fast Micro-Sleep and Yawning Detections to Assess Driver's Vigilance Level », ***the 6th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design (DA'11)***, California, USA, 2011.

Nawal Alioua, Aouatif Amine, Mohammed Rziza, Driss Aboutajdine, « Driver's Fatigue and Drowsiness Detection to Reduce Traffic Accidents on Road », ***Computer Analysis of Images and Patterns (CAIP'11), Lecture Notes in Computer Science***, Volume 6855, pp 397-404, Sevilla, Spain, 2011.

Nawal Alioua, Aouatif Amine, Mohammed Rziza, Abdelaziz Bensrhair, Driss Aboutajdine, « Head pose estimation based on steerable filters and likelihood parametrized function », ***the 21st European signal Processing conference (EUSIPCO'13)***, Marrakech, Morocco, 2013.

Aouatif Amine, **Nawal Alioua**, Frédérick Zann, Yassine Ruichek, and Nabil Hmina, « Monitoring Drivers Drowsiness Using a Wide Angle Lens », ***The 16th IEEE International Conference on Intelligent Transportation Systems (ITSC'13)***, The Hague, The Netherlands, pp. 290-295, 2013.

Bibliographie

Aggarwal, G., Veeraraghavan, A. et Chellappa, R. (2005). 3D facial pose tracking in uncalibrated videos. *In International Conference on Pattern Recognition and Machine Intelligence (PReMi)*, pages 515–520, Kolkata, India.

Al-Haj, M., Gonzalez, J. et Davis, L. (2012). On partial least squares in head pose estimation : How to simultaneously deal with misalignment. *In Proceeding of Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2602–2609.

AntiSleep 4 (2013). Smart eye antisleep 4. `www.smarteye.se/sites/smarteye/files/datasheets/smarteye_antisleep.pdf` (dernier accès : février 2014).

ASFA (2010). La sécurité, asfa (association professionnelle autoroutes et ouvrages routiers). `www.autoroutes.fr/FCKeditor/UserFiles/File/B-la_securite.pdf` (dernier accès Février 2014).

AttentionAssit (2009). Mercedes-benz attention assist. `fr.euroncap.com/fr/rewards/mercedes_benz_attention_assist.aspx` (dernier accès juin 2014).

Bay, H., Ess, A., Tuytelaars, T. et Gool, L. V. (2008). SURF : Speeded up robust features. *Computer Vision and Image Understanding*, 110:346–359.

Benjelloun, M. (2013). Journée d'etude sous le thème : Somnolence au volant. `http://fr.slideshare.net/CNPAC/s1-p2-expos-benjelloun-somnolence-18-04-2013` (dernier accès Janvier 2015).

Benoit, A. et Caplier, A. (2005). Hypovigilence analysis : Open or closed eye or mouth ? blinking or yawning frequency ? *In IEEE Conference on Advanced Video and Signal Based Surveillance*, pages 207–212.

Bergasa, L., Nuevo, J., Sotelo, M., Barea, R. et Lopez, M. (2006). Real-time system for monitoring driver vigilance. *IEEE Transactions on Intelligent Transportation Systems*, 7(1):63–77.

Berka, C., Levendowski, D., Lumicao, M., Yau, A., Davis, G., Zivkovic, V., Olmstead, R., Tremoulet, P. et Craven, P. (2007). Eeg correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine*, 78:B231–B244.

Black, J., Gargesha, M., Kahol, K., Kuchi, P. et Pan-chanathan, S. (2002). A framework for performance evaluation of face recognition algorithms. *In Proceeding of SPIE 4862, Internet Multimedia Management Systems III*, pages 242–247.

BRETZNER, L. et KRANTZ, M. (2005). Towards low-cost systems for measuring visual cues of driver fatigue and inattention in automotive applications. *In Proceeding of the International Conference on Vehicular Electronics and Safety*, pages 161–164.

BURGES, C. (1996). Simplified support vector decision rules. *In International Conference on Machine Learning*, pages 71–77.

CARRS-Q (2011). State of the road, a fact sheet of carrs-q (centre of accident research and road safety-queensland). `www.carrsq.qut.edu.au/publications/corporate/hooning_fs.pdf` (dernier accès février 2014).

CASCIA, M., SCLAROFF, S. et ATHITSOS, V. (2004). Fast, reliable head tracking under varying illumination : An approach based on registration of texture-mapped 3d models. *IEEE Transactions on on Pattern Analysis and Machine Intelligence*, 22:322–336.

CASTLEMAN, K., SCHULZE, M. et WU, Q. (1998). Simplified design of steerable pyramid filters. *In IEEE International Symposium on Circuits and System (ISCAS)*, pages 329–332.

CHAUMET, G. et PHILIP, P. (2007). Somnolence, risque d'accident de la circulation, et aspects médico-légaux : Troubles du sommeil. *La revue du praticien*, 57:1559–1560.

CHOI, S. et KIM, D. (2009). Robust head tracking using 3D ellipsoidal head model in particle filter. *Pattern Recognition*, 41:2901–2915.

CHUI, C. (1992). *An Introduction to Wavelets*. Academic Press Professional, Inc., San Diego, CA, USA.

COOTES, T., EDWARDS, G. et TAYLOR, C. (2001). Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23:681–685.

CORTES, C. et VAPNIK, V. (1995). Support vector networks. *Machine Learning*, 20:273–297.

DAC (2008). Volvo driver alert control. `http://www.verdel.ch/volvo-securite.html#DAC` (dernier accès juin 2014).

DAHMANE, A., LARABI, S., DJERABA, C. et BILASCO, I. (2012). Learning symmetrical model for head pose estimation. *In International conference on pattern recognition (ICPR)*.

DALAL, N. et TRIGGS, B. (2005). Histograms of oriented gradients for human detection. *In Proceeding of the Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 886–893.

DAWSON, D. et REID, K. (1997). Fatigue, alcohol and performance impairment. *Nature*, 388:235–237.

DEMENTHON, D. et DAVIS, L. (1995). Model-based object pose in 25 lines of code. *International Journal of Computer Vision*, 15:123–41.

DEVI, M., CHOUDHARI, M. et BAJAJ, P. (2011). Driver drowsiness detection using skin color algorithm and circular hough transform. *In 4th International Conference on Emerging Trends in Engineering and Technology (ICETET)*, pages 129–134.

Dinges, D. et Grace, R. (1998). Perclos : a valid phychophysiological measure of alertness as assessed by psychomotor vigilance. *In US Department of Transportation Federal Highway Administration*.

D'Orazio, T., Leo, M. et Distante, A. (2004). Eye detection in face images vigilance system. *In IEEE Intelligent Vehicles Symposium*, pages 14–17.

Dornaika, F. et Ahlberg, J. (2004). Fast and reliable active appearance model search for 3d face tracking. *IEEE Transactions on Systems, Man and Cybernetics Part B : Cybernetics*, 34:1838–1853.

Doucet, A., Freitas, N. et Gordon, N. (2001). *Sequential Monte Carlo methods in practice*. Springer-Verlag, New York.

Driver's Mate (2009). Effective control transport driver's mate. `http://www.24hmontreal.canoe.ca/24hmontreal/actualites/archives/2009/09/20090928-100106.html` (dernier accès : juin 2014).

Duda, R. et Hart, P. (1972). Use of the hough transformation to detect lines and curves in picture. *Communications of the ACM*, 15:11–15.

Fan, X., Bao-Cai, Y. et Yan-Feng, S. (2007). Yawning detection for monitoring driver fatigue. *In International Conference on Machine Learning and Cybernetics*, pages 664–668.

Flores, M., Armingol, J. et de la Escalera, A. (2010). Real-time warning system for driver drowsiness detection using visual information. *Journal of Intelligent & Robotic Systems*, 59(2):103–125.

Freeman, W. et Adelson, E. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 891–906.

Friedrichs, F. et Yang, B. (2010). Camera-based drowsiness reference for driver state classification under real driving conditions. *In IEEE Intelligent Vehicles Symposium*.

Fujitsu (2013). Fujitsu sleepiness detection sensor. `jad.fujitsu.com/exhibit/ceatec/pdf/id_01e.pdf` (dernier accès juin 2014).

Gee, A. et Cipolla, R. (1994). Determining the gaze of faces in images. *Image and Vision Computing*, 12(10):639–647.

Golz, M., Sommer, D., Holzbrecher, M. et Schnupp, T. (2007). Detection and prediction of drivers' microsleep events. *In 14th international conference Road safety on four continents*.

Gourier, N., Hall, D. et Crowley, J. (2004). Estimating face orientation from robust detection of salient facial features. *In International Workshop on Visual Observation of Deictic Gestures (Pointing)*.

Gourier, N., Maisonnasse, J., Hall, D. et Crowley, J. L. (2007). Head pose estimation on low resolution images. *In Multimodal Technologies for Perception of Humans*, volume 4122 de *Lecture Notes in Computer Science*, pages 270–280.

GRACE, R. (2001). Drowsy driver monitor and warning system. *In International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design (DA)*.

GUO, G., FU, Y., DYER, C. et HUANG, T. (2008). Head pose estimation : Classification or regression ? *In Proceeding of the 19th International Conference on Pattern Recognition (ICPR)*, pages 1–4.

GURBUZ, S., OZTOP, E. et INOUE, N. (2012). Model free head pose estimation using stereovision. *Pattern Recognition*, 45:33–42.

HO, H. et CHELLAPPA, R. (2012). Automatic head pose estimation using randomly projected dense sift descriptors. *In Proceeding of the 19th International Conference on Image Processing (ICIP)*, pages 153–156.

HORNG, W., CHEN, C., CHANG, Y. et FAN, C. (2004). Driver fatigue detection based on eye tracking and dynamic template matching. *In IEEE International Conference on Networking, Sensing and Control*, pages 7–12.

HORPRASERT, T., YACOOB, Y. et DAVIS, L. (1996). Computing 3-d head orientation from a monocular image sequence. *In International Conference on Automatic Face and Gesture Recognition*, pages 242–247.

HUANG, K. et TRIVEDI, M. (2003). Driver head pose and view estimation with single omnidirectional video stream. *In 1st International Workshop on In-Vehicle Cognitive Computer Vision Systems, in conjunction with the 3rd International Conference on Computer Vision Systems*, pages 44–51, Graz, Austria.

JAIN, V. et CROWLEY, J. (2013). Head pose estimation using multi-scale gaussian derivatives. *In Proceeding of the 18th Scandinavian Conference on Image Analysis*.

JEPSON, A., FLEET, D. et EL-MARAGHI, T. (2003). Robust online appearance model for visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1296–1311.

JI, Q. et YANG, X. (2004). Real-time eye, gaze, and face pose tracking for monitoring driver vigilance. *Real-Time Imaging*, 8:357–377.

JI, Q., ZHU, Z. et P.LAN (2004). Real-time nonintrusive monitoring and prediction of driver fatigue. *IEEE Transactions on Vehicular Technology*, 53:1052–1068.

JO, J., LEE, S., JUNG, H., PARK, K. et KIM, J. (2011). Vision-based method for detecting driver drowsiness and distraction in driver monitoring system. *Optical Engineering*, 5.

KARAM, W., MOKBEL, C., GREIGE, H., PESQUET-POPESCU, B. et CHOLLET, G. (2004). Un système de détection de visage et d'extraction de paramètres basé sur les svm et des contraintes. *In 9èmes journées d'études et d'échanges COmpression et REprésentation des Signaux Audiovisuels (CORESA)*.

KHAN, M. et AADIL, F. (2012). Efficient car alarming system for fatigue detection during driving. *International Journal of Innovation, Management and Technology*, 3.

KIENZLE, W., BAKIR, G., FRANZ, M. et SCHOLKOPF, B. (2005). Face detection - efficient and rank deficient. *Advances in Neural Information Processing Systems*, 17:673–680.

Kumar, V. et Poggio, T. (2000). Learning-based approach to real time tracking and analysis of faces. *In IEEE International Conference on Automatic Face and Gesture Recognition*, pages 96–101.

Lal, S. et Craig, A. (2001). A critical review of the psychophysiology of driver fatigue. *Biological Psychology*, 55:173–194.

Lamond, N. et Dawson, D. (1999). Quantifying the performance impairment associated with fatigue. *Journal of Sleep Research*, 8:255–262.

Lee, K., J.Ho, Yang, M. et Kriegman, D. (2003). Video-based face recognition using probabilistic appearance manifolds. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages I–313–I–320.

Li, B. et Chellappa, R. (2001). Face verification through tracking facial features. *Journal of the Optical Society of America A*, 18:2969–2981.

Li, L., Werber, K., Calvillo, C., Dinh, K. D., Guarde, A. et Konig, A. (2012). Multi-sensor soft-computing system for driver drowsiness detection. *In Online conference on soft computing in industrial applications*, volume 223 de *Advances in Intelligent Systems and Computing*.

Liu, X., Lu, H. et Luo, H. (2009). A new representation method of head images for head pose estimation. *In Proceeding of the 16th International Conference on Image Processing (ICIP)*, pages 3585–3588, Cairo, Egypt.

Ma, B., Chai, X. et Wang, T. (2013). A novel feature descriptor based on biologically inspired feature for head pose estimation. *Neurocomputing*, 115:1–10.

Martin, S., Tawari, A., Murphy-Chutorian, E., Cheng, S. et Trivedi, M. (2012). On the design and evaluation of robust head pose for visual user interfaces : Algorithms, databases, and comparisons. *In International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI)*, pages 149–154.

Martins, P. et Batista, J. (2008). Monocular head pose estimation. *In International conference on Image Analysis and Recognition (ICIAR)*, pages 357–368, San Francisco, USA.

Miao, J., Yin, B., Wang, K., Shen, L. et Chen, X. (1999). A hierarchical multiscale and multiangle system for human face detection in a complex background using gravity-center template. *Pattern Recognition*, 32.

Miyaji, M., Kawanaka, H. et Oguri, K. (2009). Driver's cognitive distraction detection using physiological features by the adaboost. *In International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 1–6.

Mohanty, M., Mishra, A. et Routray, A. (2009). A non-rigid motion estimation algorithm for yawn detection in human drivers. *International Journal of Computational Vision and Robotics*, 1:89–109.

Momin, B. et Abhyankar, P. (2012). Current status and future research directions in monitoring vigilance of individual or mass audience in monotonous working environment. *International Journal on Soft Computing*, 3:45–53.

MORENCY, L., WHITEHILL, J. et MOVELLAN, J. (2010). Monocular head pose estimation using generalized adaptive view-based appearance model. *Image and Vision Computing*, 28:754–761.

MUNOZ-SALINAS, R., YEGUAS-BOLIVAR, E., SAFFIOTTI, A. et MEDINA-CARNICER, R. (2012). Multi-camera head pose estimation. *Machine Vision and Applications*, 23:479–490.

MURPHY-CHUTORIAN, E., DOSHI, A. et TRIVEDI, M. (2007). Head pose estimation for driver assistance systems : A robust algorithm and experimental evaluation. *In Intelligent Transportation Systems Conference (ITSC)*, pages 709–714.

MURPHY-CHUTORIAN, E. et TRIVEDI, M. (2008). Hyhope : Hybrid head orientation and position estimation for vision-based driver head tracking. *In Intelligent Vehicles Symposium (IV)*, pages 512–517, Eindhoven, The Netherlands.

MURPHY-CHUTORIAN, E. et TRIVEDI, M. (2009). Head pose estimation in computer vision : A survey. *IEEE Transactions Pattern Analysis and Machine Intelligence*, pages 607–626.

MURPHY-CHUTORIAN, E. et TRIVEDI, M. (2010). Head pose estimation and augmented reality tracking : An integrated system and evaluation for monitoring driver awareness. *IEEE Transactions on intelligent transportation systems*, 11:300–311.

NG, J. et GONG, S. (2002). Composite support vector machines for detection of faces across views and pose estimation. *Image and Vision Computing*, 20:359–368.

NHTSA (2010). Research on drowsy driving. `www.nhtsa.gov/Driving+Safety/Distracted+Driving+at+Distraction.gov/Research+on+Drowsy+Driving` (dernier accès : février 2014).

NOGUCHI, Y., NOPSUWANCHAI, R., OHSUGA, M. et KAMAKURA, Y. (2007). Classification of blink waveforms towards the assessment of driver's arousal level - an approach for hmm based classification from blinking video sequence. *In Symposium on Biological and Physiological Engineering.*

OMI, T., NAGAI, F. et KOMURA, T. (2008). Driver drowsiness detection focused on eyelid behaviour. *In the 34th Congress on Science and Technology of Thailand.*

OSUNA, E., FREUND, R. et GIROSI, F. (1997). Training support vector machines : an application to face detection. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 130–136.

OUYANG, T. et LU, H. (2010). Vigilance analysis basedon continuous wavelet transform of eeg signals. *In International Conference on Biomedical Engineering and Computer Science (ICBECS)*, pages 1–4.

PAPAGEORGIOU, C. et POGGIO, T. (2000). A trainable system for object detection. *International Journal of Computer Vision*, 38:15–33.

QING, W., BINGXI, S., BIN, X. et JUNJIE, Z. (2010). A perclos-based driver fatigue recognition application for smart vehicle space. *In Third International Symposium on Information Processing (ISIP)*, pages 437–441.

Regan, M. (2010). Distraction du conducteur : définition, mécanismes, effets et facteurs modérateurs. www.ipubli.inserm.fr/bitstream/handle/10608/220/?sequence=21 (dernier accès Mai 2014).

Ricci, E. et Odobez, J. (2009). Learning large margin likelihoods for realtime head pose tracking. *In International Conference on Image Processing (ICIP)*, pages 2593–2596.

Romdhani, S., Torr, P., Schölkopf, B. et Blake, A. (2001). Computationally efficient face detection. *Proceeding of the 8th International Conference on Computer Vision*.

Rongben, W., Lie, G., Bingliang, T. et Lisheng, J. (2004). Monitoring mouth movement for driver fatigue or distraction with one camera. *In Intelligent Transportation Systems Conference*, pages 314–319.

Saffiotti, A. et Broxvall, M. (2005). Peis ecologies : Ambient intelligence meets autonomous robotics. *In International Conference on Smart Objects and Ambient Intelligence*, pages 275–280.

Saradadevi, M. et Bajaj, P. (2008). Driver fatigue detection using mouth and yawning analysis. *International Journal of Computer Science and Network Security (IJCSNS)*, 6.

Schneiderman, H. et Kanade, T. (1998). Probabilistic modeling of local appearance and spatial relationships for object recognition. *In International Conference on Computer Vision and Pattern Recognition (CVPR)*.

Scholkopf, B. et Smola, A. (2001). *Learning With Kernels : Support Vector Machines, Regularization, Optimization and Beyond*. MIT Press.

Senaratne, R., Hardy, D., Vanderaa, B. et Halgamuge, S. (2007). Driver fatigue detection by fusing multiple cues. *In Advances in Neural Networks*, volume 4492 de *Lecture Notes in Computer Science*, pages 801–809.

Shen, K., Ong, C., Li, X., Hui, Z. et Wilder-Smith, E. (2007). A feature selection method for multilevel mental fatigue eeg classification. *IEEE Transactions on Biomedical Engineering*, 54.

Sherrah, J., Gong, S. et Ong, E. (2001). Face distributions in similarity space under varying head pose. *Image and Vision Computing*, 19:807–819.

Shi, L. et Lu, B. (2008). Dynamic clustering for vigilance analysis based on eeg. *In International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 54–57.

Shin, H., Jung, S., Kim, J. et Chung, W. Y. (2010). Real time car driver's condition monitoring system. *In Proceeding of IEEE Sensors*, pages 951–954.

Simoncelli, E., Freeman, W., Adelson, E. et Heeger, D. (1992). Shiftable multiscale transforms. *IEEE Trans Information Theory*, 38(2):587–607. Special Issue on Wavelets.

SmartEye 5 (2013). Smart eye pro 5.10. www.smarteye.se/productseye-trackers/smart-eye-pro-2-6-cameras (dernier accès : juin 2014).

SMITH, K., BA, S., ODOBEZ, J. et GATICA-PEREZ, D. (2008). Tracking the visual focus of attention for a varying number of wandering people. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 30(7):1212–1229.

STUTTS, J., WILKINS, J. et VAUGHN, B. (1999). Why do people have drowsy driving crashes, input from drivers who just did. `https://www.aaafoundation.org/sites/default/files/sleep.PDF` (Last access : February 2014).

TAWARI, A., MARTIN, S. et TRIVEDI, M. (2014). Continuous head movement estimator for driver assistance : Issues, algorithms, and on-road evaluations. *IEEE Transactions on Intelligent Transportation Systems*, 15:818–830.

TIAN, Z. et QIN, H. (2005). Real-time driver's eye state detection. *In IEEE International Conference on Vehicular Electronics and Safety*, pages 285–289.

TOYAMA, K. et BLAKE, A. (2002). Probabilistic tracking with exemplars in a metric space. *International Journal of Computer Vision*, pages 9–19.

TRIVEDI, M., GANDHI, T. et MCCALL, J. (2007). Looking-in and looking-out of a vehicle : Computer-vision-based enhanced vehicle safety. *IEEE Transactions on Intelligent Transportation Systems*, 8:108–120.

TRUTSCHEL, U., SIROIS, B., SOMMER, D., GOLZ, M. et EDWARDS, D. (2011). Perclos : An alertness measure of the past. *In Driving Assessment 2011 : 6th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design*, pages 172–179.

VALENTI, R. et GEVERS, T. (2009). Robustifying eye center localization by head pose cues. *In IEEE conference on Computer Vision and Pattern Recognition (CVPR)*.

VAPNIK, V. (1995). *The nature of statistical learning theory*. Springer-Verlag.

VAPNIK, V. et KOTZ, S. (1982). *Estimation of Dependences Based on Empirical Data*. Springer Series in Statistics.

VIOLA, P. et JONES, M. (2001). Rapid object detection using a boosted cascate of simple features. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages I–511–I–518.

Volvo (2013). European accident research and safety report, volvo trucks. `http://pnt.volvo.com/pntclient/loadAttachment.aspx?id=27116` (dernier accès : février 2014).

WANG, J. et SUNG, E. (2008). EM enhancement of 3D head pose estimated by point at infinity. *Image and Vision Computing*, 25(12):1864–1874.

WANG, T. et SHI, P. (2005). Yawning detection for determining driver drowsiness. *In International Workshop on VLSI Design and Video Technique*, pages 373–376.

WIERWILLE, W. (1994). Overview of research on driver drowsiness definition and driver drowsiness detection. *In Technical International Conference on Enhanced Safety of Drivers (ESV)*, pages 23–26.

WILLIAMSON, A. et FEYER, A. (2000). Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. *Occupational & Environmental Medicine*, 57:649–655.

Wu, B., Ai, H., Huang, C. et Lao, S. (2004). Fast rotation invariant multi-view face detection based on real adaboost. *In International Conference on Automatic Face and Gesture Recognition (FG)*, pages 79–84.

Yang, M.-H. et Ahuja, N. (1998). Detecting human faces in color images. *In International Conference on Image Processing (ICIP)*, volume 1, pages 127–130.

Zhang, G., Cheng, B., Feng, R. et Zhang, X. (2008). A real-time adaptive learning method for driver eye detection. *In DICTA Digital Image Computing : Techniques and Applications*, pages 300–304.