

Saillance de la signalisation verticale dans les images routières : étude de la faisabilité d'un outil de diagnostic

Ludovic Simon

▶ **To cite this version:**

Ludovic Simon. Saillance de la signalisation verticale dans les images routières : étude de la faisabilité d'un outil de diagnostic. Traitement du signal et de l'image [eess.SP]. Université Pierre et Marie Curie - Paris VI, 2009. Français. NNT : 2009PA066556 . tel-00814582

HAL Id: tel-00814582

https://theses.hal.science/tel-00814582

Submitted on 17 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

École Doctorale de
« Sciences Mécaniques, Acoustique, Electronique et Robotique de Paris (ED 391) »

These de Doctorat de
l'université Pierre et Marie Curie

Spécialité
Traitement d'Image et Vision par Ordinateur

Saillance de la signalisation verticale dans les images routières

Etude de la faisabilité d'un outil de diagnostic

Présentée par
M. Ludovic SIMON

Pour obtenir le grade de
DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE

Thèse dirigée par M. Jean DEVARS
co-dirigée par M. Jean-Philippe TAREL

Soutenue le 07 décembre 2009

Devant le jury composé de :

M. Jean Devars, Professeur, Université PARIS VI	Directeur de thèse
M. Jean-Philippe Tarel, Chargé de Recherche, LEPSiS	Co-Directeur de thèse
M. Dominique Barba, Professeur émérite, Université Nantes	Rapporteur
M. Roland Chapuis, Professeur, Université Clermont	Rapporteur
M. J.M. Christian Bastien, Maître de Conférences-HDR, Université P. Verlaine	Examinateur
M. Patrick Le Callet, Professeur, Université Nantes	Examinateur
M. Maurice Milgram, Professeur émérite, Université Paris VI	Examinateur
M. Lionel Prevost, Maître de Conférences-HDR, Université Paris VI	Examinateur

Thèse préparée au Laboratoire Exploitation Perception Simulateurs et Simulations (LEPSiS UMR INRETS-LCPC)

« Lorsqu'on décrit l'œil, on le compare souvent à un appareil photo, mais les aspects les plus intéressants de la perception sont justement ceux qui diffèrent radicalement de ceux de la caméra. ».

Richard-L. GREGORY

psycho-physiologiste britannique [1].

A ma famille,
A mes amis,
A ceux qui m'ont soutenu,
A la mémoire de ceux qui sont partis trop tôt,
A mamie.

Remerciements

Le travail accompli dans le cadre de cette thèse a été effectué au sein du Laboratoire Central des Ponts et Chaussées dans la Section Éclairage et Visibilité de la Division Exploitation Signalisation et Éclairage qui est devenue, par la suite, partie intégrante du Laboratoire Exploitation, Perception, Simulateurs et Simulations (LEPSiS), nouvelle unité mixte de recherche du LCPC et de l'Institut National de Recherche sur les Transports et leur Sécurité (INRETS).

Je remercie en premier lieu l'ensemble de l'équipe pour leurs accueils et pour le plaisir que j'ai eu à travailler avec eux durant ces trois années passées au LCPC. Je témoigne ici ma profonde gratitude à M. Michel Bry, directeur du LEPSiS pour sa confiance et son soutient de tout instant.

Je souhaite remercier M. Jean Devars, mon directeur de thèse, qui m'a permis d'entrer dans le monde de la recherche. J'exprime également ma profonde reconnaissance à Messieurs Dominique Barba et Roland Chapuis qui ont bien voulu accepter le travail de rapporteur de cette thèse. Pour avoir apporté leur caution scientifique à ce travail en acceptant de faire eux aussi partie du jury de thèse, je remercie aussi Messieurs Christian Bastien, Patrick Le Callet, Maurice Milgram et Lionel Prevost.

Je veux exprimer ici mes sincères remerciements à M. Jean-Philippe Tarel, mon encadrant scientifique, qui m'a permis de travailler sur ce sujet qui me passionne. Je souhaite lui exprimer ma profonde reconnaissance pour ses qualités humaines, son amitié, son accompagnement au quotidien, pour ses conseils judicieux et également sa rigueur qu'il a su me transmettre tout en me laissant une part d'autonomie et de créativité durant mon travail de recherche. Je le remercie également pour le temps passé à relire et corriger nos nombreuses publications et documents réalisés.

Au sein du LCPC, je tiens à remercier particulièrement Roland Brémond, Eric

Dumont et Nicolas Hautière pour les innombrables conseils, discussions et relectures. Je remercie également l'équipe administrative et la direction scientifique du LCPC pour leur aide précieuse dans les diverses formalités. Merci à Henri Panjo pour l'étude statistique réalisée durant son passage au laboratoire.

Je réserve ici une place toute particulière à mon inséparable, Anaïs Mayeur, avec qui j'ai eu le bonheur de partager le bureau durant les trois années de nos thèses. Je lui exprime toute mon amitié et ma reconnaissance pour son écoute attentive, nos discussions et ses conseils en psychologie cognitive qui ne sont pas étrangers au contenu de ce mémoire de thèse.

Je remercie chaleureusement ceux de ma famille qui ont cru en moi, m'ont encouragé et soutenu, tout particulièrement mes parents Jean et Jocelyne, ma soeur Aurélie, mon frère Patrick et mon grand-père Claude Cleenewerck qui comptent beaucoup pour moi. J'associe à ses remerciements une pensée particulière pour ma grand-mère, Monique Cleenewerck née Raison, qui m'a toujours épaulé et nous a quitté, trop tôt, quelques mois avant le début de cette thèse.

Je remercie également pour leur soutient ma belle-famille et particulièrement ma concubine, Hélène Tselikis, qui m'a supporté (dans les deux sens du terme) durant tout les moments de cette thèse.

J'adresse aussi un clin d'oeil à l'ensemble de mes amis et à mes connaissances du milieu associatif, ils se reconnaîtront. Un salut particulier à mes plus proches et fidèles amis de toujours, Nicolas Milenovic et Fabien Wielgosz pour leur présence de chaque instant et à mon entraîneur de tir sportif, Bernard Grevillot, pour nos discussions et ses conseils avisés.

Je remercie enfin, tous ceux que j'ai rencontrés durant ces trois ans, professeurs, chercheurs, doctorants, ingénieurs techniciens, administratifs et qui ont contribué, par leur soutien des près ou de loin, à l'élaboration de cette thèse.

Mes excuses aux personnes que j'aurais éventuellement oubliées dans ces remerciements.

Résumé

La signalisation routière a un rôle qui n'est plus à démontrer dans la sécurité et l'exploitation des infrastructures. L'ingénieur et le gestionnaire routier recherchent, respectivement, à fabriquer et à disposer de manière pertinente, des panneaux de signalisation suffisamment saillants pour attirer l'attention de l'usager, focalisé sur sa tâche de conduite. Le gestionnaire routier est aussi garant de la qualité de service de l'infrastructure. De ce fait, il est demandeur d'outils d'aide à la décision qui lui permettent de hiérarchiser les priorités des interventions à effectuer et ainsi, gérer au mieux le budget qui lui est alloué. L'automatisation de ce type de diagnostic apparaît donc nécessaire. L'application majeure visée est le diagnostic automatique de la signalisation verticale le long des réseaux routiers, grâce à l'utilisation d'une caméra embarquée dans un véhicule. Ainsi, la saillance des panneaux pourrait être améliorée et d'éventuels risques accidentogènes diminués. Pour cela, nous proposons dans le cadre de cette thèse, d'étudier la faisabilité d'un algorithme d'estimation automatique de la saillance de la signalisation verticale dans les images routières, corrélé au comportement de l'usager en situation simulée de conduite automobile.

Aucun modèle complet de la saillance visuelle n'est actuellement disponible, encore moins de modèle prenant en compte les processus cognitifs mis en jeu lors de la tâche de conduite. Deux types de saillance visuelle sont à distinguer. D'une part, la saillance attentionnelle dans laquelle l'information saillante s'impose au conducteur et guide son attention. Et d'autre part, la saillance de recherche, dans laquelle le conducteur recherche (volontairement ou inconsciemment) l'information pertinente par rapport à la tâche en cours. Des expérimentations ont montré que la saillance attentionnelle serait moins utilisée lors de la prise d'information en conduite que la saillance de recherche. Or, il n'existe pas de modèle complet de cette dernière.

Nous proposons donc une modélisation de la saillance de recherche d'un objet défini. Pour détecter les zones saillantes, nous utilisons un algorithme d'apprentissage nommé "Support Vector Machines" (SVM) connu pour son efficacité. Cet algorithme modélise les variations de l'aspect de l'objet que nous cher-

chons à détecter et, de plus, donne une estimation de sa confiance en la détection. Notre approche est de relier cette confiance à la saillance de recherche. L'utilisation d'une signature (représentation des données) sous forme d'histogramme des couleurs normalisées et du noyau (mesure de dissimilarité) triangulaire permet, outre une rapidité accrue du SVM, d'être corrélée à la vérité terrain et à la performance visuelle du conducteur. La valeur de décision du SVM calculée à plusieurs échelles donne la saillance intrinsèque de l'objet détecté. Une compensation prend en compte la saillance du fond. La taille de l'objet détecté est ensuite introduite dans le modèle.

Nous avons vérifié que l'estimation de la saillance que nous proposons est corrélée à la perception du conducteur. La mise au point d'un protocole d'expérimentation d'oculométrie cognitive nous a permis de recueillir des données sur le comportement visuel en situation de conduite simulée. La saillance réelle de chaque panneau est obtenue par une évaluation objective (points de fixations du regard) et une évaluation subjective (score). Le traitement statistique des données recueillies nous a permis d'évaluer certains paramètres du modèle. La corrélation du modèle avec le comportement humain en situation de conduite simulée prouve ses qualités afin de mesurer la saillance de la signalisation verticale.

Mots-clefs

Saillance de recherche, modèle «top-down», Saillance attentionnelle, modèle «bottom-up», Signalisation Verticale, Panneaux, Évaluation de la qualité de service de l'infrastructure, Apprentissage, SVM, Détection, Classification, Expérimentation, Oculométrie cognitive.

Road signs saliency in road images
Feasibility study of a diagnostic tool

Abstract

Road signs play a significant role in traffic control and road safety trough driver's guidance, warning, and information. They are the main communication media towards the drivers. To be effective, they must be visible, legible and comprehensible. Road engineers and traffic manager, respectively aim at designing and placing in the road environment, road signs enough salient so that they attract the driver's attention, regardless of the driver's preoccupation who is focused to the driving task. The road manager is also responsible for the maintenance of the quality of service of the road infrastructure. As a result, he needs tools to help him prioritize his interventions, so as to optimize the budget he is allocated. The automation of such diagnostic tools is desirable. The diagnostic of road signs along a road network may be automated by the processing of images acquired by a camera onboard a dedicated vehicle. Eventually, the road sign saliency could be enhanced and possibles accidents risks diminished. For this purpose, in this thesis, we study the feasibility of an algorithm to automatically estimate the saliency of road signs in road scene images with a criterion correlated to the visual performance of drivers in near-driving situations.

To date, no complete model of visual saliency is available, let-alone model which considers the cognitive processes implied during the driving task. Two kinds of saliency need to be distinguish. On the one hand, there is the attentional saliency, in which salient information establishes itself to the driver and guides his attention. On the other hand, there is the search saliency, in which the driver seeks (voluntarily or unconsciously) the information relevant for his ongoing task. Some experimentations tend to show that attentional saliency is less used during the information taking in driving than search saliency. For which there is, unfortunately, no available complete model.

We propose a model of the search saliency for an object of interest. In order to detect salient areas, we rely on a learning algorithm called " Support Vector Machines " (SVM) which proves quite reliable performances for learning object

appearance in many pattern recognition applications. This algorithm is able to infer the frontier that splits the feature space into regions associated to the appearance of the object of interest and regions associated to the background. The SVM models the variability in the appearance of the object we try to detect, a (set of) road sign(s), and moreover, computes a confidence value about the estimated detections. Our seminal thinking is to link this confidence with the search saliency. The use of a feature vector composed by color histogram in the normalized space, and the triangular kernel, enables, besides an increased speed of the SVM, to obtain a saliency indicator correlated to the ground truth and to the driver's visual performance. The confidence value of the SVM computed at different scales, gives the so-called intrinsical saliency of the detected object. An off-setting processes on this value is done considering the saliency of the related background. The size of the detected object is next introduced in the model.

We checked that the estimate of the road sign saliency that we proposed is correlated to the driver's perception. The adjustment of an experimental protocol in cognitive eye tracking enabled us to collect visual behavior data in near-driving situations. The actual saliency of every road signs is assessed by an objective evaluation (detection rates and fixation durations) and a subjective evaluation (scoring). A statistical analysis enabled us to tune some parameters of the model. The correlation of the proposed model with the human behavior in near-driving situation proof is quality in order to valuate the road sign saliency.

Key-words

Search saliency, «top-down» models, Attentional saliency, «bottom-up» models, Conspicuity, Road signs, Traffic signs, Road environnement, Road manager, Learning, SVM, Detection, Classification, Experimentation, Cognitive Eye tracking.

Table des matières

1 Introduction **12**

1.1 Contexte 12

1.2 Objectifs : La détection de panneaux pour l'estimation de la saillance corrélée avec le comportement humain 16

1.3 Plan de la thèse 18

2 La signalisation verticale, une nécessité et des besoins **20**

2.1 Conception et certification de la signalisation verticale 20

2.2 Importance de la conception d'outils de diagnostic pour le gestionnaire routier 25

3 Vision, saillance de la signalisation verticale dans l'activité de conduite **28**

3.1 Éléments de physiologie du système visuel humain 29

3.2 Activité de conduite et prise d'information visuelle 36

3.3 Deux types de saillance en conduite 40

3.3.1 Les modèles informatiques de la saillance attentionnelle . 42

3.3.2 Les modèles informatiques de saillance de recherche . . . 45

3.3.3 Vers une modélisation de la saillance de recherche en conduite pour la caractérisation de la saillance de la signalisation de police 46

3.4 L'oculométrie, une méthode pour étudier la prise d'information visuelle en conduite 48

3.4.1 Fonctionnement et limite de l'oculométrie 49

3.4.2 Les indicateurs du comportement visuel humain utilisés en oculométrie cognitive 54

4 Analyse d'images pour la détection de la signalisation verticale 59

4.1 État de l'art sur la détection de panneaux, contraintes spécifiques liées à notre problème 59

4.2 Apprentissage par SVM 61

5 Un SVM pour détecter les zones saillantes cohérentes à la signalisation verticale 64

5.1 Choix des SVMs 64

5.2 Construction des bases d'exemples pour l'apprentissage et les tests 66

5.3 Procédure d'apprentissage et de recherche de panneaux 69

5.3.1 Apprentissage standard 69

5.3.2 Influence du paramètre de régularisation dissymétrique . . 72

5.3.3 Détection par fenêtres en multi-échelles 73

5.4 Etude sur les signatures 76

5.4.1 Histogramme des angles et des couleurs des contours . . . 77

5.4.2 Histogramme des couleurs 79

5.4.3 Histogramme des couleurs normalisées 79

5.4.4 Histogramme des couleurs normalisées sur les contours . . 80

5.4.5 Histogramme de la distribution des couleurs normalisées autour des contours 82

5.4.6 Histogramme des angles des contours OEH 84

5.4.7 Histogramme de la concaténation OEH et couleurs 84

5.5 Etude sur les Noyaux 85

5.5.1 Noyau Gaussien 85

5.5.2 Noyau Polynomial 86

5.5.3 Noyau Laplacien 86

5.5.4 Noyau Triangulaire 86

5.6 Choix des couples noyau/signature de meilleures performances . . 88

5.6.1 Vérité Terrain, la signalisation verticale 89

5.6.2 Données de comportement humain, expérimentation exploratoire sur les fixations du regard dans un contexte routier 89
5.6.3 Choix du couple en fonction des deux vérités 96

6 Un modèle d'estimation de la saillance de recherche basé sur les SVMs 101
6.1 Utilisation des valeurs de confiance du SVM comme estimateur de la saillance intrinsèque 101
6.2 Prise en compte du fond environnant 104
6.3 Prise en compte de la taille de la signalisation verticale 106
6.4 Calcul de saillance des objets routiers à partir du modèle proposé . 108

7 Expérimentation de validation par oculométrie cognitive 111
7.1 Méthodologie 112
7.1.1 Matériel et dispositif expérimental 113
7.1.2 Participants 118
7.1.3 Procédure 120
7.1.4 Variables étudiées 133
7.2 Analyse des données et modèle statistique 134
7.2.1 Evaluation objective 134
7.2.2 Evaluation subjective 135
7.3 Résultats et Interprétation : Une corrélation positive entre le modèle et le comportement humain 137

8 Conclusion et perspectives 142

Chapitre 1

Introduction

1.1 Contexte

La signalisation routière a un rôle qui n'est plus à démontrer dans la sécurité et l'exploitation des infrastructures routières. Elle est, depuis ses lointaines origines (une ordonnance royale de 1539 créant la police de la circulation et interdisant le dépassement et le demi-tour sur certaines voies), jusqu'à aujourd'hui, et pour de nombreuses années, le principal média d'information entre, d'une part le gestionnaire de voirie et l'autorité de police, et d'autre part les usagers de la route.

Visibilité, lisibilité, uniformité, homogénéité, simplicité, continuité des directions signalées, cohérence avec les règles de circulation et avec la géométrie de la route constituent les grands principes de la signalisation. Ils sont intangibles pour que l'usager puisse toujours la comprendre, lu accorder confiance et la respecter.

Ces principes sont déclinés dans la réglementation de la signalisation routière qui trouve ses fondements dans la convention internationale signée à Vienne en 1968, les accords européens signés à Genève le 1er mai 1971, le code de la route (articles R.411-25 à R. 411-28), l'arrêté interministériel du 24 novembre 1967 modifié relatif à la signalisation des routes et autoroutes, l'instruction interministérielle sur la signalisation routière [2] et l'instruction interministérielle relative à la signalisation de direction.

Comme le précisent les textes de loi et les recommandations de la sécurité routière [3], ce corpus juridique s'applique à l'ensemble des voies ouvertes à la circu-

lation publique et tous les maîtres d'ouvrages et gestionnaires routiers doivent s'y conformer. La mise en place d'une signalisation non conforme à la réglementation est interdite.

Cette réglementation est décrite dans l'instruction ministérielle sur la sécurité routière (dont la dernière version consolidée est en date de novembre 2008 [2]), afin de répondre aux besoins des usagers de la route et à ceux du gestionnaire. Cette instruction fixe la nature des signaux, ainsi que les conditions et les règles de leur implantation. Elle donne les principes qui permettent au gestionnaire de vérifier l'ensemble de son infrastructure et la cohérence de la signalisation. Afin de répondre aux nouveaux besoins des usagers de la route et à ceux du gestionnaire, l'évolution de la réglementation de la signalisation routière se fait par des ajouts réguliers dans l'instruction ministérielle sur la sécurité routière. Par exemple, la dernière version consolidée, en date de novembre 2008 [2]), précise les modalités de mise en oeuvre et de fonctionnement de la signalisation spécifique au transport en commun (tramway, bus), récemment apparue sur le réseaux routiers urbain. Elle ajoute également quelques recommandation sur la prise en compte de l'activité de conduite pour l'implantation des panneaux.

L'importance du rôle de la signalisation routière s'accroît avec le développement de la circulation. Face à l'évolution des règles, il est nécessaire pour le gestionnaire de vérifier régulièrement que la signalisation implantée dans son réseau routier est conforme aux normes en vigueur. En amont, analyser et comprendre la perception visuelle de l'infrastructure par l'usager est primordial pour définir des règles générique et améliorer la sécurité routière [4].

Les chercheurs-ingénieurs et le gestionnaire routier cherchent respectivement à concevoir et à disposer de manière pertinente, dans l'environnement routier, des panneaux de signalisation, suffisamment saillants pour attirer l'attention du conducteur en dépit de sa focalisation sur ses différentes tâches de conduite. En effet, les ressources attentionnelles du conducteur sont plus ou moins sollicitées selon son niveau d'expertise en conduite [5], mais aussi selon l'environnement dans lequel il se déplace [6]. Si l'on souhaite que certaines informations lui parviennent de manière efficace et adaptée pour induire un comportement routier sûr, il faut que ces informations soient saillantes dans son champ visuel [7]. De nombreuses définitions de la saillance ont été proposées [8]. Pour [9], un objet est

« conspicious » s'il est détecté avec une probabilité de plus de 90% pendant un temps très court (250ms) sur un fond complexe quelconque et pour une position quelconque par rapport à la direction du regard. Pour [10], la saillance est la propriété d'un objet de paraître évident dans un environnement complexe. Nous nous accordons à la définition suivante : la saillance est le degré selon lequel un objet attire l'attention selon un fond donné [11]. Elle dépend de divers paramètres tels que la taille, la couleur et l'emplacement dans l'environnement routier, et ceci en fonction des processus cognitifs inhérents à l'être humain.

Il est donc intéressant de réaliser des outils de mesure apportant aux gestionnaires une aide au diagnostic du réseau routier dont il est responsable et décisionnaire. Notamment, de tels outils trouvent leur intérêt dans la vérification de la bonne implantation des panneaux dans l'environnement routier, par rapport aux principes de la signalisation (Visibilité, lisibilité, uniformité, homogénéité, simplicité, continuité des directions signalées, cohérence avec les règles de circulation et avec la géométrie de la route).

Au sein du Réseau Scientifique et Technique (RST) du Ministère de l'Écologie, de l'Énergie, du Développement Durable et de la Mer (MEEDDM) auquel le Laboratoire Central de Ponts et Chaussées (LCPC) et l'Institut National de Recherche sur les Transports et leur Sécurité (INRETS) appartiennent, les chercheurs et ingénieurs sont force de proposition dans les applications à destination du gestionnaire (cf. partie 2.2). Certains outils, déjà développés, permettent de caractériser la visibilité photométrique des objets routiers (marquage, chaussées, panneaux, obstacles, etc.), d'autres, en cours de développement, caractérisent la visibilité géométrique de la chaussée (visibilité offerte par l'infrastructure) et la visibilité météorologique (par temps de brouillard). En revanche, le gestionnaire ne dispose d'aucun outil lui permettant de réaliser un diagnostic de la qualité de la saillance des objets routiers. En effet, ce type de critère de visibilité fait intervenir l'aspect cognitif de la tâche de conduite du conducteur, et la prise de connaissance de l'information portée par les signaux routiers dans laquelle l'information visuelle est prédominante (cf. chapitre 3). La perception de l'infrastructure est alors déterminée par divers paramètres, tels que :

- les caractéristiques géométriques (le tracé de la chaussée, etc.) ;
- les caractéristiques photométriques de l'environnement (conditions d'éclai-

rage, état de la chaussée, usure des équipements de la route) ;
- les caractéristiques atmosphériques (pluie, brouillard, neige, etc.) ;
- les caractéristiques du système visuel humain ;
- les caractéristiques des processus cognitifs liés à la tâche de conduite.

Pour réaliser le diagnostic de la saillance de la signalisation verticale, le gestionnaire ne peut, à ce jour, que se fier à l'avis des services chargés des voiries, qui estiment visuellement si un panneau donné est suffisamment saillant pour un conducteur. Cette évaluation est, de ce fait, elle-même biaisée par l'appréciation subjective de l'évaluateur. Le relevé de la signalisation et de ses propriétés, par cette méthode, est coûteux en moyens humains et, de part ce caractère humain, l'évaluation ne peut être assurée fiable et répétable. De plus, face à la taille du réseau, ces moyens restent souvent insuffisants pour assurer complètement l'inspection de la signalisation et donc sa qualité de service. Il est donc souhaitable d'aller vers une automatisation de ce type de diagnostic, au moyen de véhicules instrumentés avec des caméras, ce qui permettrait d'améliorer la saillance de la signalisation et donc d'augmenter le confort et la sécurité des usagers.

Pour cela, nous proposons dans le cadre de cette thèse, *d'étudier la faisabilité d'un algorithme d'estimation automatique de la saillance de la signalisation verticale dans les images routières, corrélé au comportement de l'usager en situation simulée de conduite automobile*. L'objectif est, à long terme, d'intégrer les indicateurs issus de l'algorithme dans les outils dont dispose le gestionnaire, tels que les dispositifs de type IRCAN (imagerie routière par caméra numérique). Ces dispositifs acquièrent des séquences d'images numériques en couleur de la route et de son environnement, qui une fois associées au logiciel d'exploitation IREVE (imagerie routière, étalonnages, visualisations, exploitations), permettent aux gestionnaires de constituer et d'exploiter des banques d'images de leurs réseaux routiers.

1.2 Objectifs : La détection de panneaux pour l'estimation de la saillance corrélée avec le comportement humain

Si le concept visibilité photométrique d'un panneau seul est bien défini [12], il n'existe pas de définition claire de celle-ci quand ce même panneau est implanté dans un environnement routier complexe [11]. On parle alors de saillance visuelle. Mais, on ne dispose pas de modèle complet de la propension à être perçu, encore moins de modèle prenant en compte les tâches à effectuer, tandis que de nombreux processus cognitifs sont mis en jeu lors de la conduite [13].

L'objectif de cette thèse est de mettre au point et d'évaluer un système d'estimation automatique de la saillance des panneaux de signalisation à partir d'images numériques de scènes routières. Une application majeure visée est l'utilisation de ce système avec une caméra embarquée dans un véhicule (IRCAN), dans le but de faciliter le contrôle de la qualité de service de la signalisation verticale le long des réseaux routiers.

Avant d'estimer la saillance d'un panneau, il faut d'abord le détecter automatiquement dans les images. Le choix de traitements d'image idoine permet d'obtenir aussi un indice de confiance sur la présence du panneau dans la scène. Notre paradigme de base est d'utiliser cet indice pour définir la saillance. Notre référence étant la saillance perçue par les conducteurs, un autre objectif est donc d'analyser le comportement visuel humain devant des scènes routières afin de valider ou d'infirmer nos différentes approches du calcul de la saillance. Cette étude nécessite l'utilisation d'un matériel spécifique (oculomètre et salle photométriquement contrôlée) et d'un protocole expérimental permettant le contrôle des différentes variables.

Un panneau, même s'il est un objet bien normalisé, peut avoir des aspects assez variés, du fait de son implantation, des variations d'éclairage et des conditions météorologiques. De ce fait, nous choisissons de traiter le problème de la détection avec des méthodes d'apprentissage, qui permettent de modéliser ces différents aspects. Ainsi, les quatre questions principales de cette thèse sont :

- Comment expliquer la saillance afin de lui trouver un procédé de calcul dans les images ?
- Quelles techniques d'apprentissage seront à adopter pour effectuer la détection de la signalisation verticale ?
- Quel estimateur de confiance en la détection pourra-t-on en tirer pour en extraire un indicateur de saillance ?
- Une fois cet indicateur conçu, comment évaluer cet indicateur vis-à-vis du comportement visuel d'un usager de la route ?

Pour répondre à ces questions, le travail de la thèse a été décomposé en huit étapes :

1. Étude de la spécificité du sujet d'étude, la signalisation verticale, cf. Chap. 2.
2. Comprendre le fonctionnement du système visuel humain et de la perception du conducteur afin de définir ce qu'est la saillance de la signalisation verticale, cf. Chap. 3.
3. Etude de l'état de l'art sur la détection des panneaux afin de trouver la méthode qui nous donne de bons résultats et nous permet d'approcher une estimation de la saillance lors de la détection, cf. Chap. 4.
4. Choix du processus d'apprentissage et de classification pour effectuer la détection de la signalisation : les Machines à Vecteurs de Support (SVM), cf. partie. 5.1.
5. Constitution de la base de données de référence. Avant d'effectuer un apprentissage, il faut en effet une base contenant des panneaux sous de multiples aspects, cf. partie 5.2.
6. Représentation des données. Il est très coûteux d'utiliser les panneaux extraits de l'image directement, lors de l'apprentissage. Il est plus astucieux de choisir une représentation intermédiaire des objets à apprendre et reconnaître, et de choisir une mesure de dissimilarité adaptée aux données, cf. partie. 5.6.
7. Modèle et estimation de la saillance. Une fois l'algorithme de détection mis au point, nous avons utilisé la valeur de décision fournie par le SVM afin

de proposer un modèle calculatoire de la saillance de l'objet détecté, cf. Chap. 6

8. Validation du modèle. L'estimation de la saillance de la signalisation devant être corrélée à sa perception par le conducteur, une partie importante de cette thèse a consisté en la mise au point d'un protocole d'expérimentation d'oculométrie cognitive afin de recueillir des données de comportement visuel en situation la plus écologique possible, par rapport à la tâche de conduite. Ensuite, le traitement statistique des données recueillies nous a permis de vérifier la présence d'une corrélation entre la saillance mesurée informatiquement et celle évaluée par les sujets, cf. Chap. 7.

1.3 Plan de la thèse

Le manuscrit de cette thèse est organisé comme suit :

– L'introduction présente le contexte dans lequel s'inscrit cette thèse, ainsi que les objectifs fixés pour concevoir et valider le modèle d'estimation de la saillance de la signalisation verticale.
– Le deuxième chapitre aborde la conception et la certification de la signalisation verticale, ainsi que les besoins des gestionnaires routiers en termes d'outils de diagnostic.
– Le troisième chapitre rappelle le fonctionnement du système visuel humain et introduit la notion de saillance visuelle, en la resituant dans l'activité de conduite.
– Le quatrième chapitre présente l'état de l'art sur la détection de panneaux par traitement d'images, ainsi que l'analyse des contraintes spécifiques à notre application et le principe de l'apprentissage par SVM utilisé.
– Le cinquième chapitre détaille notre approche de détection de la signalisation verticale par SVM. Nous y présentons les différents choix de paramètres, de signatures et de noyaux intéressants notre cas. La première expérimentation exploratoire, qui nous a permis d'avancer dans nos choix, est également présentée.

- Nous présentons, dans le sixième chapitre, notre paradigme d'utilisation des valeurs de confiance du SVM comme noyau du modèle de saillance. Le processus de compensation pour prendre en compte la saillance du fond environnant du panneau, ainsi que la prise en compte des aspects de taille, qui complètent le modèle, y sont également présentés.
- Le septième chapitre décrit l'expérimentation d'oculométrie cognitive qui a été développée et qui est nécessaire pour récolter les données de comportement visuel humain en situation proche de l'acte de conduite. Les principes de l'oculométrie et les indicateurs du comportement visuel humain sont au préalable introduits. En fin de ce chapitre, le protocole de l'expérimentation psycho-cognitive et l'analyse statistique des données sont présentés. Nous y mettons également en avant les résultats de corrélation obtenus entre notre modèle d'estimation de la saillance de la signalisation verticale et les indicateurs de saillance issus des sujets de l'expérimentation.
- Enfin, nous récapitulons dans la conclusion les contributions dans les divers domaines de cette thèse et présentons des pistes de recherche et de développements pour des travaux futurs.

Chapitre 2

La signalisation verticale, une nécessité et des besoins

2.1 Conception et certification de la signalisation verticale

Bien conçue et réalisée, la signalisation routière réduit les causes d'accident et facilite la circulation. Insuffisante, trop abondante ou impropre, elle est facteur de gêne et d'insécurité.

Les catégories de signalisation sont d'après [2] :

- la signalisation verticale :
  - par panneaux,
  - par feux,
  - par balisage,
  - par bornage.
- la signalisation horizontale, c'est à dire la signalisation par marquage des chaussées.

Les principaux critères d'efficacité sont, toujours d'après [2] :

- *L'uniformité,* qui impose l'usage exclusif de signaux réglementaires, sur toutes les voiries.

- *L'homogénéité*, qui exige que, dans des conditions identiques, l'usager rencontre des signaux de même valeur et de même portée, implantés suivant les mêmes règles.
- *La simplicité*, qui s'obtient en évitant une surabondance de signaux qui surcharge l'attention de l'usager, lequel tend alors à négliger les indications données ou même ne peut les lire, les comprendre ou les mémoriser.
- *La continuité des directions signalées*, qui est assurée sur les routes importantes par la coordination effectuée à l'échelon de l'Administration Centrale. Cette continuité doit être recherchée sur toutes les autres routes en réalisant, localement, entre services, les liaisons nécessaires.

La signalisation routière a pour objet de rendre plus sûre la circulation routière, de faciliter cette circulation, d'indiquer ou de rappeler diverses prescriptions particulières de police et de donner des informations relatives à l'usage de la route. Toutefois, la signalisation n'a et ne saurait avoir le caractère d'une garantie assurée par les pouvoirs publiques aux usagers de la route contre les aléas et les dangers de la circulation. Ces usagers circulent toujours à leurs risques et périls.

Dans la conception et l'implantation de la signalisation verticale, les conditions de perception de l'usager ne doivent jamais être perdues de vue : il se déplace à une vitesse non négligeable et son attention est sollicitée par les exigences de la conduite [14]. De par l'article 4 de l'instruction interministérielle sur la signalisation routière [2], outre les règles d'implantation et de normalisation des panneaux, le gestionnaire doit s'appuyer sur trois grands principes de base de la signalisation, afin qu'elle soit efficace et ainsi utile. Ces trois principes sont :

- *Le principe de valorisation*. L'inflation des signaux nuit à leur efficacité (rmq : et réduit leur saillance). Il ne faut donc en placer que lorsqu'ils sont jugés vraiment utiles. Ceci rejoint le critère de simplicité.
- *Le principe de concentration*. Lorsqu'il est indispensable que plusieurs signaux soient vus en même temps ou à peu près en même temps, on doit les implanter de façon à ce que l'usager puisse les percevoir d'un seul coup d'œil, de nuit comme de jour. Il y a un intérêt à grouper deux signaux sur un même support lorsque les deux indications se rapportant au même point se complètent l'une à l'autre. Le principe de concentration trouve toutefois

une limite dans le principe suivant.

– *Le principe de lisibilité*. Il ne faut pas demander à l'usager un effort de lecture ou de mémoire excessif. On doit donc réduire et simplifier les indications au maximum et le cas échéant répartir les signaux sur des supports échelonnés.

Les panneaux de signalisation français, cf. Fig. 2.1, se répartissent en 18 types, pour un total de 576 panneaux [3] :

– Type A - Panneaux de danger (29 panneaux).
– Type AB - Panneaux d'intersection et de priorité (9 panneaux).
– Type B - Panneaux de prescription (88 panneaux) se subdivisant en cinq catégories : Panneaux d'interdiction (40 panneaux) - Panneaux d'obligation (18 panneaux) - Panneaux de fin d'interdiction (6 panneaux) - Panneaux de fin d'obligation (8 panneaux) - Panneaux de prescription zonale (16 panneaux).
– Type C - Panneaux d'indication utiles pour la conduite des véhicules (52 panneaux).
– Type CE - Panneaux d'indication d'installations pouvant être utiles aux usagers de la route (39 panneaux).
– Type D - Panneaux de direction (56 panneaux).
– Type Dp - Panneaux de jalonnement piétonnier (4 panneaux).
– Type Dv - Panneaux de jalonnement des aménagements cyclables (13 panneaux).
– Type Dc - Panneaux de signalisation d'information locale (2 panneaux).
– Type E - Panneaux de localisation (26 panneaux).
– Type EB - Panneaux de début et de fin d'agglomération (2 panneaux).
– Type G - Panneaux de position des passages à niveau (10 signaux).
– Type H - Panneaux d'information (10 panneaux).
– Type ID - Idéogrammes, emblèmes et logotypes (101 panneaux).
– Type AK, K, KC et KD - Panneaux de signalisation temporaire (35 panneaux).
– Type M - Panneaux additionnels ou panonceaux (59 panneaux).
– Type SE - Symboles (37 signaux).

– Type SR - Panneaux d'information de sécurité routière (4 panneaux).

Les autres signaux routiers constituant le reste de la signalisation verticale (bornes, balises et feux de signalisation) sont répartis en 3 types, pour un total de 28 signaux différents :

– Type F - relatif aux bornes de jalonnement kilométrique et hectométrique qui figurait dans l'arrêté du 24 novembre 1967 initial, mais a été supprimé depuis.
– Type J - Balises (13 signaux).
– Type R - Feux et signaux lumineux (15 signaux).

D'autre part, on distingue quatre grandes catégories de panneaux dénommées comme suit :

– SP (signalisation de police permanente).
– SD (signalisation directionnelle permanente).
– TP (signalisation de police temporaire).
– TD (signalisation directionnelle temporaire).

On cherche ici à réaliser un outil de diagnostic de la signalisation en place dans l'infrastructure du réseau routier habituel, nous ne prenons donc pas en compte l'étude des spécificités de la signalisation temporaire. Par ailleurs, comme détaillé dans le chapitre 3, les panneaux de direction et de police ne requièrent pas le même type de saillance.

Les panneaux de police sont les types de panneaux mis en jeu dans la sécurité routière, nous limitons donc l'outil de diagnostic aux panneaux de Signalisation Permanente de Police (catégorie SP), comprenant tout panneau utilisé en signalisation permanente de types A, AB, B, C, CE, G, J et M conformément à la norme [15], soit un total de 299 panneaux et panonceaux.

Dans cette catégorie, comme le montre la Fig.2.1, on trouve des panneaux qui prennent, selon leurs significations, des formes et couleurs différentes arrêtées grâce à un consensus international (ce qui n'est pas le cas pour la signalisation de direction) :

– panneaux de danger (Type A) : forme triangulaire, fond blanc, listel rouge.

– panneaux relatifs aux intersections et notifiant les régimes de priorité (Type AB) : forme triangulaire, fond blanc et listel rouge ; forme carrée (panneaux placés sur pointe) fond jaune et listel blanc ; et forme octogonale fond rouge et listel blanc.
– panneaux de prescription (Type B) : forme circulaire ; fond blanc et listel rouge pour les interdictions et fond bleu pour les obligations.
– autres panneaux (Type C, CE, G, J et M) : rectangulaire à côtés horizontaux et verticaux, fond bleu, ou blanc avec listel bleu.

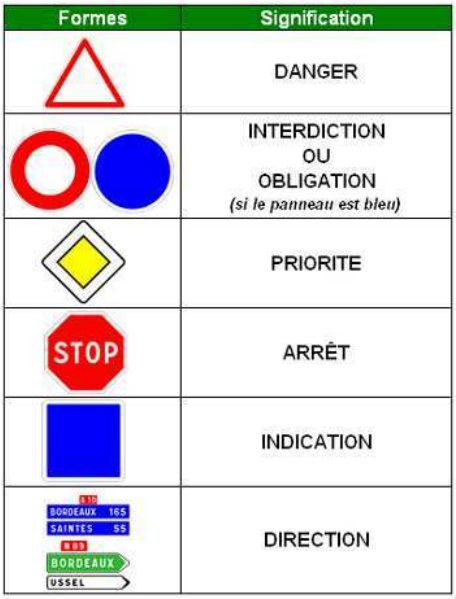


FIGURE 2.1 – Caratéristiques principales de formes et de couleurs de la signalisation verticale en fonction de la signification. Source wikipédia.

Ces panneaux ont également des tailles différentes en fonction des vitesses pratiquées (autorisées). Ainsi, un panneau circulaire peut présenter différents diamètres, idem pour les les panneaux triangulaires.

2.2 Importance de la conception d'outils de diagnostic pour le gestionnaire routier

De par les lois et règlements concernant la signalisation routière, le gestionnaire doit respecter les divers principes, critères et règles énoncés afin d'assurer la conformité de son réseau routier.

Ainsi, l'article 2 de l'instruction interministérielle sur la signalisation routière [2], énonce qu'elle « complète l'arrêté du 24 novembre 1967 et précise les règles à suivre, tant pour l'implantation que pour la nature des signaux à adopter. Elle s'impose dans les conditions qu'elle édicte à tous ceux qui sont à un titre quelconque habilités à mettre en place la signalisation routière, sur les voies ouvertes à la circulation publique ». Or, il est aussi indiqué dans son préambule que cette instruction « représente l'idéal vers lequel on doit tendre. Toutefois, la signalisation effectivement mise en place peut être moins dense pour des raisons d'ordre pratique (faible circulation, nécessité d'éviter la multiplication de panneaux, choix de la meilleure affectation des crédits). La responsabilité de l'Administration ne saurait être mise en cause en pareil cas ». Le gestionnaire est donc finalement le seul décisionnaire de l'implantation sur son réseau et reste légalement responsable de l'exploitation et de la maintenance de celui-ci, l'article 18 appelant son attention sur la nécessité d'assurer l'entretien des signaux et de leurs supports.

L'article 4 de l'instruction interministérielle sur la signalisation routière, précise à propos du principe de visibilité que « l'expérience prouve qu'il y a intérêt à inspecter périodiquement la signalisation de jour et de nuit, avec un « œil neuf » pour faire disparaître les panneaux superflus ou remédier aux insuffisances éventuelles ». A propos du principe de lisibilité, elle précise également que « des expériences ont montré que l'observateur moyen ne peut d'un seul coup d'œil percevoir et comprendre plus de deux symboles ».

Le gestionnaire, voulant s'assurer de la conformité du réseau routier à sa charge, se trouve alors confronté à l'analyse de son patrimoine alors que cela requiert une connaissance fine des instructions et une approche d'ergonomie cognitive auxquelles les services techniques ne sont pas ou peu formés.

La perception visuelle de la route et de son environnement constitue un objet d'applications et d'études important en vue de l'amélioration de la sécurité [4]. Cette thématique est étudiée depuis de nombreuses années au sein du Réseau Scientifique et Technique (RST) du MEEDDM. Ceci, depuis le début des années 80 au Laboratoire Centrale des Ponts et Chaussées (LCPC) dans la division « Exploitation, Signalisation, Eclairage » et plus récemment dans plusieurs Équipes de Recherche Associées (ERA), Laboratoires Régionaux (LRPC) et Centre d'Etude Technique de L'Équipement (CETE) situés à Strasbourg, Saint-Brieuc, Rouen, Angers et Clermont-Ferrand.

Divers outils d'aides à la mesure et au diagnostic ont déjà été mis au point et proposé aux gestionnaires, pour caractériser la visibilité photométrique de différents types d'objets routiers, mais pas leur saillance :

- CYCLOPE : Véhicule instrumenté par une caméra en position de conduite, permettant de savoir si les niveaux de visibilité des objets routiers (panneaux, marquages) situés dans la scène routière respectent bien le modèle d'Adrian, ou bien encore de contrôler la luminance (cd/m^2) moyenne produite par l'éclairage public ainsi que son homogénéité transversale et longitudinale.
- COLUROUTE : COefficient de LUminance des ROUTEs, appareil $mlpc$® permettant la mesure in situ des caractéristiques photométriques des chaussées (regard porté à 60m).
- VECLAP : Véhicule de contrôle de l'ÉCLAirage Public, permettant une mesure en continu de la qualité de service de l'éclairage public. Ceci afin d'effectuer une analyse de la visibilité, du confort et de la sécurité nocturne.
- ECODYN : Appareil $mlpc$® à grand rendement pour le contrôle, en continu, de jour, de la visibilité diurne et nocturne de la signalisation horizontale (marquages).

Il est aussi proposé aux gestionnaires de caractériser la visibilité géométrique de l'infrastructure en développant des outils d'analyse de la visibilité offerte, que ce soit par :

- Comparaison directe de la visibilité requise (distance d'arrêt, visibilité en

carrefour, ...) avec la visibilité offerte. Ceci par l'utilisation de VISULINE (LRPC St Brieuc), qui est un dispositif de saisie embarquée des conditions de visibilité (agent dans le véhicule) entre deux véhicules localisés par GPS.

– Mesure indirecte de la visibilité géométrique de la chaussée (sans *a priori* d'un opérateur) que l'on retrouve dans le projet SARI/VIZIR. La distance de visibilité peut être estimée par analyse d'images (détection de la zone chaussée, estimation de la hauteur de la route dans l'image et conversion en distance métrique). Cette distance de visibilité peut également être calculée par l'utilisation d'un modèle 3D de la route, obtenu par un véhicule instrumenté d'un laser rotatif vertical (CAOR, Mines Paris). Le logiciel "Qt-Ballad" (ERA 27, LRPC Strasbourg) utilisant ce modèle 3D, permet de calculer alors la visibilité offerte et de la comparer à celle requise.

Certains systèmes, issus de travaux de thèse et brevetés [16], utilisant le traitement d'images sont également en cours de transfert vers des matériels *mlpc*®, afin de caractériser la visibilité météorologique, en perception dégradée (vision dans le brouillard). D'autres systèmes permettant le relevé du patrimoine routier sont eux aussi en cours de réalisation. Ils reposent également sur des techniques de traitement d'images afin d'effectuer, par exemple, la détection des panneaux (LRPC Strasbourg).

En terme d'applications visées, cette thèse s'inscrit dans la même lignée. Toutefois, comme présenté précédemment, nous nous limitons à la saillance de la signalisation de police permanente de jour.

Chapitre 3

Vision, saillance de la signalisation verticale dans l'activité de conduite

Les modèles de la visibilité photométriques proposent la prise en compte d'un seul panneau sur un fond homogène [12]. Ceux-ci ne considèrent donc pas la complexité de l'environnement routier (fond complexe), ni le fait que le conducteur réalise une multitude de tâches et sous-tâches lors de la conduite. On parle alors de perception et de saillance visuelle des objets routiers, plutôt que de vision. A l'heure actuelle, aucun modèle, satisfaisant, de la saillance de la signalisation n'existe, du fait que les phénomènes et les caractéristiques de la saillance de la signalisation verticale en conduite ne sont pas totalement connus [11] : beaucoup de recherches sont encore à mener pour la compréhension des mécanismes perceptifs et cognitifs sous-jacent à la saillance, en particulier de la signalisation verticale. En effet, la question de savoir comment combiner et sélectionner les différentes caractéristiques de la saillance est toujours ouverte. La considération de l'activité de conduite en lien avec la saillance est très peu modélisée. Ceci est partiellement du à la connaissance limitée que l'on a du système visuel humain, ainsi qu'aux difficultés de mesurer l'attention des conducteurs, même avec les techniques oculométriques actuelles. Ce constat explique la difficulté de la modélisation de la saillance, que l'on se place du coté du domaine informatique ou du domaine des sciences humaines.

Selon [17], la conduite automobile est un type de locomotion, psychologique-

ment analogue à la marche, mais qui s'effectue au moyen d'un outil, comportant un niveau de risque. Ainsi, conduire un véhicule consiste à effectuer un déplacement au moyen d'un outil particulier, le véhicule, dans un environnement en perpétuelle évolution [18]. Ce déplacement, orienté vers un but, est soumis à différentes règles explicites ou implicites.

Comme le présente [19], la conduite d'une voiture (ou de tout véhicule routier) est une activité familière qui est réputée dangereuse, mais en même temps facile car ouverte à tous, hormis quelques exceptions médicales. Conduire est un acte qui peut paraître anodin pour le conducteur expérimenté, cette tâche devenant fortement automatisée. Pour autant, ce n'est pas le cas pour un conducteur novice. Dans les deux cas, le conducteur mets en jeu divers processus attentionnels et traitements cognitifs, afin de traduire les différentes prises d'informations effectuées (notamment visuelles) en des suites d'actions qui permettent d'assurer le contrôle de la vitesse, la planification et le suivi de la trajectoire, ainsi que le respect de la signalisation de police lui indiquant ses droits, interdictions et obligations.

3.1 Éléments de physiologie du système visuel humain

L'attention visuelle entretient des relations étroites avec la perception et le mouvement des yeux [20]. Ces mouvements oculaires peuvent être intentionnels ou réflexes. Ils ont pour but d'amener le centre de la rétine vers le stimulus prédétecté afin de procéder à son analyse et son identification avec plus de netteté et de résolution. D'où la nécessité, pour le conducteur, de pré-sélectionner des zones pertinentes de la scène routière (en fonction de la tâche) afin d'y fixer le regard et ainsi de placer la zone sélectionnée dans le champ visuel de la fovéa et par la même de permettre le traitement perceptif (identification et catégorisation) des objets dans cette zone [21].

Dans le système visuel humain, l'œil est le capteur qui permet le codage de l'information qui sera traitée par les différents constituants du cortex. Or, contrairement à un capteur d'image (ex : appareil photo), l'œil n'a pas une réponse spatiale uniforme. La conduite exige l'utilisation simultanée de deux zones de vision :

la vision centrale et la vision périphérique. Selon, [22, 23], on peut distinguer, comme il est présenté dans la Fig.3.1, quatres régions du champ visuel en fonction de l'angle d'ouverture par rapport au centre de la rétine :

- La vision fovéale, inférieur à 1° de demi angle visuel ;
- La vision para-fovéale qui complète la fovéale jusque 4° à 5° d'excentricité ;
- Vient ensuite la vision péri-fovéale qui d'étend jusque 9° ou 10° ;
- Au delà, et jusqu'à la limite du champs visuel humain (90° temporal pour chaque œil soit 180° au total), c'est la vision périphérique.

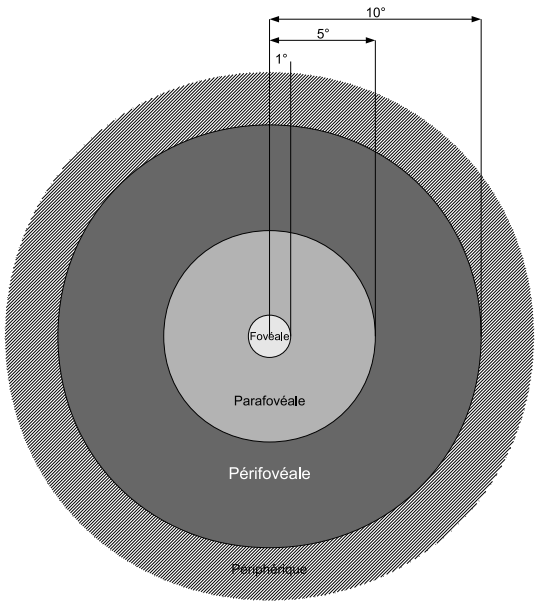


FIGURE 3.1 – Découpage spatial des différentes zones de vision

Dans le langage courant, la vision centrale correspond à la vison fovéale et para-fovéale tandis que la vision dite périphérique correspond elle à l'association de la vision péri-fovéale et du reste du champ visuel périphérique. Il est aussi commun de rencontrer la notion de proche-périphérie, qui regroupe les visions para-fovéale et péri-fovéale.

Comme le montre la Fig. 3.2, l'œil est un instrument optique de haute précision dont la fonction est de focaliser les flux lumineux sur la rétine, partie photosensible située sur son fond. La fovéa, au centre de la rétine, correspond à la zone d'acuité maximale, cf Fig. 3.3. On remarque également une zone appelée

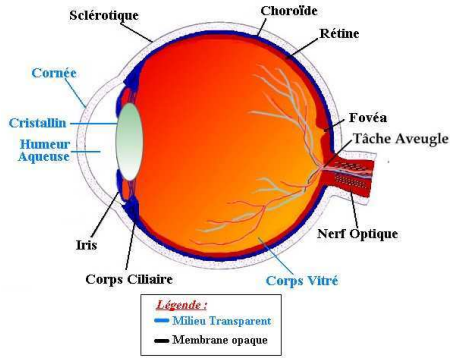


FIGURE 3.2 – Schéma de l'œil. Source http ://wiki.univ-paris5.fr/.

tâche sombre qui correspond à la zone de ramification du nerf optique sur la rétine. dans cette zones, vierge de tout photorécepteurs, la rétine est « aveugles ». Les photorécepteurs constituant la rétine sont de deux types et sont inégalement répartis :

- Les cônes, qui permettent la vision, en couleur, dans le domaine photopique (vision de jour). Chaque type de cône a une bande passante limitée, dans des longueurs d'ondes correspondant à ce qui est perçu comme des couleurs. Dans l'œil humain, il y a généralement 3 types de cônes, cf. Fig. 3.4 et 3.5, réagissant principalement aux couleurs :
  - jaune (560 nm), cônes L, communément appelés *cônes rouges* du fait que les longueurs d'ondes rouges activent ces cônes sans activer les cônes verts.
  - vert (530 nm), cônes M, communément appelés *cônes verts*.
  - bleu (424 nm), cônes S, communément appelés *cônes bleus*.

  Cette catégorisation a servi à la décomposition de la lumière en vidéo, et de manière complémentaire en impression numérique et en peinture. Les cônes sont principalement présents dans la zone centrale, sauf pour les cônes bleus qui sont absents de la fovéa, cf. Fig.3.5. Cette densité se réduit rapidement à mesure que l'on s'éloigne du centre de l'œil.
- Les bâtonnets, qui permettent la vision, en niveaux de gris, dans le domaine mésopique et scotopique (vision de nuit). Au contraire des cônes, la densité des bâtonnets est la plus forte en périphérie de la rétine et diminue à mesure

que l'on s'approche du centre de l'œil, cf. 3.5. Cette densité est quasi nulle au centre de la fovéa.

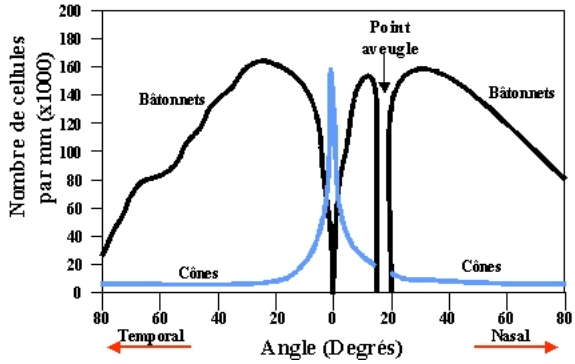


FIGURE 3.3 – Distribution des récepteurs dans la rétine. Source http ://www.rennes.supelec.fr/ren/perso/jweiss/tv/perception/percept1.html.

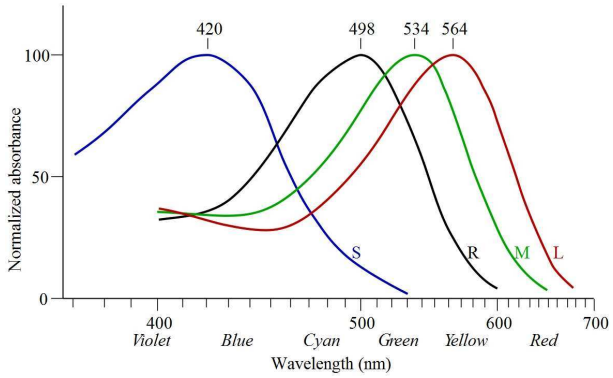


FIGURE 3.4 – Courbes d'absorbance des cônes et des bâtonnets pour l'Homme Source wikipédia..

Les bâtonnets captent la distribution des luminosités de la scène. Leurs performances, notamment en basse lumière, nous permettent de voir dans l'obscurité presque complète mais seulement en nuance de gris. C'est pour cette raison que l'on ne voit pas les couleurs quand la lumière n'est pas assez forte, les performances des cônes (seuls capables de percevoir les couleurs) étant beaucoup plus faibles que celles des bâtonnets pour les éclairages faibles. La fovéa, la zone centrale de la macula, est la zone de la rétine où la vision des détails est la plus

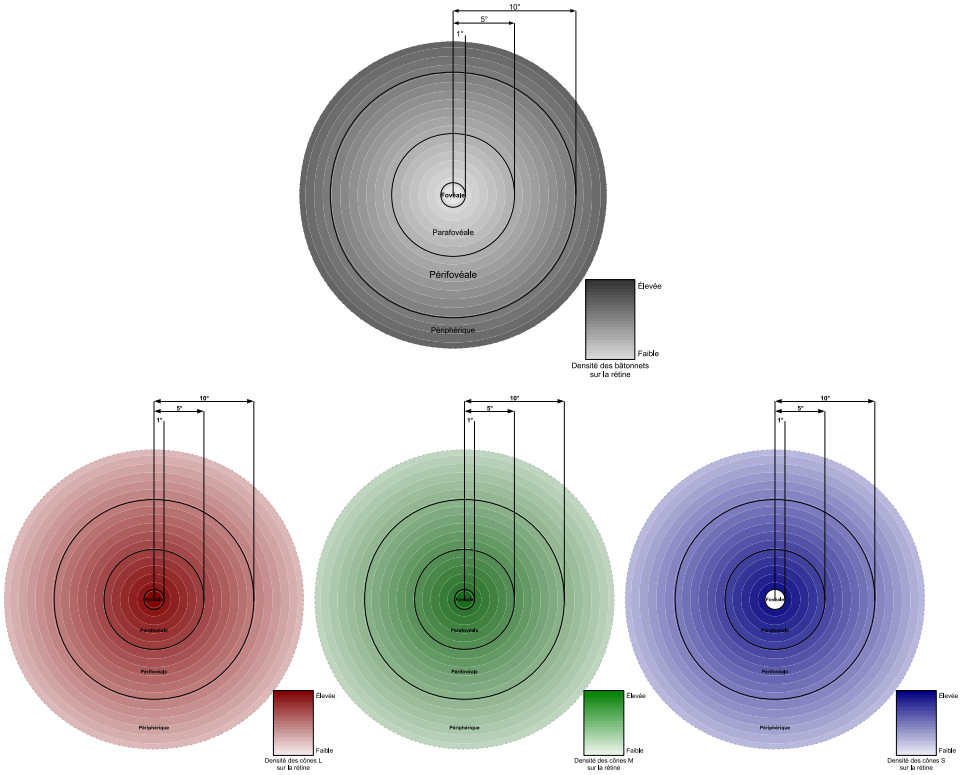


FIGURE 3.5 – Répartition des photorécepteur sur la rétine en fonction des zones de vison. En haut répartition des bâtonnets. En bas, répartition des trois types de cônes

précise. Elle est située dans le prolongement de l'axe optique de l'œil. La fovéa est peuplée quasi uniquement de cônes, les bâtonnets étant répartis sur la rétine périphérique, et c'est dans cette zone que la majeure partie de l'appréciation des couleurs se fait. Malgré ce que nous suggère notre perception, nous sommes donc quasiment "aveugles" aux couleurs hors de cette zone. Ce sont les mouvements de l'œil qui permettent d'avoir une impression globale de la couleur d'une scène. Naturellement, la vision humaine est floue et peu colorée en périphérie. Dans cette région le système visuel est surtout sensible aux mouvements grâce aux bâtonnets qui sont directement associés à leurs détection par le cortex visuel.

On comprend que même si la zone de la vision centrale (fovéale et parafovéale) où l'acuité est la meilleure, est restreinte, cf. 3.3, il est également possible

de percevoir (de manière floue) les constituants de l'espace périphérique. Ceci est à prendre en considération dans le traitement des données oculométriques, comme expliqué par la suite dans les parties 3.4.1 et 7.2.1.

Le système visuel est un système stéréoscopique. Les deux images ne se forment pas exactement sur des points correspondants, c'est la disparité rétinienne, qui est responsable de la vision du relief. Pour que la vision soit satisfaite, le système oculomoteur contrôle le mouvement perpétuel des yeux par six muscles.

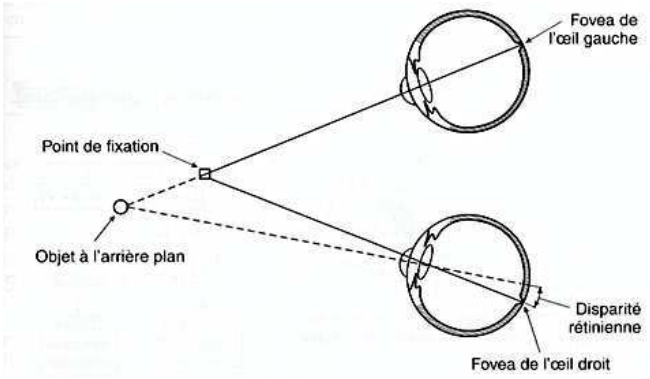


FIGURE 3.6 – Disparité rétinienne. Source http ://mon.univ-montp2.fr/ cours de Jean Valat Module de psychophysiologie.

Comme le montre la Fig. 3.7 qui décrit le processus itératif d'exploration visuelle, expliqué par Goldberg [24], l'activité oculaire que représente le chemin visuel est une succession de séquences saccades-fixations-saccades :

– Les saccades, sont le fruit de la vision pré-attentive. Une saccade est un mouvement rapide de l'œil qui le projette, de manière inconsciente, sur une nouvelle position dans la scène visuelle [25]. Elles placent dans la zone fovéale une zone précise au préalable détectée comme zone d'intérêt. Ce sont des mouvements balistiques (non interruptibles) très rapides, de l'ordre d'une dizaine de millisecondes par degré d'angle visuel. Un mécanisme correcteur, exprimé par la loi de Fitts [26] permet d'ajuster la destination par de petites saccades qui suivent une saccade importante. Les saccades sont automatisées, mais peuvent être considérées comme des mouvements

volontaires traduisant l'activité cognitive.

– Les fixations, s'intercalent entre les saccades. Elles sont la finalité de la vision pré-attentive qui devient alors attentive. A ce moment, les stimulus sont analysés dans la zone fovéale. Comme expliqué dans [27], la durée totale d'une fixation est de l'ordre de quelques centaines de millisecondes et au moins trois processus sont mis en action durant ce temps, cf. Fig. 3.7. Premièrement, l'information visuel est encodée. Quelques millisecondes plus tard, en parallèle, le champ périphérique de la fixation courante est capturé pour déterminer les zones intéressantes du point de vue de l'information saillante (recherchée ou attentionnelle). Enfin, dès le début de cette acquisition et analyse, la saccade suivante est préparée afin d'amener le centre du regard sur la fixation suivante. Une fixation n'est pas une réelle fixation au sens physique, car pendant celle-ci, des micro-saccades, très rapides, d'amplitude inférieure à un degré d'angle visuel et involontaires, appelées nystagmus, se produisent à fréquence constante. Ce mécanisme sert à réactiver les récepteurs visuels (cônes et bâtonnets) qui envoient l'image au cerveau. Sans ce mécanisme, une image fixée par le regard s'estompe progressivement.

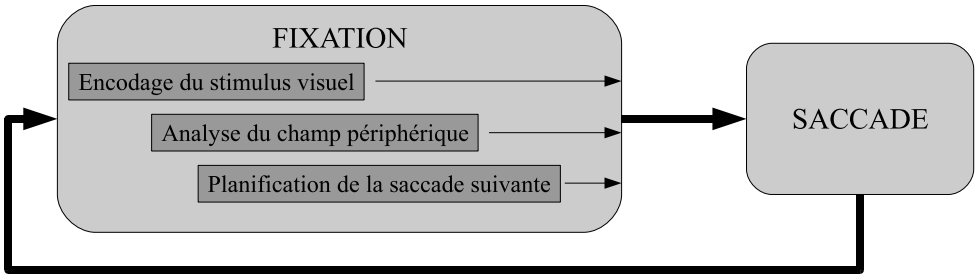


FIGURE 3.7 – Processus itératif d'exploration visuelle, selon Goldberg [24]

Afin d'avoir un flux constant d'images à traiter et garder en éveil le mécanisme de vision pré-attentive et de successions saccades-fixations, un autre mécanisme cérébral, assimilé à la persistence rétinienne, entre en action lors des clignements des yeux. En effet, l'information visuelle est alors supprimée pendant environ

$1/5^{ème}$ de seconde, durant cette période, le mécanisme cérébral inhibe la transmission de l'information visuelle qui est figée sur la dernière image.

3.2 Activité de conduite et prise d'information visuelle

Durant l'acte de conduite, 65 tâches principales et 1700 sous-tâches ont été identifiées par [28, 29]. Cette multitude de sous-tâches correspond aux actions à réaliser pour une situation définie [30], et se retrouve dans la description du « modèle hiérarchique de la tâche de conduite », modèle cognitif basé sur le recueil d'observations de l'activité de conduite, par [14].

L'activité de conduite est complexe et peu structurée [31, 32]. En effet, le conducteur doit traiter de nombreuses tâches diverses (action sur le véhicule et tâches annexes) alors qu'il ne dispose pas d'ordre précis et clair qui lui indique son champ des actions possibles et que l'environnement dans lequel il se déplace est, par nature, instable et variable (la dynamique des autres usagers comporte toujours une part d'incertitude). Ainsi, [33] décrit la conduite automobile comme « un contrôle de processus individualisé réalisé dans un environnement dynamique partagé » pendant lequel « le conducteur est amené à prélever en permanence de l'information dans son environnement afin d'acquérir une conscience de la situation pertinente pour satisfaire ses objectifs de performance et de sécurité ».

A ce jour, il n'existe pas de consensus sur un modèle complet du conducteur. Comme l'explique [34], les modèles proposés se focalisent sur une, ou un ensemble, restreint de composantes impliquées dans la conduite. Par ailleurs, lors de l'activité de conduite, le conducteur se repose surtout sur son système visuel [35]. Selon [7], et nombreux autres auteurs s'y référant, cette modalité sensorielle rend compte de 90% de l'information perçue, mais ce chiffre n'est pas vérifié, et l'on peut juste affirmer que la majeure partie des informations sensorielles utilisées en conduite sont visuelles. Toutefois, la modélisation des informations visuelles utilisées en conduite reste peu prises en compte par les modèles de conducteur qui incorporent souvent des composants visuels fondés sur des comportements idéalisés et dans certain cas des suppositions discutables [36]. Bien sûr, si la modéli-

sation des composants du système visuel du conducteur est incorrecte, le modèle complet est, dans une certaine mesure, compromis.

Selon [19], la perception et l'exploration visuelle, le traitement et la mémorisation des indices sont, par de nombreux aspects, liés à la vigilance et à l'attention. Cette dernière est, pendant la conduite, dirigée vers une source d'information ou la résolution d'un problème en fonction des représentations mentales activées par le conducteur. Elle est, par ailleurs, sujette aux interférences internes (préoccupations, distractions) et externes (lecture des instruments de bord, téléphone, GPS, ou autres tâches multiples tels que la conversation). Les recherches dans le domaine de l'exploration visuelle font donc intervenir des domaines divers.

Pour avancer dans notre problème, il est impératif de mettre en perspective différentes disciplines qui sont complémentaires, afin de comprendre et maîtriser les effets des caractéristiques photométriques des objets, des distances de visibilité, de la signification et la mémorisation des indices en fonction des contextes de conduite, de la symbolique et de la syntaxe de la signalisation ou de l'information embarquée. Dans un champ plus restreint, étudier la saillance de la signalisation verticale fait intervenir également plusieurs disciplines, telles que les sciences cognitives et informatiques.

Comme expliqué précédemment, selon [18], l'activité de conduite est une activité d'adaptation à un environnement en évolution continue. Elle exige alors le prélèvement et le traitement d'informations afin de prendre connaissance de l'état courant de l'environnement et du véhicule en vue de maintenir ce dernier dans une trajectoire et une vitesse compatible avec le contexte routier. Cet auteur sépare les activités d'exploration visuelle, d'identification, de prévision et de décision d'action. Ce formalisme a été repris récemment par [13] qui considère que le conducteur doit :

- sélectionner les informations issues de l'environnement routier afin de ne retenir que celles pertinentes pour la tâche de conduite (sous-tâche de détection et d'identification) ;
- comprendre la situation présente pour anticiper la situation futur ;
- prendre des décisions pour interagir de la manière la plus sûre et convenable, à travers les commandes de son véhicule, avec l'environnement routier et les autres usagers ;

– gérer ses propres ressources (physiques, perceptives et cognitives) afin de rester opérationnel dans la dynamique de la tâche de conduite.

Cet auteur insiste sur l'importance de la prise de l'information (sous-tâche de détection) et de la sélection de celle idoine à la conduite (sous-tâche d'identification).

Par ailleurs, la conduite est une tâche qui se déroule dans un environnement dynamique, c'est à dire en évolution continue, même en absence d'action de la part du conducteur. La conduite se définie comme une activité d'adaptation à l'environnement pour laquelle le conducteur doit mettre en oeuvre des traitements cognitifs variés. L'un des principaux est le mécanisme d'anticipation (activité de diagnostic) [37] que celle-ci soit consciente ou inconsciente. L'anticipation en conduite porte principalement sur les effets des actions effectuées par le conducteur, sur l'état futur du système Véhicule-Infrastructure-Conducteur, et sur les priorités à accorder (en fonction du temps) aux différentes tâches et sous-tâches de conduite. Le mécanisme anticipatif permet donc de prédire le comportement des autres usagers, les directions à prendre, les actions à effectuer, en respect avec l'analyse de la signalisation, indiquant l'ensemble des actions possibles et réalisables par les autres usagers et le conducteur lui même [38]. Sans prise d'information, ce mécanisme est naturellement inopérant.

Pour [18], en conduite automobile, les activités prévisionnelles peuvent ainsi être reliées aux activités perceptives, en l'occurrence visuelles. On distingue alors trois types d'activités :

– le recueil anticipé d'indices visuels (détection et sélection de l'information pertinente) ;
– l'anticipation des indices visuels (prédiction des futurs indices possibles et conditionnement de la recherche à effectuer) ;
– l'utilisation des indices pour la prévision d'événements futurs (manoeuvres possibles, dangers potentiels, ...).

Lors de l'activité de conduite, le prélèvement de l'information est donc un moment privilégié. C'est à la fois le point de départ d'une décision, par l'évaluation de l'état actuel du système dynamique Véhicule-Infrastructure-Conducteur, et le résultat (en terme de stratégie d'exploration visuelle) de l'état futur de ce système.

Ainsi, [39] rejoint ce formalisme, en dégageant deux fonctions principales de la vision égocentrée, qui fournit :

- des informations immédiates sur le véhicule lui même (position, direction, vitesse) ;
- des informations utiles à la prédiction, en particulier sur le déplacement du véhicule par rapport à l'environnement routier et aux autres usagers, ainsi que leurs déplacements propres.

La psychologie cognitive ne s'éloigne pas des formalismes présentées précédemment. Pour [40] la perception est conçue comme un traitement de l'information fournie par les sens. C'est un processus selon lequel le conducteur sélectionne, organise et interprète les données de l'environnement routier. Les informations portées pars les objets peuvent être plus ou moins saillantes dans la scène routière. Ceci peut rejoindre la théorie de la détection du signal qui prend en compte le rapport signal/bruit, [41, 42].

En accord avec les autres auteurs cognitivistes [43] découpe la perception en trois niveaux de tâches :

- La tâche de détection, qui correspond à la discrimination de l'information pertinente (dans notre cas la signalisation de police) par rapport au bruit de fond (dans notre cas l'environnement routier complexe et ses distracteurs). La détection dépend alors directement de la visibilité ou plus exactement de la saillance de l'objet vis à vis de son environnement proche.
- La tâche d'identification, qui correspond à la reconnaissance de l'objet détecté. Ceci par la comparaison des caractéristiques de l'objet détecté avec les représentations apprises, en mémoire et associées à un nom. Dans notre cas, ceci correspond à la phase de discrimination entre la certitude de reconnaître un panneau et de rejeter un distracteur au préalable détecté comme étant un panneau possible. L'identification dépend, quand à elle, de la saillance intrinsèque de l'objet, l'observateur étant fixé sur cet objet. La saillance intrinsèque se définit ici comme la propension d'un objet à être identifié, c'est à dire reconnu, lorsqu'il est le seul stimuli analysé, en dehors du monde environnant.
- La tâche de catégorisation sémantique, qui correspond à la reconnaissance de la classe, dans le réseau sémantique, à laquelle appartient l'objet identifié.

Cette approche, en trois niveaux de tâches perceptives, à été largement étudiée et utilisée dans le cadre du *paradigme de l'attention visuelle* [44].

3.3 Deux types de saillance en conduite

Les mécanismes attentionnels dépendent de la tâche effectuée, ce qui a été mis en évidence par différents paradigmes tels que le *paradigme de l'amorçage spatial* [45] ou le *paradigme de la recherche et de la détection visuelle* [46, 47].

Le *paradigme de la recherche et de la détection visuelle* montre que l'attention est « une ressource aux capacités limitées qui permet le traitement de l'information » [47]. Les travaux de [46], ont permis de démontrer quelques propriétés des processus contrôlés et automatiques de l'attention. Le traitement des processus contrôlés est sériel, puisque le conducteur traite un à un chacun des éléments constituant la zone de stimulus. Le traitement des processus automatiques est, quant à lui, parallèle, puisque toute l'information présente est traitée simultanément. Par ailleurs, l'attention peut être déclenchée automatiquement et capturée par une cible qui « saute aux yeux », on parle d'effet « pop-out ». Le guidage de l'attention visuelle peut être considérée comme une vision pré-attentive [48]. [49], décrit les caractéristiques attirantes qui régissent ces mouvements pré-attentifs, tels que les mouvements, les couleurs, la luminance, l'orientation, etc.

Le paradigme de l'amorçage spatial [45], met en avant que l'attention est « un contrôleur du système de traitement de l'information ». De ce paradigme découlent deux constats, concernant le « coût attentionnel » selon la région considérée et concernant « deux distinctions cognitives de l'attention ».

Le « coût attentionnel » concerne l'aspect spatial en lui même :

– Le fait de diriger l'attention (en vision fovéale) sur une zone a pour effet de renforcer perceptivement l'information contenue dans celle-ci. Les décisions (identification, catégorisation) sur l'information (objet détecté) sont alors plus efficaces et les actions correspondantes exécutées plus rapidement.

– Les régions voisines (para-fovéale et péri-fovéale) et périphériques sont, dans le même temps, inhibées. Le « coût attentionnel » équivaut à une complication des processus cognitifs mis en jeu. En effet, pour ne pas passer à coté d'autres informations dans la scène, l'attention doit être désengagée de la zone focalisée, pour pouvoir se déplacer vers une autre zone détectée et ensuite se réengager sur cette zone.

En conduite, le coût attentionnel, réside dans le fait que le conducteur doit analyser rapidement toutes les informations pertinentes. Ainsi, il doit démobiliser son attention d'un objet qu'il a détecté comme pertinent pour sa tache en cours si le traitement de l'information de celui-ci commence à prendre trop de temps. Dans ce cas, la stratégie adoptée est de ne pas identifier l'objet, au profit de la mobilisation de l'attention vers d'autres objets informatifs pertinents. La saillance intrinsèque est de ce fait liée au coût attentionnel.

Les « deux distinctions cognitives de l'attention » concernent une catégorisation des processus attentionnels selon qu'ils soient :

– objets de modulations involontaires, dépendantes du milieu extérieur. On parle alors d'attention exogène, de processus ascendant ou « bottom-up ». Cette catégorie d'attention est rapide, brève et capturée par l'arrivée inattendue d'un stimulus dans le champ visuel (distracteur).
– objets de modulations volontaires, dépendantes de l'individu et de sa tâche (ou plus précisément sous-tâche) en cours. On parle alors d'attention endogène, de processus descendant ou « top-down ». Cette catégorie d'attention est plus lente, dure plus longtemps mais permet, une fois engagée de ne plus prendre en compte les distracteurs.

Hughes et al. dans [8, 50, 51] ont démontré que la conduite se caractérise par l'utilisation en alternance de ces deux types de processus qui régissent deux modes d'acquisition de l'information visuelle dans l'environnement :

– L'« attention conspicuity », traduit ici par **saillance attentionnelle**, qui correspond à la modulation involontaire de l'attention. C'est l'information qui, du fait d'une saillance élevée, s'impose au conducteur et guide son attention en changeant ses priorités (par exemple, un signal lumineux de type warning ou l'allumage de feux stop) ;
– La « search conspicuity », traduit ici par **saillance de recherche**, qui

correspond à la modulation volontaire (qu'elle soit consciente ou fortement automatisée). C'est le conducteur qui recherche l'information pertinente (correspondante aux caractéristiques recherchées) en fonction de la tâche en cours (par exemple, recherche des indices pour prendre le chemin voulu).

On ne peut évaluer la proportion de chacun de ces deux types de saillance (de recherche ou attentionnelle) dans la perception de la signalisation verticale. Des expérimentations [52, 53], ont permis de récolter des indicateurs sur le fonctionnement de ces deux processus et de les mettre en confrontation avec le comportement visuel et la prise d'information dans des tâches de conduite simulée. Ces expérimentations n'ont montré aucune corrélation entre les modèles de saillance attentionnel (« bottom-up ») et l'activité de conduite. Pour autant, nous pensons que la prise en compte par le conducteur de la signalisation de police, en particulier des panneaux de police, mobilise dans une plus grande proportion la saillance de recherche (processus « top-down ») par rapport à la saillance attentionnelle (processus « bottom-up »).

3.3.1 Les modèles informatiques de la saillance attentionnelle

Un exemple de modèle informatique de saillance attentionnelle (« bottom-up ») est le modèle d'Itti [54]. Ce modèle, trouve ses origines dans les travaux de Koch et Ullman [55], qui ont proposé une architecture, biologiquement plausible, pour la modélisation de l'attention visuelle. Ils ont également introduit le concept de carte de saillance, qui exprime l'encodage spatial du degré d'intérêt de chaque pixel d'une image.

Comme le montre l'architecture générale du modèle d'Itti sur la Fig. 3.8, le principe de déroulement de ce modèle consiste à :

- Calculer quatre types de cartes de caractéristiques sur l'image :
  - contrastes rouge-vert,
  - contrastes bleu-jaune,
  - contrastes en luminance,
  - variation de l'angle du gradient.
- Calculer ces cartes à différentes échelles.

– Normaliser toutes les cartes sur la même échelle des ordonnées, cf . 3.9.
– Combiner les cartes de façon linéaire, avec une pondération différente pour chaque carte. La pondération est fonction de la présence d'un ou de quelques pics prépondérants dans la carte. La combinaison obtenue est nommée carte de saillance, même s'il n'y a pas de preuve de la présence d'une telle carte dans le système cognitif humain.
– Définir les zones de forte saillance comme des disques de taille fixe autour des maximums locaux de la carte de saillance.
– Utiliser la stratégie "Winner take all" pour ordonner les zones saillantes dans l'image.

En 2001, ce système a été modifié [56], afin de prendre en compte une combinaison pondérée des cartes de saillance, les poids étant issus d'un apprentissage pour spécialisé la carte selon quelques caractéristiques prépondérantes, cf. Fig. 3.8. Une normalisation itérative est également apparue, cf. Fig. 3.9. Le modèle final, est prouvé par l'auteur comme valide pour modéliser l'attention visuelle humaine.

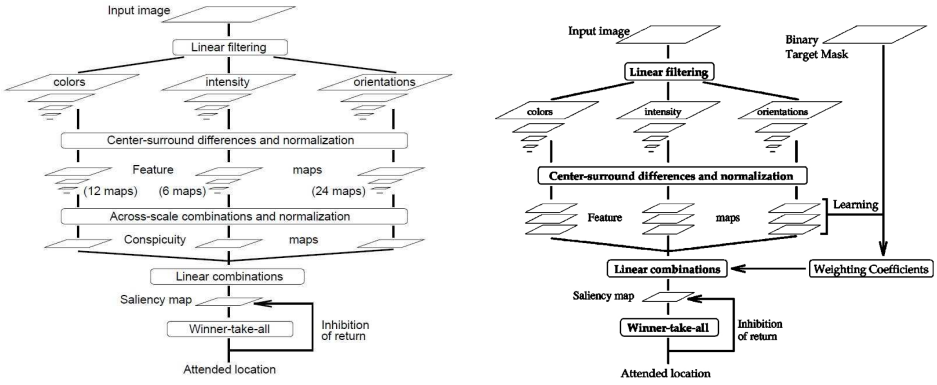


FIGURE 3.8 – Architecture générale du modèle d'Itti, à gauche version de 1998, à droite version de 2001. Source [54] et [56].

De nombreux autres modèles de la saillance attentionnelle ont été présentés [57, 58, 59, 60, 61], et bien d'autres encore. La plupart des modèles se fondent sur la théorie d'intégration des caractéristiques présenté par Treisman et al. dans [47]. Ces modèles ont été prouvé comme corrélés au comportement visuel hu-

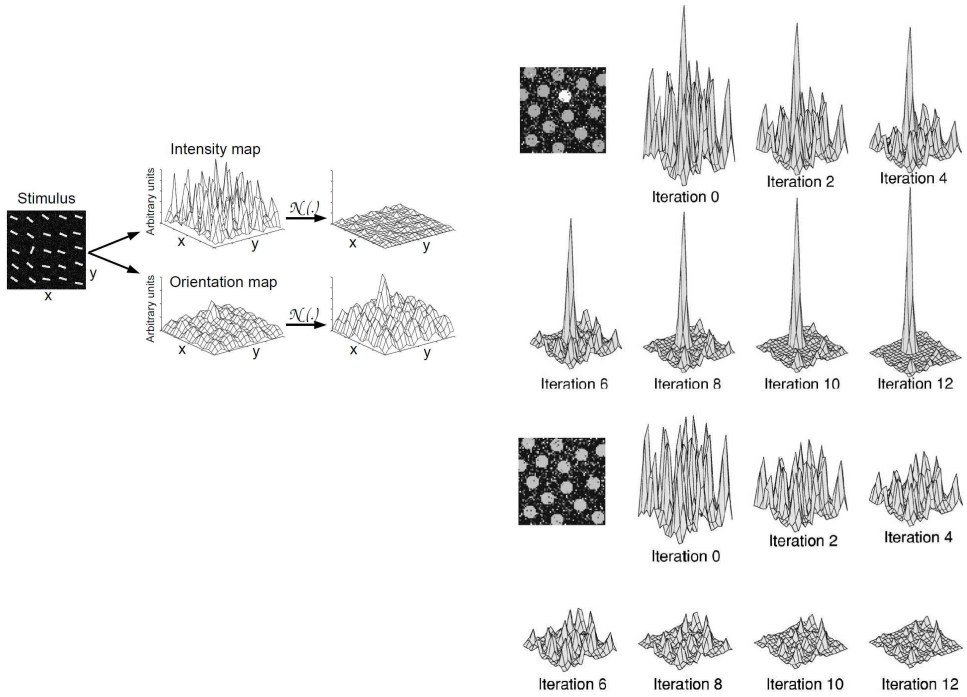


FIGURE 3.9 – L'opérateur de normalisation du modèle d'Itti, à gauche version de 1998, à droite version de 2001. Source [54] et [56].

main dans le cadre d'une exploration libre d'image [62] et dans le cadre de l'analyse de scène rapide [63]. Il est à noter une approche intéressante de modélisation spatio-temporelle de l'attention visuelle (modèle « bottom-up ») proposée par LeMeur, Le Callet, Barba et Thoreau [64, 65] dont les résultats validés par des expérimentations oculométriques sont bien meilleurs que ceux obtenus avec le modèle d'Itti sur les mêmes images. Ce modèle ce distingue notamment par l'utilisation de l'espace des couleurs de Krauskopf, qui est une représentation luminance-chrominance des couleurs proche de la vision humaine [66, 67]. Cet espace (A, Cr_1, Cr_2) est obtenu par une étape intermédiaire de transformation de l'espace *RGB* en *LMS* qui représente le signal issu de l'absorption des cônes.

Les modèles de type « bottom-up » sont conçus pour modéliser la saillance attentionnelle ou attention visuelle. Ce type de modèle, sur la base de celui d'Itti, modélise relativement le comportement visuel humain sur ces scènes courantes et lorsque l'exploration est totalement libre. En revanche, comme montré dans [68]

et [69] l'ajout d'un objet incongru, même peu saillant, dans la scène présentée déclenche des mécanismes cognitifs qui modifient le comportement humain et rend les modèles « bottom-up » inefficaces dans ce cas là. Or, le système visuel humain est rarement dirigé par la seule attention visuelle, un individu étant en permanence en cours de réalisation de tâches cognitives. Ainsi dans ce cas, [70, 71, 72] prouvent que la saillance attentionnelle pure est peu utilisée et que seul la tâche et le contexte déterminent les fixations visuelles, tout comme l'a montré Yarbus [73] en 1967.

3.3.2 Les modèles informatiques de saillance de recherche

Tout le problème est donc de modéliser l'influence de la tâche sur l'attention, notamment la tâche de recherche d'un objet. Pour ce faire, nombre d'auteurs prennent le formalisme de moduler les modèles « bottom-up » par les effets du contexte et des effets « top-down ». Ces effets ont été, à ce jour limité à la tâche de détection ou de reconnaissance d'objet. Ainsi, Navalpakkam et Itti [74], ont proposé un modèle de ce type. En partant de mots clés spécifiant la tâche à accomplir, la carte de saillance d'Itti est biaisée par les caractéristiques connus de l'objet correspondant recherché. Torralba et al. [75] considèrent également une tâche de détection d'objet et proposent une modulation spatiale de la carte de saillance par une carte additionnelle qui représente les positions probables de l'objet recherché dans la scène. Cette carte est obtenue par un apprentissage supervisé sur une base d'image dans lesquels les positions les plus probables de certains objets ont été pointées manuellement. D'autres modèles font de même en modelant la carte de saillance par des *a priori* sur le contexte ou les objets d'études [76] [77]. Toutefois, ces approches proposées restent essentiellement théoriques plutôt que pratiques et informatiques. Elles sont donc peu applicables à notre problème.

Des modèles informatiques de la propension de détection d'un objet donné ont été proposés dans les travaux de Gao et al [78, 79] et Witus [80] ou Wolfe [81]. Toutefois, ces travaux sont pour l'instant limités à des situations de laboratoire. Par exemple, un de ces modèles intéressant est celui de Gao. Ce modèle est basé sur la sélection des caractéristiques qui sont les plus discriminantes pour une reconnaissance de l'objet. Les parties de l'image qui contiennent la plus grande partie des

caractéristiques sélectionnées sont considérées comme saillantes. Ce modèle a été testé lors d'une expérimentation oculométrique où le sujet avait pour tâche de de reconnaître si un objet donné était présent dans l'image affichée. Mais cette expérimentation était en simple tâche, sur des images synthétiques et très simples. Ce modèle ne peut pas être appliqué à des situations compliquées où l'objet recherché peut avoir des apparences variables. De plus, la dépendance entre les caractéristiques n'est pas assez informative.

Une approche intéressante se retrouve dans les travaux de Chapuis, Chausse et al. [82, 83, 84] sur la reconnaissance d'objet focalisé capable de s'auto-adapter à l'instance de l'objet et qui défini une approche « top-down » comme processus de focalisation de l'attention dans l'espace des caractéristiques.

3.3.3 Vers une modélisation de la saillance de recherche en conduite pour la caractérisation de la saillance de la signalisation de police

La question reste ouverte sur la proportion de la mobilisation de l'un ou l'autre des deux processus présentés précédemment, selon la sous-tâche en cours lors de la conduite. Au sens ergonomique, les processus perceptifs sont déterminés à la fois par l'objectif de la tâche, la structure de la représentation externe et le contexte global dans lequel s'effectue l'activité cognitive [85]. Ce qui rejoint certains auteurs, [86], [87], [72] et [71, 70], selon lesquels, les processus de type « top-down » (saillance de recherche) priment sur ceux de type « bottom-up » (saillance attentionnelle), spécifiquement en conduite pour certains de ces auteurs. Ceci pourrait expliquer le fait qu'un conducteur s'engage à contre-sens dans une voie à sens unique, le panneau de sens interdit ayant été « vu mais pas regardé ». Cela peut s'expliquer par le fait que le panneau de sens interdit, n'ayant pas une saillance suffisante, n'a peut être pas été détecté et sûrement non identifié par le conducteur, d'où sa non-reconnaissance (cf. partie 3.2). Heureusement, dans certains cas, la non-reconnaissance des objets routiers n'induit pas systématiquement un accident, le conducteur et les autres usagers adaptant instantanément leurs comportements pour tenter de résoudre rapidement la situation pré-accidentogène. Ces

situations peuvent avoir une conséquence dramatique, il est donc nécessaire de les minimiser en comprenant et agissant sur la saillance de la signalisation verticale et plus précisément sur la signalisation de police.

Il a été montré par [88], dans une expérimentation où les sujets devaient lire toute la signalisation routière (panneau de police, de direction, feux, marquages confondus) que seulement 11.2% d'entre elle était fixée. Mais, comme présenté dans la partie 3.1, une fixation correspond à la vision fovéale (centre du regard). Elle ne prend pas en compte les objets vus, détectés et identifiés par la vision périphérique. Le taux de signalisation pris en compte doit être, en fait, bien plus important. [89] a, quant à lui, montré que les objets routiers étaient globalement bien vus et mémorisés, grâce à une expérimentation d'oculométrie cognitive en conduite sur un parcours de 50 km. Les sujets étaient questionnés sur les différents objets qu'ils ont rencontrés. En étudiant la relation entre les fixations visuelles et la perception « consciente » des sujets, cet auteur a noté que certains des objets fixés (vision fovéale, ie. centre du regard) n'ont pas été mémorisés, donc certainement non identifiés ou reconnus. De plus, certains objets vus et mémorisés l'ont été grâce à la vision périphérique (que l'oculomètre ne peut pas enregistrer cf.3.4.1). Cet auteur a également montré des différences selon le type d'objets mémorisés (signalisation, marquage, piétons, chevaux, ...). Les signaux de vitesse sont rapportés correctement à 78% par les conducteurs, 63% pour les signaux de police et 17% pour les passages piétons.

Les panneaux de direction, par exemple, n'ont pas besoin d'avoir une forte saillance attentionnelle, car ils ne sont pas nécessaires à tous moments lors de la tâche de prise d'information visuelle en conduite. Les conducteurs ont recours aux panneaux indicateurs de direction seulement lorsqu'ils s'orientent et ceci de manière active. Pour de tels panneaux, le conducteur considère alors volontairement cet objet routier, il mobilise donc des processus « top-down » réalisés consciemment et dans un temps correspondant au temps nécessaire à la prise de décision et au traitement des informations. Cette stratégie, répondant à une tâche de recherche, est une stratégie endogène de l'attention dans laquelle les mouvements oculaires sont volontaires, avec une mobilisation de l'attention vers la cible recherchée.

Par contre, pour un panneau de police (prescription, obligation ou interdic-

tion), la relation entre saillance attentionnelle et saillance de recherche est complexe. Les deux mécanismes entrent en jeux, mais nous pensons que le processus « top-down » est prépondérant sur le processus « bottom-up ». Les caractéristiques visuelles des panneaux de police sont choisies pour permettre une forte saillance attentionnelle (« bottom-up ») afin de faciliter la vision pré-attentive du conducteur (localisation de l'information). Toutefois, l'une des sous-tâches de conduite est la prise d'information à partir de la signalisation, afin d'en déduire le champ des actions possibles. Cette sous-tâche s'apparente à une tâche de recherche (« top-down »). En revanche, cette tâche est automatisée par les conducteurs dès lors qu'ils ne sont plus novices. On se place alors dans un caractère involontaire de cette « tâche de recherche », qui par son caractère automatique, est réalisée rapidement et inconsciemment. Nous formulons alors l'hypothèse que la saillance de la signalisation de police, objet de notre étude, correspond à un processus « top-down » de l'attention dans lequel les informations visuelles sont sélectionnées selon des caractéristiques connues (modèle mémoriel).

La revue de littérature présentée ci-dessus nous amène à proposer un algorithme de la saillance de recherche qui représentera le comportement du conducteur en réaction à la signalisation de police [90, 91]. A notre connaissance, il n'existe aucun modèle informatique de la saillance de recherche à ce jour. Notre démarche nous amène à vérifier la validité écologique de l'algorithme que nous proposons. Ainsi, nous vérifions la corrélation de notre modèle avec le comportement humain en situation écologique à la conduite, par des tâches (induites ou explicites) de focalisation sur les panneaux, au sein d'un contexte routier. Ce que nous avons réalisé, par le biais d'expérimentations psyco-cognitive qui utilisent un système oculométrique, cf. partie 5.6.2, et chapitre 7.

3.4 L'oculométrie, une méthode pour étudier la prise d'information visuelle en conduite

Les mouvements oculaires et l'attention étant reliés par les mêmes processus de sélection attentionnelle [92], les méthodes d'oculométrie sont majoritairement utilisées pour étudier les mécanismes de perception visuelle.

3.4.1 Fonctionnement et limite de l'oculométrie

L'oculométrie s'appuie sur le paradigme que les processus cognitifs et la vision, notamment les mouvements oculaires sont associés. Ainsi, selon [93], les fixations témoignent des traitements cognitifs tandis que les saccades relèvent des mécanismes perceptifs et oculomoteurs.

Le principe de l'oculométrie remonte à la fin du $19^{ème}$ siècle. Le premier système utilisant l'analyse du reflet cornéen a été développé par Dodge et Cline en 1901. Cette technique, utilisée par notre matériel [94], reste la plus répandue à présent, mais une autre technique, basée sur le traitement d'images [95], plus précise, car non soumise à la dérive du au larmoiement, fait son apparition et semble très prometteuse. Historiquement, on retrouve les premiers travaux utilisant l'oculométrie avec ceux de Fitts [96], en 1947, dans le domaine de l'aéronautique. La technologie se répand à partir du développement de la psychologie cognitive au début des années 1970, notamment par l'expérimentation marquante de Yarbus [97, 73], qui a mis en évidence les différences de trajectoire oculaire lors de l'exploration d'un tableau, selon que l'exploration est libre ou guidé par une problématique, cf. Fig. 3.10.

Depuis le début de l'oculométrie, différentes techniques ont été mises en oeuvre, celles-ci décrites dans [98, 99, 100, 101, 93, 102]. On retient principalement :

- *La technique de Huey* en 1897, considéré comme le premier système oculométrique. Le principe est de coller un amer sur le blanc de l'œil du sujet afin de le suivre. Naturellement très invasive, cette méthode ne fonctionne que pour les mouvement horizontaux. Le sujet est contraint à l'immobilité.
- *La technique du Limbe*, développée par Torok en 1951, dont le principe est d'éclairer le limbe (séparation entre la sclère, c'est à dire le blanc de l'œil, et l'iris, la partie sombre), la quantité de lumière réfléchie dépend de la surface relative de la sclère et de l'iris dans le champ de mesure, et donc permet d'identifier la position de l'œil. Cette technique a suscité plusieurs applications dans le domaine du handicap. Cette technique peu coûteuse (simple source de lumière et détecteur d'intensité lumineuse) peut être fixé sur une monture. Or, elle nécessite l'alignement du capteur par rapport au globe

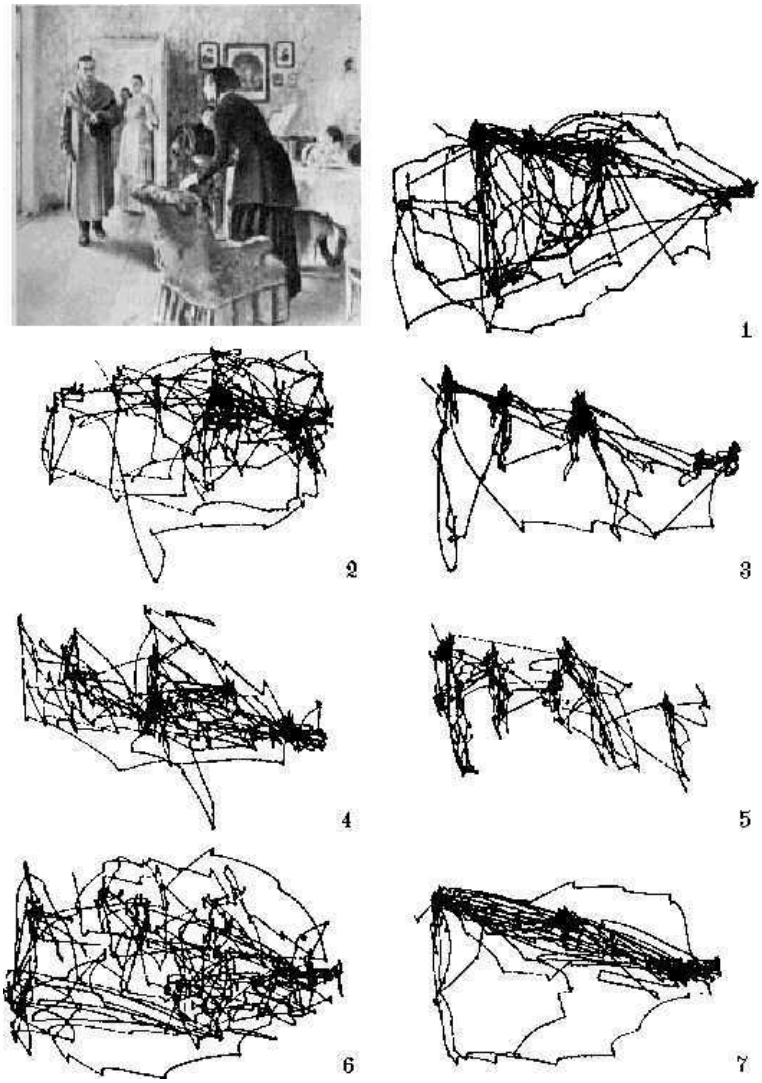


FIGURE 3.10 – Expérimentation de Yarbus, 1967. Sept enregistrements de chemins visuels de trois minutes, pour le même sujet, en fonction de la tâche demandée. 1- Exploration libre. Avant les autres enregistrements, le sujet devait : 2- Évaluer la situation matériel de la famille ; 3- Donner l'âge des personnes ; 4- Prévoir ce que la famille avait fait avant l'arrivée du « Visiteur inattendu » ; 5- Mémoriser les vêtements portés par les personnes ; 6- Mémoriser la position des personnes et des objets dans la chambre ; 7- Estimer combien de temps « Le visiteur inattendu » a été loin de la famille. Source [97, 73].

oculaire, ce qui n'est pas toujours évident ; les mouvements de la tête perturbent la mesure ; les paupières supérieures perturbent la fiabilité de la mesure des mouvement verticaux, seuls les mouvements horizontaux restent fiables ; enfin, cette technique est particulièrement sensible aux variations brutales de luminosité et fonctionne mal avec des yeux très clairs.



FIGURE 3.11 – Exemple d'oculomètre utilisant la technique du limbe. Source [101].

– *La technique galvanométrique*, qui consiste à mesurer les variations d'intensité d'un signal électrique généré par une lentille de contact équipée d'une bobine d'induction, le sujet étant placé dans un champ électromagnétique, cf ; Fig. 3.12. Cette technique est très précise, mais elle est aussi très contraignante et invasive pour le sujet.

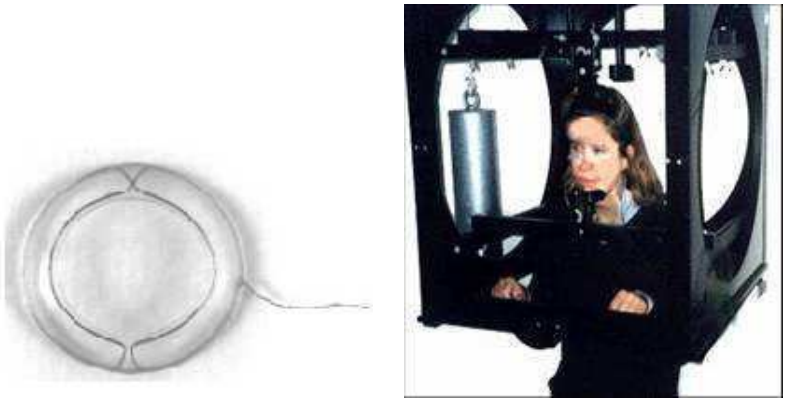


FIGURE 3.12 – Exemple d'oculomètre utilisant la technique de la galvanométrie. Source [93].

– *La technique électro-oculographique*, une des plus ancienne technique dont le principe, mis au point par Marg en 1951, est de mesurer les différences

de potentiels bioélectriques induits par la rotation oculaire (muscle oculomoteur) par des électrodes disposées autour de l'œil, cf. Fig ; 3.13. Cette technique, invasive du fait des électrodes, a une bonne résolution temporelle (pas de dérive de la mesure) mais une faible résolution spatiale. De plus, la mesure des mouvements verticaux est peu fiable et la tête doit être maintenue immobile pour avoir une indication fiable de la direction du regard.



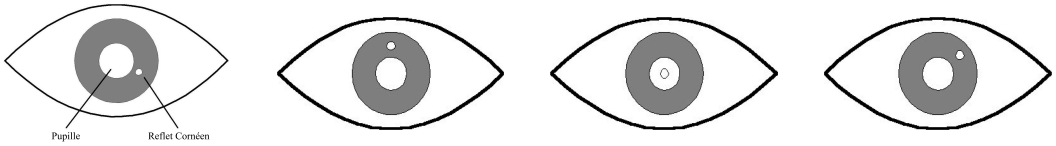
FIGURE 3.13 – Exemple d'oculomètre utilisant la technique de l'électro-oculographie. Source [93].

– *La technique du reflet cornéen* est basée sur le principe de Hirschberg (strabologue de la fin du $19^{ème}$ siècle) qui a montré que l'on peut déterminer l'orientation du regard en repérant la position du reflet d'une source de lumière sur la cornée de l'œil (reflet cornéen) par rapport à la pupille : une caméra infrarouge détecte ainsi le mouvement de l'œil explorant une image. Depuis, la source de lumière est remplacé par une ou deux sources infrarouge mais le principe reste le même. La position du regard est déterminé par la position relative de centre de la pupille par rapport au reflet cornéen cf. Fig. 3.14.
  Une analyse quantitative du mouvement oculaire peut ensuite être effectuée (nombre et durée des fixations du regard, nombre et amplitude des saccades, etc.). Cette méthode permet des mesures absolues des différentes positions de l'œil indépendamment des mouvements de la tête.
  Cette technique nécessite un calibrage du sujet par la fixation de points préa-

lablement choisis. Selon les fabricants, ont trouve des systèmes déportés (constitués d'un écran, de diodes électroluminescentes et de caméra ou de type « monté sur la tête » sous la forme de casque ou lunette. Pour les systèmes déportés, il existe deux classes. Les systèmes dit 2D permettent seulement l'analyse du regard en tolérant de légers mouvements de tête (il ne faut pas que les yeux sortent du champ de la caméra). Les systèmes dit 3D permettent d'obtenir, en plus de la direction du regard, la direction de le tête. Toutefois, ceci nécessite un calibrage de la géométrie du système et un calibrage plus long pour le sujet (fabrication d'un modèle). Les systèmes « tête porté » permettent une plus grande liberté de mouvement, mais sont plus invasifs de par leur poids.

Globalement, pour tous ces systèmes, la technique étant assez précise, on obtient une précision du regard de l'ordre de $0,5°$ à $1°$. La position du regard est couramment échantillonné à 50 ou 60Hz (de 20ms à 17ms). Certains systèmes plus rapide fonctionnent à des fréquences de 250Hz à 1000Hz. Toutefois, ces systèmes nécessitent beaucoup de calculs et donc une machine puissante, c'est pourquoi dans certains cas, il est utilisé deux machines, une pour les calculs et l'autre pour l'environnement informatique. Le principal défaut de ces systèmes réside dans la perte du suivi du regard si l'œil sort du champ de la caméra. De plus, la fiabilité du système dépend des variations lumineuses, ce qui rend les systèmes embarqués plus ou moins fonctionnels. Pour les sujets porteurs de verres correcteurs, il peut y avoir des reflets parasites ; c'est notamment le cas des montures de lunettes métallisées. Il peut également y avoir de nombreux dysfonctionnements avec les lunettes, de par les traitements des verres (anti-reflets, etc.). Il est parfois nécessaire pour les personnes porteuses de lentilles de contact ou les personnes ayant une tendance à avoir les yeux secs, d'appliquer des larmes artificielles régulièrement.

– *La technique de traitement d'image* est une technique utilisé par l'oculomètre de la société Pertech [95]. On ne connaît pas le principe utilisé : Ce système, EyeTechSensor, cf Fig. 3.15 n'utilise pas le reflet cornéen. Il est robuste aux changements d'illumination par l'utilisation d'une source infra rouge de faible puissance. Il n'occasionne pas de larmoiement, ce qui per-



Disposition relative du reflet cornéen et de la pupille | Le sujet regarde en bas | Le sujet regarde au centre | Le sujet regarde en bas à gauche

FIGURE 3.14 – Principe du reflet cornéen. Source [103].

met de ne pas avoir de dérive de la mesure. Sa précision est de $0,25^\circ$ pour un échantillonnage à 25Hz. Ce système fixé sur un support de type lunette ou bandeau, pèse moins de 70 grammes et est non invasif.



EyeTechSensor type L | EyeTechSensor type B

FIGURE 3.15 – Oculomètre de la société Pertech. Source [95].

3.4.2 Les indicateurs du comportement visuel humain utilisés en oculométrie cognitive

La pertinence de l'oculométrie a été démontré par de nombreux travaux. Or l'oculomètre ne fournit que des données brutes, qu'il faut par la suite analyser

pour en retirer des indicateurs pertinents par rapport aux objectifs de recherche. Le choix de ces indicateurs est donc primordial.

Les données brutes produites par un oculomètre sont les coordonnées des positions successives du centre du regard (fovéale) selon un pas de discrétisation et une précision donnée. Généralement, le « timecode », c'est-à-dire l'index de temps, accompagne chaque donnée. Il est également possible, pour certain système, d'obtenir des informations sur le diamètre de la pupille. Les fixations sont définies par le regroupement de positions de regard proche. Selon le système, il existe plusieurs méthodes pour déterminer les fixations [104] selon les considérations spatiales et / ou temporelles prises en compte. Dans le cas le plus courant, les systèmes actuels définissent les fixations selon deux critères, cf. Fig.3.16 :

- *La durée minimum* requise pour considérer la fixation, qui est d'environ 100ms, soit 5 points de regard pour un oculomètre à 50Hz.
- *La dispersion maximum* qui s'exprime en degrés ou en pixels. Elle se calcule en fonction des contraintes physique de l'expérimentation. Par exemple, si l'on veut prendre en compte les fixations qui correspondent au champ de la fovéale (2°), le sujet étant à une distance de 60cm de l'écran (de résolution 1280×1024 pour une largeur de 37 cm), on obtient une dispersion maximale de 64 pixels. Ce seuil D_{max} est par la suite utilisé pour considérer les points de regard comme appartenant à la fixation en cours, du moment que la dispersion D_{fix} de l'ensemble de ses points respecte la contrainte $D_{fix} \leq D_{max}$. D_{fix} est défini par la relation, cf. Fig.3.16 :

$$D_{fix} = (X_{max} - X_{min}) + (Y_{max} - Y_{min}) \tag{3.1}$$

Avant de déterminer les fixations, il est généralement réalisé un filtrage des données brutes afin de supprimer les points aberrants. Ces points peuvent résulter d'une perte de suivi du regard, d'une erreur due à un reflet (lunettes, etc.), ou d'un clignement des yeux. Le principe, cf. Fig. 3.17 est d'enlever les points qui ne sont pas logiques, dus à une trop grande vélocité de la saccade, ou à un éloignement trop important par rapport aux autres points. Ce point est alors remplacé par la position moyenne entre le point qui le précède et celui qui le suit.

L'ensemble des données oculométriques brutes (séquence de points de regard) est ainsi analysé afin de révéler le chemin visuel (« scanpath ») qui est constitué

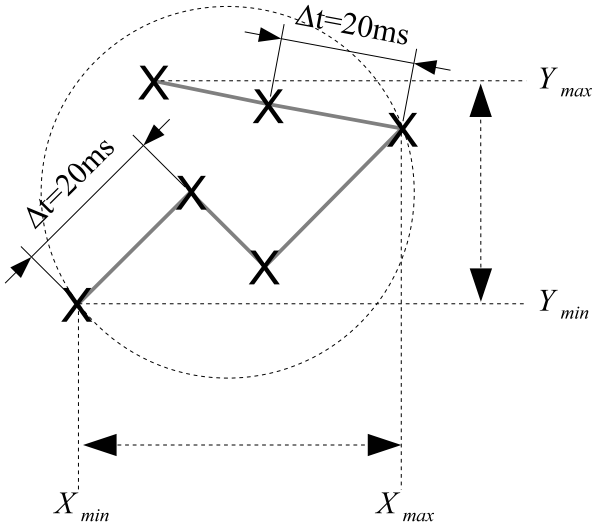


FIGURE 3.16 – Critère de dispersion d'une fixation.

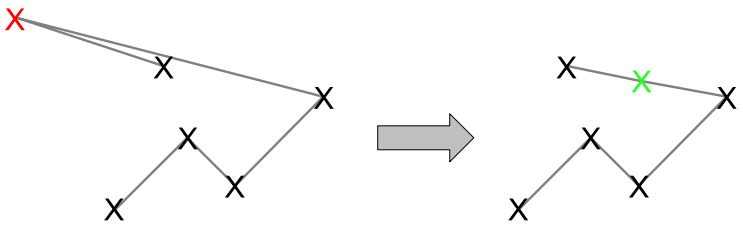


FIGURE 3.17 – Filtrage d'un point aberrant.

de la succession de saccades et de fixations, cf. Fig. 3.18.

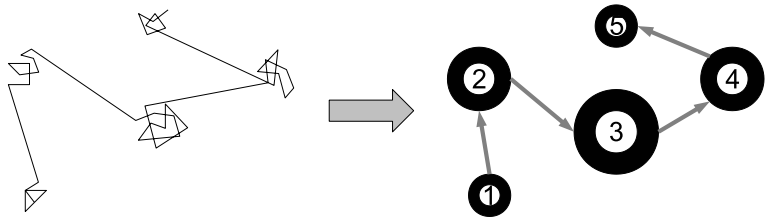


FIGURE 3.18 – Chemin visuel typique : Après analyse des données oculométriques brutes, à gauche, les fixations (cercles noirs) et les saccades (flèches grises) sont mises en évidence, à droite.

Une fois les fixations et les saccades identifiées, on peut alors effectuer des statistiques sur celles-ci afin de mesurer des indicateurs qui seront des variables dépendantes pertinentes pour l'expérimentation. Dans la littérature, une multitude d'indicateurs ont été utilisés, [105] et [103] présente un état de l'art en s'appuyant sur celui de [24]. On peut résumer les différents indicateurs en les classant comme dans le tableaux 3.1.

En règle générale, ces indicateurs sont mis en relation avec des zones définies appelées zones d'intérêts (en anglais « *AOI, Area Of Interest* »).

Processus Cognitif	**Métrique Oculaire**	**Définition de la métrique**	**Notion mesurée**	**Unité**
Recherche de l'information	Longueur du chemin visuel	Somme des distances entre les points de regards	Efficacité de l'exploration	*Pixels*
	Durée du chemin visuel	Temps passé depuis le début de l'exploration	Efficacité de l'exploration	*ms*
	Densité spatiale du chemin visuel	Dispersion des points du chemin visuel dans une grille	Efficacité de la recherche	%
	Matrice de transition	Fréquence de transition entre les zones d'intérêts	Inefficacité de la recherche ; stratégie d'inspection	%
	Recouvrement spatiale par l'aire convexe de Hull	Surface du plus petit ensemble convexe contenant le chemin visuel	Niveau de dispersion de la recherche (à normaliser par la longueur)	$Pixels^2$
	Nombres de saccades	Nombres de saccades d'amplitude supérieur à un seuil défini	Volume de recherche et organisation de la recherche	*N*
	Amplitudes des saccades	Amplitude angulaire	Efficacité du ciblage des recherches	*Degrés*
	Taille moyenne des saccades	Moyenne des distances entre fixations	Étendue de l'exploration	*Pixels*
Traitement de l'information	Durée du chemin visuel	Temps passé depuis le début de l'exploration	Complexité du traitement des informations	*ms*
	Durée des fixations	Durée de chaque fixation	Facilité de compréhension intrinsèque de l'élément fixé	*ms*
	Nombre de fixations	Nombre de fixations	Nombre d'éléments traités	*N*
	Ratio fixation/saccade	Rapport entre la durée de fixation et la durée de la saccade précédente	Durée du traitement de l'information proportionnellement à la durée de sa localisation	–

TABLE 3.1 – Tableau récapitulatif des principaux indicateurs utilisés en oculométrie en fonction des processus cognitif étudiés.

Chapitre 4

Analyse d'images pour la détection de la signalisation verticale

Les techniques d'analyse d'images sont couramment utilisées pour effectuer la detection et la reconnaissance automatique de la signalisation verticale. Outre l'inventaire et la maintenance du réseau routier, la reconnaissance des panneaux permet, en lien avec la robotique, de concevoir des véhicules automatisés ; ou bien de réaliser des aides à la conduite (indication en temps réel de certains panneaux).

4.1 État de l'art sur la détection de panneaux, contraintes spécifiques liées à notre problème

La détection de la signalisation verticale, comme sa reconnaissance et sa classification est toujours un problème d'intérêt majeur. Ce type d'applications se retrouve souvent dans bon nombre de conférences et travaux, que celles ci soient axées sur le domaine routier, comme Intelligent Vehicle Symposium (par exemple [106, 107, 108, 109, 110, 111, 112, 113, 114]) ou dans le domaine du traitements d'images et de la vision par ordinateur. Cette multitude de publications dans diverses conférences ou revues prouve l'importance et l'intérêts de cette problématique. Toutefois, à ce jour, peu de travaux ont présenté un algorithme complet pour la détection et la reconnaissance de la signalisation verticale. Lorsque l'un d'eux atteint un taux de reconnaissance et de fausse détection acceptable, c'est souvent

au prix d'un temps de calcul rendant impossible le traitement de grandes bases d'images. A cela s'ajoute aussi la complexité induite par la détection et la reconnaissance de plusieurs (voir tous) types de panneaux. En effets, comme présenté dans la partie 2.1, ils sont nombreux, de formes et de couleurs variés, ce qui rend difficile la généralisation d'un modèle et requiers alors l'utilisation de fusion ou de cascades de classificateurs.

Or, une application comme la notre nécessite que l'algorithme de détection respecte quelques contraintes. En temps qu'outil de diagnostic, même si le traitement et l'analyse n'ont pas besoin d'être effectués en temps réel, le temps de calcul de l'estimation de la saillance doit rester raisonnable, et par voie de conséquence, la détection en amont se doit d'être rapide. D'autre part, le taux de détection doit être aussi bon que possible (on ne doit pas rater de panneaux) et le taux de fausse détection doit rester assez bas. Ce dernier point ne constitue pas un objectif principal, car les fausses détections seront rejetées, dans notre cas, lors de l'analyse de la saillance et de l'exploitation des résultats par le gestionnaire routier. Atteindre ces taux implique la résolution de quelques difficultés. Par exemple, la visibilité d'un panneau peu diminuer à cause de la pollution de l'air, des conditions climatiques (pluie, brouillard, nuage, soleil). Les conditions de lumières (ombres, réflexions spéculaire, ...), variant du jour à la nuit, affectent également l'apparence de la signalisation verticale. D'autres facteurs comme la vétusté ont aussi un impact sur la saillance d'un panneau, la peinture peut peler ou s'écailler avec le temps, et en raison de l'exposition au soleil et aux intempéries, les couleurs peuvent passer. Par ailleurs, selon leurs implantations, les panneaux peuvent apparaître partiellement occultés par des obstacles comme des arbres, des piétons, d'autres véhicules ou d'autres panneaux. Enfin, un algorithme de détection fonctionnant à partir de séquences d'images acquises à partir d'un véhicule doit être robuste au flou de bougé dû aux vibrations et aux mouvements du véhicule. Deux publications, [115] et [116], présentent un état de l'art et les problèmes liés à la détection de panneaux. Ces deux travaux tentent de résoudre ces différents problèmes et explorent quelques approches fondées sur différentes techniques de reconnaissance et de détection. Pour la détection et la reconnaissance de la signalisation verticale, on peut distinguer deux grandes classes de techniques, la détection par des méthodes dites classiques de traitement d'image et les méthodes de

détection via la reconnaissance d'objet. Dans ces deux catégories, on trouve alors différentes approches, certaines utilisent les informations de contours, d'autres ont une approche couleurs et d'autres utilisent ces deux informations. Enfin certaines, y ajoutent l'information de mouvement. Dans les techniques les plus utilisées pour la détection, on retrouve les algorithmes génétiques, l'indexation de couleur, les transformations de distance, les méthodes morphologique, les réseaux de neurones, la logique floue, le seuillage de couleur, les espaces de couleur, la détection de région, l'analyse de forme et bien d'autres.

4.2 Apprentissage par SVM

Les SVMs sont de plus en plus utilisés en vision par ordinateur et en traitement d'images pour résoudre des problèmes de reconnaissance et de classification. Le principe du SVM est de séparer les données de façon à maximiser la marge entre deux classes d'exemples représentant les données, les exemples positifs et ceux négatifs [117]. La base d'apprentissage est alors constituée de couple $\{(x_1,y_1),\ldots,(x_l,y_l)\}$ où x_i est le vecteur de $\mathbb{R}^n$ représentant l'exemple i, et où $y_i \in \{-1,+1\}$ est son label.

Lorsque les exemples positifs et négatifs sont linéairement séparables, l'algorithme SVM estime les paramètres $w \in \mathbb{R}^n$ et $b \in \mathbb{R}$ de l'hyperplan $w \cdot x + b = 0$ qui les séparent (voir Fig. 4.1). On défini les vecteurs support (SV) comme les exemples les plus proches de l'hyperplan qui sont seuls utiles à caractériser cet hyperplan (voir Fig. 4.1).

Dans le cas où les exemples ne sont pas linéairement séparables (voir Fig. 4.2), une fonction mathématique $k(x,x')$, appelée noyau, est utilisée afin de projeter implicitement les exemples dans un espace de plus grande dimension où le problème devient linéairement séparable. Toutefois, pour ce faire, le noyau choisi doit remplir certaines conditions mathématiques. Notamment, il doit remplir la condition de Mercer, c'est à dire être défini positif. Ce point est développé lors de l'étude du choix d'un noyau spécifique, cf partie 5.5.4 ainsi que dans une des publications de cette thèse [118] où un noyau mis au point à partir de la distance d'Hausdorff est utilisé pour faire de la comparaison de forme.

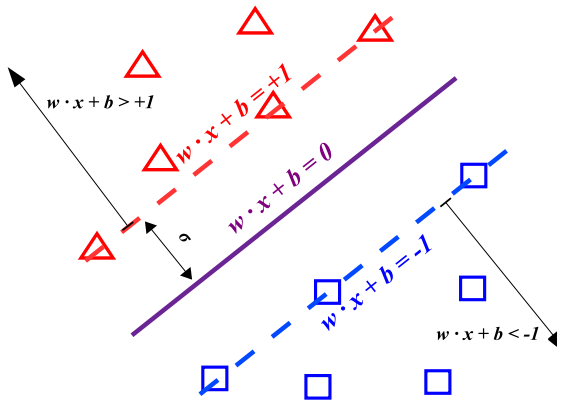


FIGURE 4.1 – Classification linéairement séparable entre les △ (exemples de classe positive) et les □ (exemples de classe négative). L'hyperplan séparateur est défini par $w \cdot x + b = 0$. La décision est donnée par le signe de $w \cdot x + b$. Les exemples situés sur les deux lignes discontinues sont des vecteurs support (SV). La marge σ est la distance minimale des SV à l'hyperplan.

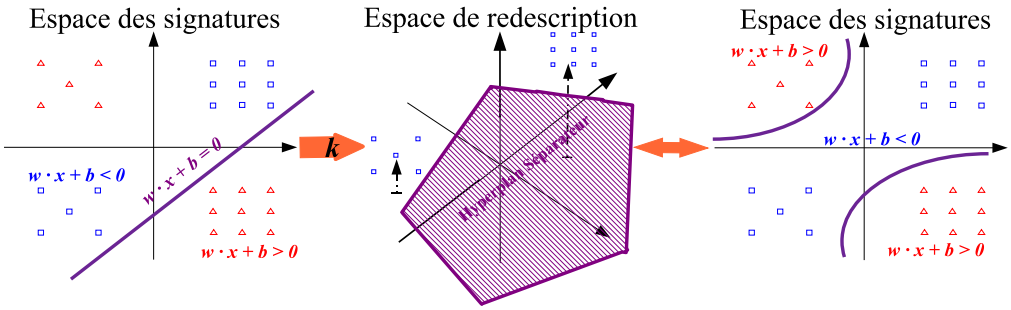


FIGURE 4.2 – Les données 2D sur la gauche n'étant pas linéairement séparables, elles sont projetées dans $\mathbb{R}^3$, par l'utilisation d'un noyau k. Le SVM sépare alors, par un hyperplan, les exemples positifs et négatifs projetés dans $\mathbb{R}^3$.

D'autre part, on introduit un terme de régularisation, paramétré par C, qui autorise des erreurs de classification lors de l'apprentissage. Ceci est utile lorsque

des exemples d'apprentissage ont des labels erronés.
En pratique, l'algorithme SVM consiste à minimiser le problème quadratique suivant par rapport aux α_i, $1 \leq i \leq l$:

$$W(\alpha) = -\sum_{i=1}^{l} \alpha_i + \frac{1}{2} \sum_{i,j=1}^{l} \alpha_i \alpha_j y_i y_j k(x_i, xj) \tag{4.1}$$

sous les contraintes $\sum_{i=1}^{l} \alpha_i y_j = 0$ et $0 \leq \alpha_i \leq C$.
Après le calcul des α_i, le b est calculé [117].

La fonction de décision du SVM est donnée par :

$$f(x) = sgn(\sum_{i=1}^{l} \alpha_i y_j K(x_i, x) + b) \tag{4.2}$$

Toutefois, on peut moduler cette fonction de décision en prenant un paramètre de seuil σ sur la fonction de decision. Cela revient à poser une fonction de décision telle que :

Soit $d(x) = \sum_{i=1}^{l} \alpha_i y_j K(x_i, x) + b$ la valeur de décision du SVM pour l'exemple de test x.

- Si $d(x) \geq \sigma$ Alors, x est de classe positive.
- Si $d(x) < \sigma$ Alors, x est de classe négative.

Dans le cas le plus courant ($\sigma = 0$), la frontière de séparation des deux classes se trouvant pour $d(x) = 0$ et les supports vecteurs négatifs et positifs respectivement pour $d(x) = -1$ et $d(x) = +1$. Plus la distance $d(x)$ est élevée, plus la confiance de l'algorithme en la classification est élevée. Il est d'usage de définir pour $-1 \leq d(x) \leq 1$ la zone d'indécision du SVM. L'exemple se trouvant dans cette zone étant hors de la marge définie par les supports vecteurs, la confiance de l'algorithme en sa décision se trouve réduite.

Chapitre 5

Un SVM pour détecter les zones saillantes cohérentes à la signalisation verticale

Afin d'estimer la saillance de la signalisation verticale dans les images routières, il est nécessaire auparavant, que celle-ci soit détectée, comme expliqué dans le chapitre 2. Notre approche se fonde sur la réalisation de deux actions (détection et évaluation), à partir du même algorithme, et ce afin de minimiser les temps de calcul. Nous nous sommes donc naturellement orientés vers les méthodes d'apprentissage, ceci dans le but de concevoir un détecteur de panneaux capable de donner une estimation de la qualité de sa performance de détection.

5.1 Choix des SVMs

Le but du détecteur est de sélectionner les parties de l'image qui peuvent être des « panneaux », afin de pouvoir, par la suite, évaluer ces zones au sens de la saillance de la signalisation verticale.

Or, un panneau, selon sa position dans l'image et dans la scène routière, peut avoir des caractéristiques géométriques relativement différentes. De plus, même si c'est un objet bien normalisé, il peut prendre des aspects assez variés à cause des variations d'éclairage, des conditions météorologiques et de l'usure, comme

expliqué dans la partie 4.1. Il nous faut donc élaborer un système qui soit invariant à la rotation, à l'échelle et à l'illumination tout en restant robuste aux occlusions, déformations et détériorations des panneaux. Les variations de l'objet que nous cherchons à détecter nous ont conduit vers les méthodes d'apprentissage.

L'objectif étant l'évaluation de la saillance des panneaux, notre détecteur doit avoir un taux de détections manquées le plus faible possible, même si les panneaux sont partiellement cachés. De plus, pour que l'outil de diagnostic soit rapide et efficace, le taux de fausses détections doit être peu élevé, afin que l'évaluation se porte majoritairement sur les panneaux. L'utilisateur de l'outil, peut dans un second temps rejeter les éventuelles fausses détections restantes. Pour autant, ces fausses détections sont intéressantes en termes de saillance relative aux panneaux et sont utiles au diagnostic du gestionnaire (cf. Chap. 5.6, 6 et 7.3).

Pour faire cet apprentissage, nous choisisons d'utiliser un algorithme, nommé "Support Vector Machines" (SVM), connu pour son efficacité et dont la convergence vers un minimum unique est garantie, contrairement à d'autres techniques d'apprentissage. Cette méthode est détaillée dans la partie 4.2. En résumé, durant la phase d'apprentissage, les différents aspects de chaque panneau sont mémorisés par le choix d'exemples pertinents (nommés Supports Vecteurs) et ils seront utilisés pour inférer la fonction de décision de classification des régions « panneaux et non-panneaux ».

L'utilisation du SVM, hormis le réglage de divers paramètres, nécessite les choix, d'une part d'une signature, représentant les données dans l'espace des caractéristiques et d'autre part d'une mesure de dissimilarité idoine à ces dernières, le noyau. La signature et le noyau doivent, tacitement, être adaptés à la détection des panneaux. Ces panneaux ont des particularités de formes mais surtout des attributs de couleurs spécifiques, il faut donc prendre en compte ces informations dans les signatures, comme nous l'étudierons dans la partie 5.4.

En amont, afin de déterminer les Supports Vecteurs et les paramètres du SVM, nous devons :

- Choisir un ensemble de modèles de panneaux qui constituerons notre base d'exemples de labels positifs ;
- Et, afin que l'algorithme puisse converger, un ensemble d'exemples négatifs, représentant quant à eux les parties de l'image qui ne contiennent pas

de panneaux.

5.2 Construction des bases d'exemples pour l'apprentissage et les tests

Afin de créer les exemples utiles à la base d'apprentissage et à la vérité terrain de référence pour les bases de test, nous avons utilisé un programme nommé SAFOR (SAisie de FORme) développé au LRPC de Strasbourg par Pierre CHARBONNIER et que nous avons amélioré durant la thèse. Le principe est de générer un masque binaire de la forme que l'on a sélectionné dans une image. Ainsi, en chargeant une image comme le montre la Fig. 5.1, on a alors la possibilité de choisir le type de panneaux que l'on va sélectionner dans l'image, celui-ci pouvant être de type Triangle, Ellipse, Rectangle, Losange ou Octogone. Une fois le type de panneaux choisi à l'aide des icônes, il apparaît un contour de la forme au centre de l'image, qu'il suffit de déplacer et déformer, afin de coller au mieux au panneau voulu. Ce programme permet d'effectuer, rapidement, par la sélection des coins du cadre (jaune) de présélection, des rotations ou de changer la taille du contour généré (vert) afin d'adapter la forme du modèle au panneau dans l'image. On peut, par ailleurs effectuer des transformations affines de cette forme afin de pouvoir entourer correctement des panneaux d'allures diverses, comme dans le cas de ceux qui ne sont pas fronto-parallèle à la caméra.

Une fois ce contour mis en place autour du panneau sélectionné et validé, grâce à une de nos modifications apportées, la génération du masque se fait automatiquement ainsi que son enregistrement, selon une norme prédéfinie. Cela permet, par la suite, une recherche rapide des masques d'une image, selon le type de panneau, par nom de fichier. Le masque créé, comme le montre la Fig. 5.1 est sous la forme d'une image binaire, qui nous sert ensuite à extraire le panneau de son environnement.

Ainsi, grace à ce procédé, nous avons réalisé une base contenant des panneaux routiers de taille, d'illumination, d'éclairage et de condition météo différentes, ceci afin de faire apprendre au SVM les panneaux dans tout l'éventail des apparences possibles qu'ils peuvent prendre. Pour cela, nous avons découpé la base en

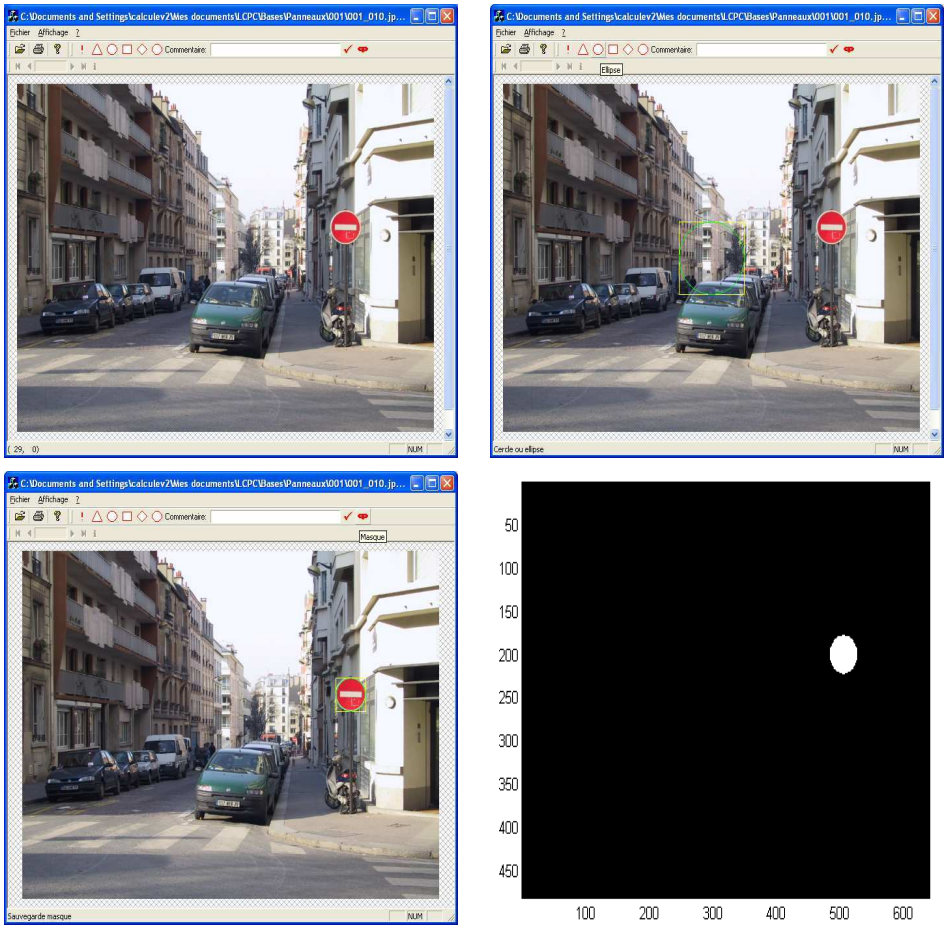


FIGURE 5.1 – Extraction d'un masque de Panneau via SAFOR.

deux parties, une pour l'apprentissage et l'autre pour la base de test.

Ce programme, durant le stage de Messaoud Guerrida que nous avons encadré [1], a été modifié afin de le rendre plus ergonomique et de générer une base de vérité terrain codée en fonction du type de panneau, comme le montre la Fig. 5.2. La sélection de la forme du panneau et la génération du nom du masque se fait automatiquement en fonction du choix, par l'utilisateur, du panneau à extraire. Par exemple, la sélection d'un panneau de sens interdit proposera une forme circulaire et le code B1, en respect avec la nomenclature de la signalisation.

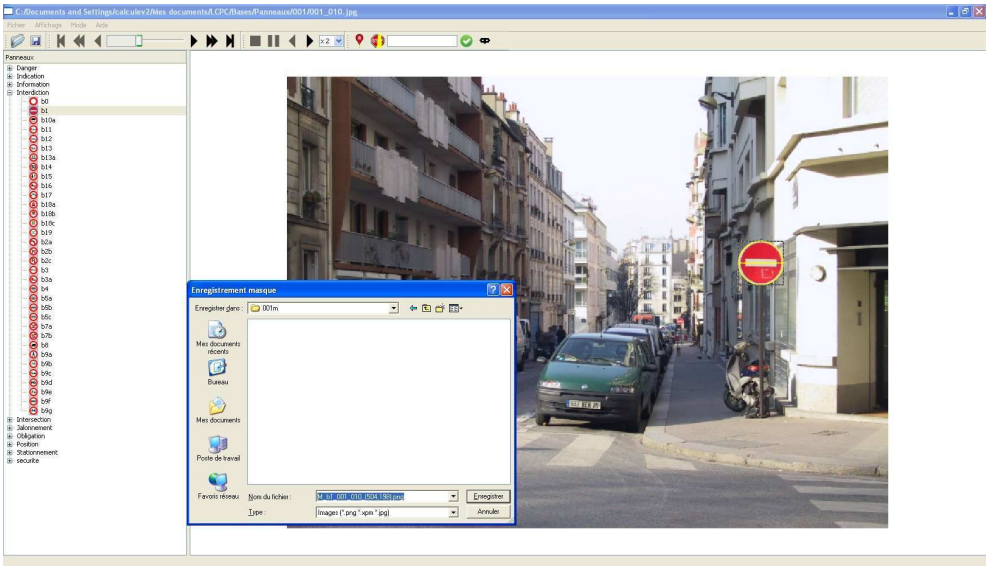


FIGURE 5.2 – Nouvelle version ergonomique de SAFOR.

Ce type d'application, plus ergonomique et ayant un formalisme dans l'enregistrement des masques, est utile à tout le réseau scientifique et technique dont le LCPC fait parti. Sur la base de ce travail et des propositions d'amélioration que nous avons suggérées, une collaboration avec le LRPC de Strasbourg (Philippe FOUCHET) et le LCPC-Nantes (Philippe NICOLLE) a permis de recoder ce programme en un code open source (QT), fournissant un utilitaire gratuit.

1. Ce stage de 4 mois s'inscrivait dans le cadre de son parcours de première année de Master Sciences de l'Ingénieur, mention Informatique Industrielle et Systèmes Automatisé de l'Université Paris VI.

Durant le début de la thèse, nous avons étoffé la base en réalisant de nombreuses prises de vues (photographie par un piéton) et saisies avec SAFOR. Nous avions une base de travail convenable avec toutefois une prépondérance en panneaux de sens interdit. De nouvelles acquisitions, plus valides écologiquement [2] par rapport à une prise de vue (cone de vision) en conduite, à l'aide d'un appareil professionnel (Nikon D200) fixé dans une voiture, ont été réalisées. Celles-ci ont complété la base et certaines images ont été utilisées pour l'expérimentation psycho-visuelle de validation des modèles décrite dans le chapitre 7.

5.3 Procédure d'apprentissage et de recherche de panneaux

Comme il est indiqué précédemment, la détection de panneaux par SVM s'effectue en deux phases. La première est l'apprentissage, qui permet de sélectionner les supports vecteurs et ainsi de définir la fonction de classification. La deuxième phase, appelée phase de détection, utilise cette fonction afin de déterminer si un échantillon d'image donné correspond à un panneau ou non.

5.3.1 Apprentissage standard

L'apprentissage automatique peut se classifier en deux types, l'apprentissage supervisé et non-supervisé. Dans le cas du non-supervisé, il y a en entrée un ensemble de données collectées. Ensuite, le programme traite ces données comme des variables aléatoires et construit un modèle de densités jointes pour cet ensemble de données. Dans le cas du supervisé, on cherche à établir automatiquement les règles de classification à partir d'une base de données d'apprentissage contenant des exemples de cas déjà traités, ce qui est communément appelé la vérité terrain. L'apprentissage par SVM, de par la labélisation des exemples d'apprentissage, fait partie de cette dernière catégorie. En effet, l'utilisation de SAFOR

2. La notion d'écologie se réfère ici à son sens le plus strict, utilisé en sciences ergonomique, qui est le respect des conditions environnementales [119]. Ici, on cherche la reproduction la plus fidèle de l'environnement visuel des conducteurs.

(cf. partie 5.2) nous permet d'avoir cette vérité terrain et de savoir si une fenêtre de l'image contient ou non, tout ou partie, d'un panneau de l'image.

Pour l'apprentissage, un expert fait le choix des images contenant des panneaux qui seront pris en compte pour l'apprentissage des exemples positifs. Ceci permet de choisir seulement quelques panneaux représentatifs de chaque types de forme, d'illumination, de déformation, etc., en ne prenant pas deux fois le même panneau sur deux images consécutives. Cela permet de contrôler l'apprentissage et de réduire d'emblée l'espace des données, d'où une rapidité accrue du SVM.

Ensuite, à partir de cette liste et des masques de panneaux correspondant à chaque image, les exemples de chaque classe sont extraits des images routières, puis les signatures voulues sont calculées sur ceux-ci.

Dans un premier temps, nous ne prenions que le panneau par lui même, éliminant ainsi le fond environnant du panneau. Or cette méthode est inadéquate, puisque l'apprentissage se réalise sans le fond. Or, lors de la recherche des panneaux dans une image, ceux-ci sont entourés d'un fond variable. La répartition de l'information dans la description de la fenêtre en cours d'analyse est difficilement comparable avec celles issues de l'apprentissage. Les performances de classification se trouvent donc réduites. Nous avons donc pris une zone plus large autour des panneaux lors de la construction des exemples de la base d'apprentissage. Nous avons, tout d'abord, élargi cette zone à un cadre englobant de plus de 10 pixels autour du masque généré par SAFOR. Or cette procédure n'est pas optimale car le ratio entre l'aire occupée par le panneau et celle occupée par le fond dans le modèle dépend de la taille du panneau. En effet, dans les images utilisées d'une résolution de 640×480, les plus petits panneaux considérés comme visibles dans l'image avaient une taille d'environ 8 pixels de coté et les plus grands une taille de 70 pixels de coté. Nous avons donc finalement opté pour un élargissement du cadre englobant autour du masque de SAFOR proportionnellement à la taille de ce dernier.

Ainsi, on prend en compte pour l'apprentissage des modèles comprenant 70% de panneau pour 30% de fond, comme dans la Fig. 5.3. Les différents types de fond sont inclus et par la suite, la performance de classification sur les fenêtres de recherche dans les images sont meilleures. Ceci s'explique par le fait que le fond

est très variable, donc lors de l'apprentissage, cela permet implicitement au SVM de se focaliser sur la partie de la signature qui caractérise un panneau de part la comparaison des signatures entre elles, via le noyau.

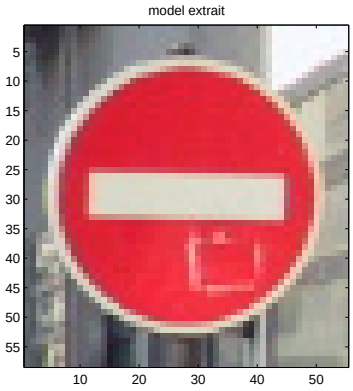


FIGURE 5.3 – Modèle d'un panneau après extraction selon son masque SAFOR (cf. Fig.5.1).

Cette première approche, prenant des panneaux entiers avec fonds variables en tant qu'exemples positifs (et des morceaux d'images ne contenant pas de panneaux en tant qu'exemples négatifs) est une approche globale. Par la suite, les résultats étant meilleurs et le nombre de support vecteur n'augmentant pas énormément, nous avons ajouté une approche locale. Ainsi, lors de l'apprentissage, se basant sur la vérité terrain SAFOR, les images de la base d'apprentissage sont traitées par des fenêtres glissantes de six tailles différentes (10×10, 16×16, 20×20, 30×30, 40×40, 60×60), comme montré dans la Fig. 5.4. Lorsqu'une fenêtre contient au moins 70% de panneaux, elle est labélisée comme un exemple positif. Si elle contient moins de 70% de panneaux, elle n'est pas prise en compte. Si elle ne contient pas de morceaux de panneaux, mais exclusivement du fond, elle est labélisée comme un exemple de la classe négative.

Un paramètre d'échantillonnage dépendant de la taille de la fenêtre glissante nous permet de limiter le nombre d'exemples dans chaque classe, afin de ne pas déséquilibrer l'apprentissage. Cela permet également de gagner du temps lors de la phase d'apprentissage, car selon la théorie des SVM, il est inutile de mettre dans

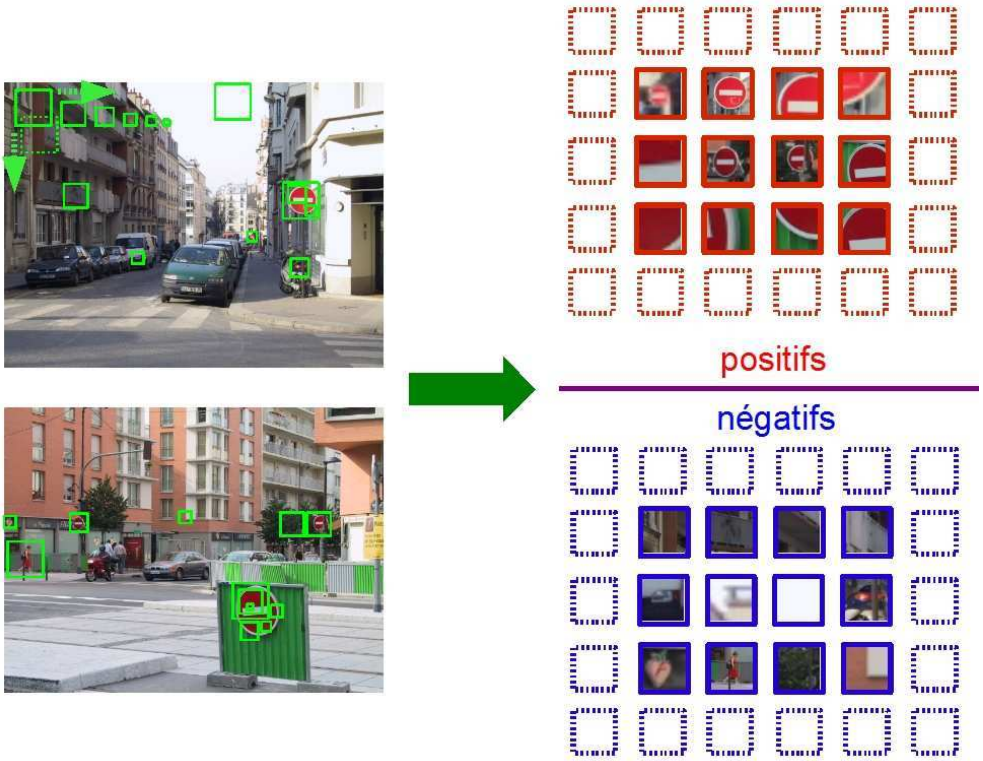


FIGURE 5.4 – Création des exemples de la base d'apprentissage

la base d'apprentissage des exemples très proches, car seuls les supports vecteurs restent représentatifs de chaque classe.

5.3.2 Influence du paramètre de régularisation dissymétrique

Comme indiqué dans la partie 4.2, il est possible d'introduire un terme de régularisation paramétré par C qui nous permet d'autoriser des erreurs de classification, en ne considérant pas certains exemples isolés, s'ils sont trop proches de la classe opposée. Ceux-ci sont alors appelés « points aberrants », « exemples aberrants » ou singularité (en anglais « *outlier* »).

Dans notre cas, notre but étant une détection, puis l'évaluation de toute la signalisation verticale, il ne faut pas rater des zones où il y a des panneaux et donc avoir un taux de reconnaissance proche de 1. Par contre, nous pouvons concéder

un taux de fausses détections un peu élevé, s'il reste corrélé avec les performances du conducteur (cf. partie 5.6.3). Les fausses détections peuvent être ignorées par l'opérateur dans l'outil final et pour la plupart, peuvent être filtrées en amont.

Nous avons donc utilisé un paramètre C de type dissymétrique, ce dernier étant proche de l'infini pour la classe positive (panneaux) et restant variable, à notre convenance, pour la classe négative. De ce fait aucun exemple aberrant entouré de positifs ne sera pris en compte, tout en gardant la possibilité d'avoir des singularités dans l'espace négatif.

Ce choix nous assure de prendre en compte tous les panneaux que l'on veut introduire lors de l'apprentissage et donc d'avoir le meilleur taux de reconnaissance possible. En effet, les exemples positifs qui auraient été considérés comme aberrants, sans le C de la classe positive à l'infini, deviennent logiquement des Supports Vecteurs. Ils définissent alors des nouvelles zones de la classe positive à eux seuls. Toutefois, cette méthode augmente naturellement le taux de fausses détections et il y a également un risque important de hausse de ce taux, si un exemple non-panneau venait à être mal labelisé en tant qu'exemple de la classe positive. Pour autant, un panneau labelisé comme appartenant à la classe négative aura moins d'incidence. Le paramètre de régularisation de la classe négative étant peu élevé, il ne sera pas considéré comme un exemple valide, mais comme aberrant.

5.3.3 Détection par fenêtres en multi-échelles

Les exemples pris en compte lors de l'apprentissage par le SVM sont créés via des fenêtres glissantes de taille variable. Il faut donc que les signatures qui seront utilisées par la suite lors de la classification aient les mêmes caractéristiques.

Du fait d'un effet de perspective et d'éloignement, les panneaux sont vus avec différentes tailles dans les images de scène routière. Pour chaque nouvelle image, la détection des panneaux est donc réalisée par des fenêtres de recherche glissante de tailles différentes. La valeur de décision $d(x)$ du SVM sur chaque fenêtre est calculée. Prenant comme paramètre $\sigma = 1$, les fenêtres ayant une valeur de prédiction supérieure à 1 contiennent donc tout ou partie d'un panneau et celles en dessous de -1 n'en contiennent pas. Les panneaux dans nos images de résolution

640 ayant des tailles variant d'environ 10 à 70 pixels de coté, nous avons procédé à plusieurs tests essayant des combinaisons de fenêtres de tailles différentes et observant l'évolution des capacités de détection. A l'issu de cette étude, nous avons sélectionné, pour l'apprentissage et la recherche, six tailles de fenêtres différentes (10×10, 16×16, 20×20, 30×30, 40×40, 60×60). Le facteur de translation des fenêtres en verticale et horizontale est d'$1/4$ de la taille de la fenêtre de recherche.

Globalement, les panneaux sont détectés dès que le ratio entre le panneau et le fond dans la fenêtre est proche de 50%, ce qui implique que l'on a bien pris en compte la variance du fond mais aussi du panneau.

Combinant ceci au fait que, lors de l'apprentissage, les panneaux ont été appris dans les différentes conditions d'éclairage et météorologiques, le système est donc capable de repérer des petits panneaux d'une taille de 8 pixels cachés dans l'ombre, aussi bien que des panneaux de 70 pixels sur-éclairés (reflets spéculaire). De plus, cette méthode est peu coûteuse en temps de calcul, car les fenêtres se déplacent par saut, ce déplacement étant fonction de leur taille ($1/4$ de la taille de la fenêtre).

L'utilisation de plusieurs échelles différentes est rendue possible par le fait que les signatures choisies (cf. partie 5.4) sont robustes à la taille et à la résolution de la fenêtre.

Chaque échelle utilisée fournie une carte de décision, ce qui résultent en six cartes différentes. Comme montré dans la Fig. 5.5, ces six cartes sont fusionnées en prenant le maximum de la valeur de classification (décision) du SVM pour chaque pixel sur les six cartes. En seuillant cette carte, on obtient alors un masque des zones détectées comme contenant un panneau.

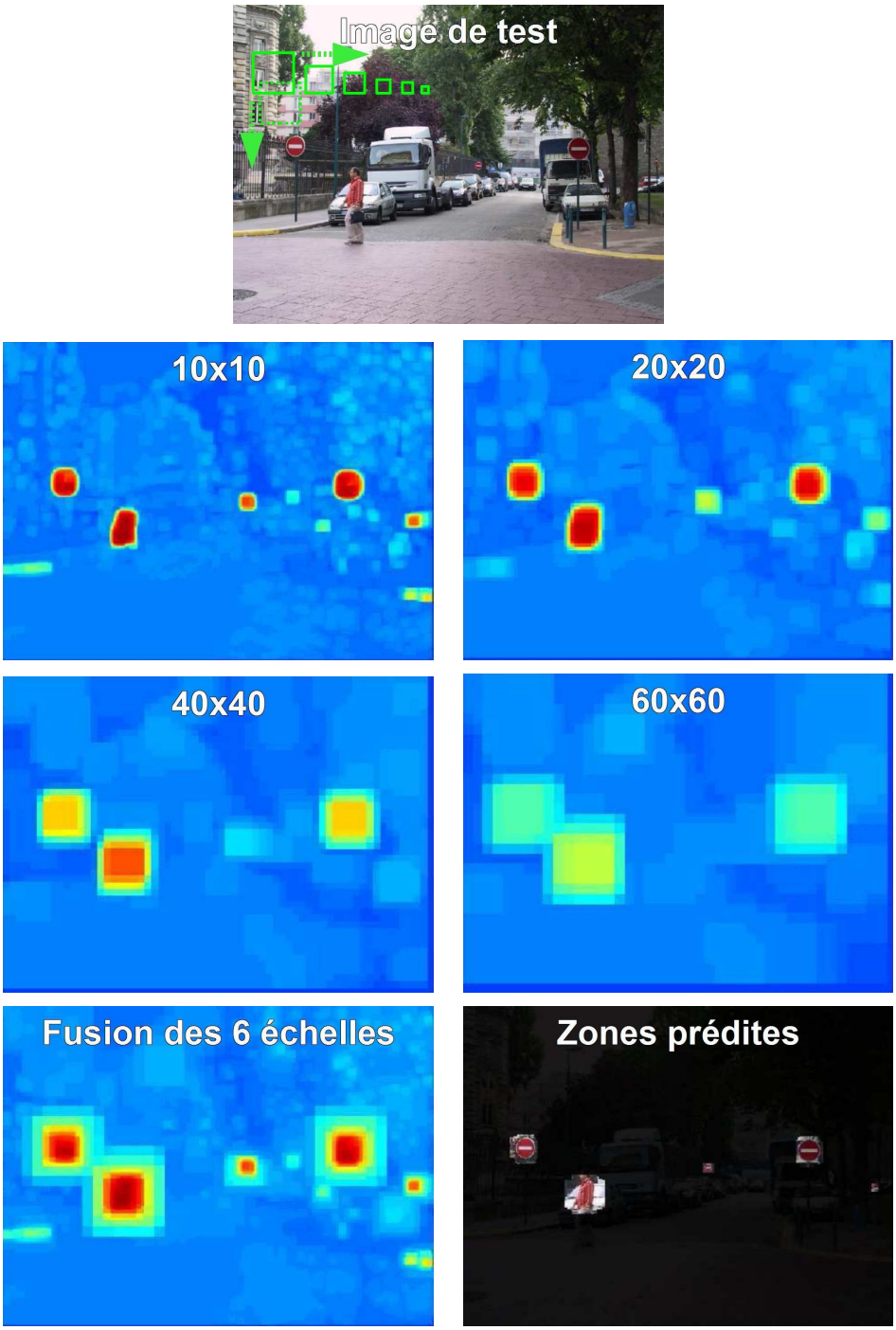


FIGURE 5.5 – Procédure de détection des panneaux. En haut, l'image originale. Au milieu le résultat de classification pour quatre échelles différentes. En bas, à gauche, le résultat de la fusion des cartes ; à droite, les parties de l'image détectées comme panneaux (seuillage de la carte de décision à la valeur $\sigma = 1$)

5.4 Etude sur les signatures

Pour que la classification par SVM soit optimum, il faut choisir une signature qui représente de manière correcte l'aspect des objets que l'on veut détecter dans les images.

La taille de la signature, et donc la richesse de l'information qu'elle contient, affecte le nombre de Supports Vecteurs, et donc conditionne la capacité de classification du SVM, mais également sa rapidité. Comme montré dans la Fig. 5.6, si on prend trop peu d'information, les données deviennent difficiles à distinguer. Inversement, plus on introduit d'informations, plus on doit introduire d'exemples dans la base d'apprentissage pour garder un pouvoir de discrimination, ce qui a pour effet de générer plus de support vecteur.

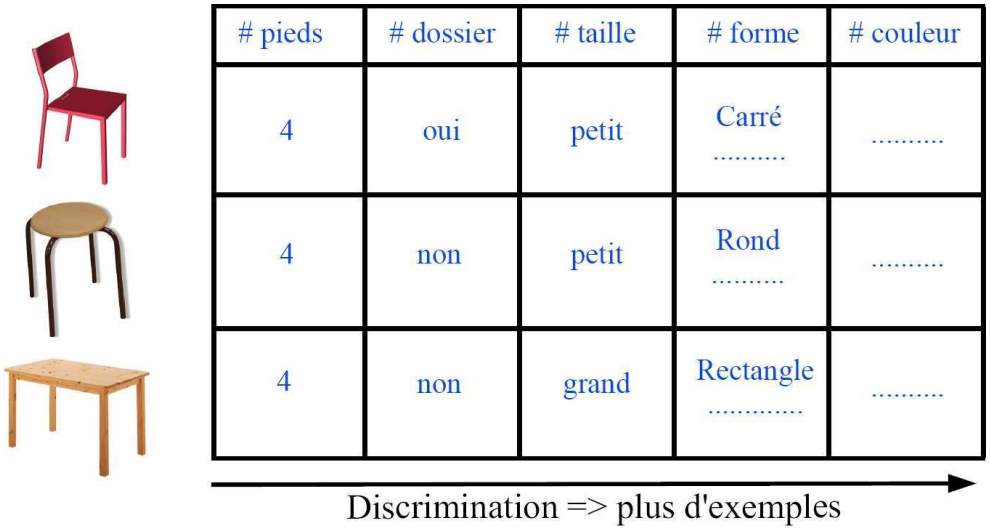


FIGURE 5.6 – Importance du choix de la signature. Exemple de la distinction entre une chaise, un tabouret et une table

Pour faire la détection des panneaux, nous avons choisi de concevoir une signature de taille fixe, rapide à calculer, du type histogramme, du fait du grand nombre d'échantillon à parcourir lors de la classification. Comme les panneaux n'ont pas tous la même taille dans les images, les signatures ont été conçues de façon à ce qu'elles soient invariantes à la taille du modèle, c'est-à-dire de la fenêtre utilisée pour l'extraction de l'imagette.

5.4.1 Histogramme des angles et des couleurs des contours

La première signature que nous avons créée est fondée sur l'extraction des contours dans l'image par un algorithme de type Deriche en couleur. Cette signature s'inspire du fonctionnement bas niveau du système visuel humain (cf. Chap. 3), ce dernier ayant une propension à fixer les contours (angles, ...) et les transitions de couleurs Rouge-Verte et Jaune-Bleue. S'inspirant des travaux de Itti [54] sur la saillance attentionnelle, nous avons adapté son extraction de caractéristiques visuelles en une signature.

Ainsi, d'une image jpeg au format *RGB*, on obtient une carte sur trois plans, comme le montre la Fig. 5.7. Le premier plan contient les valeurs des angles des contours ; le deuxième contient le gradient normalisé des $R-G$ (transition de couleurs Rouge-Vert) au point de contour et le dernier le gradient normalisé des $Y-B$ (transition de couleur Jaune-Bleu). Ces trois plans prennent des valeurs échantillonnées de 1 à 255. La valeur 0 est réservée pour les points qui n'appartiennent pas au contour.

Afin de prendre en compte le bruit, les petites variations de forme, d'orientation et de couleur, nous avons effectué une quantification des 3 plans sur *NbAngle*, *NbRG* et *NbYB* valeurs respectivement, qui sont des paramètres fixés au début de l'algorithme.

Comme nous pouvons le voir sur la Fig. 5.7, une imagette est représentée par trois histogrammes : celui des angles, celui des gradients R-G, et celui des gradients Y-B, de tailles *NbAngle*, *NbRG* et *NbYB* respectivement. Ensuite, chaque histogramme est normalisé en fonction du nombre de pixels de contour dans l'imagette. Puis ils sont concaténés afin de former la signature. Cette dernière est alors invariante à la taille de la fenêtre dont est issue l'imagette.

Par nos tests, nous avons observé que cette signature offre les meilleures performances lorsque les valeurs de *NbAngle*, *NbRG* et *NbYB* sont prises à 64, on obtient alors une signature assez longue avec 192 composantes. Toutefois, avec cette signature, les performances restaient moyennes et le taux de fausses détection restait assez élevé avec des détections visuellement très différentes des panneaux (par exemple, fausses détections sur des fenêtres de bâtiments).

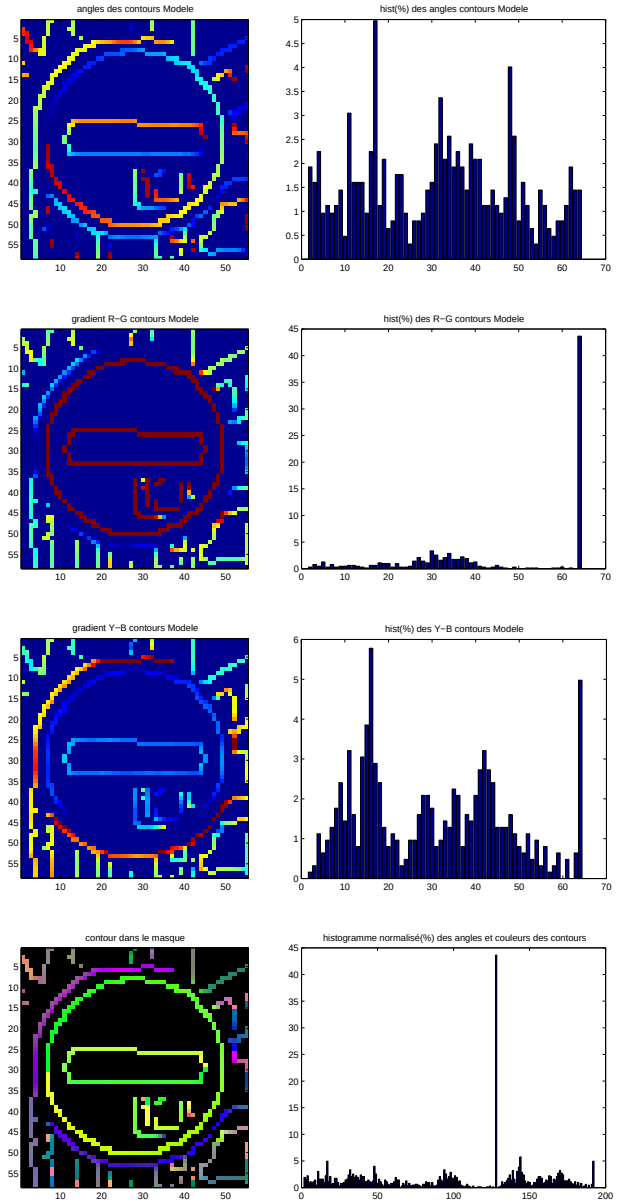


FIGURE 5.7 – Création de la signature des angles et couleurs dans la fenêtre du modèle de la Fig. 5.3. Les trois premières lignes correspondent respectivement à l'angle du gradient, au gradient $R-G$ et au gradient $Y-B$, contenus dans l'image de gauche. La colonne de gauche montre les plans correspondants. Celle de droite montre l'histogramme normalisé de ces valeurs. Sur la dernière ligne : à gauche, les trois plans forment une image ; à droite l'histogramme qui représente la signature finale, concaténation des trois histogrammes précédents.

5.4.2 Histogramme des couleurs

Utilisant des images au format *RGB*, une signature naturelle est de calculer l'histogramme des couleurs *RGB* dans la fenêtre. Afin de réduire la dimension de la signature, nous effectuons une quantification de chaque plan couleur. De par nos tests, nous avons obtenu les meilleurs résultats pour une quantification en 6 classes (ou intervalles, « *bins* » en anglais) par plan de couleur. La signature obtenue est de taille 6^3, soit 216 composantes. Malgré tout, les résultats restent assez faibles, le taux de détection étant éloigné de 1 pour un taux de fausses détections encore assez élevé. Ces résultats peuvent être expliqués par les problèmes d'illumination qui font varier les valeurs de couleurs dans l'image. Par exemple, une composante bleue est présente sur une partie de l'image dans l'ombre.

5.4.3 Histogramme des couleurs normalisées

Nous avons aussi testé des signatures couramment proposées pour les images couleurs en reconnaissance d'images, en particulier le classique travail de Swain & Ballard [120] sur l'indexation couleur.

En prenant une image au format *RGB*, nous effectuons une normalisation des couleurs selon :

$$r = \frac{R}{R+G+B} \tag{5.1}$$

$$g = \frac{G}{R+G+B} \tag{5.2}$$

$$b = \frac{B}{R+G+B} \tag{5.3}$$

Comme $r+g+b=1$, en pratique, on ne conserve que r et b. Les valeurs de r et b sont quantifiées pour réduire la taille des signatures et introduire la robustesse au bruit et aux petites variations de couleur.

De par nos tests, nous sommes arrivés à la conclusion que le taux de reconnaissance est meilleur lorsque que nous prenons un pas de quantification large sur chaque plan. Il suffit donc de 8 à 16 valeurs par couleurs normalisées. Comme le montre par ailleurs Swain & Ballard [120] et le confirme la partie application

du [117], nous avons les meilleurs résultats avec une quantification en 16 valeurs par couleur. La Fig. 5.8 montre la quantification obtenue.

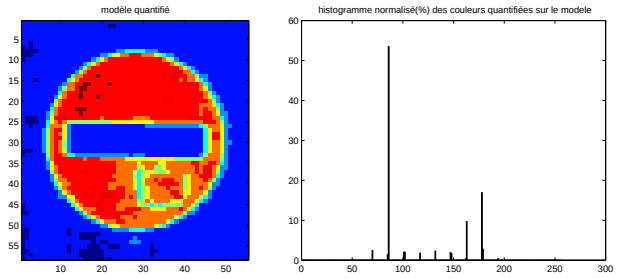


FIGURE 5.8 – A gauche, modèle quantifié d'un panneau après extraction selon son masque SAFOR (cf Fig.5.3 et Fig.5.1). A droite, la signature de type histogramme normalisé des couleurs quantifiées de ce modèle.

Par la suite, sur chaque fenêtre, nous calculons un histogramme des couleurs normalisées. Il est aussi normalisé par rapport aux nombres de pixels dans l'imagette. Ainsi, nous obtenons donc une signature de taille fixe à 256 composantes, représentant le pourcentage de chaque couleur dans la fenêtre, comme le montre la Fig. 5.8.

Cette signature est à la fois rapide à calculer et ne change pas trop suivant la taille de son échantillonnage. De plus, elle est naturellement invariante à la rotation, à l'occlusion d'une partie du panneau ainsi qu'à ses déformations. Par ailleurs, elle donne de très bons taux de reconnaissance. Toutefois, le taux de fausses détection est un peu élevé. En effet, n'étant basée que sur la couleur, deux fenêtres différentes peuvent avoir le même histogramme, comme le montre la Fig. 5.9. En revanche, le fait que cette signature ne comporte que 256 composantes permet une grande rapidité lors de l'optimisation du SVM pour choisir ses SV et lors de la classification.

5.4.4 Histogramme des couleurs normalisées sur les contours

Disposant des contours de l'image (cf partie 5.4.1), il semblait intéressant d'utiliser cette information afin de rendre la signature précédente plus sélective

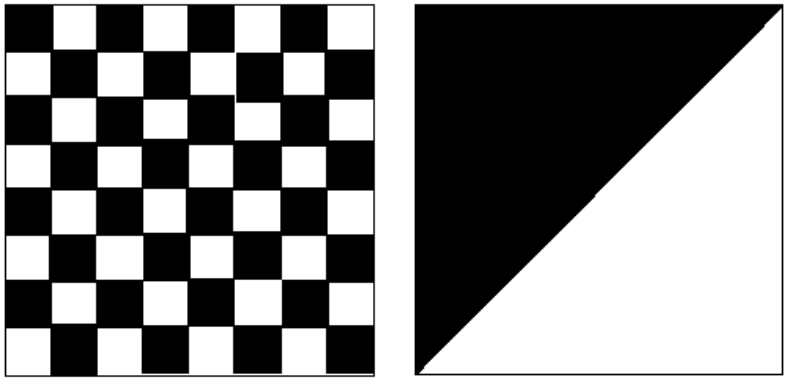


FIGURE 5.9 – Deux fenêtres différentes ayant le même histogramme couleur. Extrait de [121].

dans le but de réduire le taux de fausses détections. On a donc restreint l'approche à la carte des contours sous sa forme binaire issue de l'algorithme de détection de contour de type Deriche. Puis, on a calculé l'histogramme des couleurs normalisées *r* et *b* quantifiées sur les contours, comme le montre sur la Fig. 5.10, l'histogramme de couleurs normalisées quantifiées des pixels appartenant aux contours dans la fenêtre du modèle. Enfin, cet histogramme a été normalisé par le nombre de pixels de contours présent dans l'imagette.

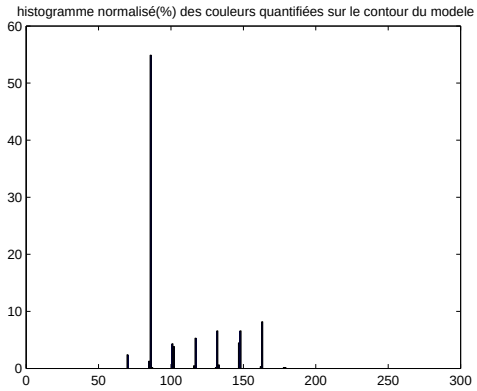


FIGURE 5.10 – Signature de type Histogramme normalisé des couleurs quantifiées sur les contours dans la fenêtre du modèle de la Fig.5.3.

Cette signature ne change pas trop selon la taille de la fenêtre et l'angle de la rotation. De plus, elle reste rapide à calculer, sa longueur étant de 256 composantes. En revanche, les taux de reconnaissance et de fausses détections obtenus avec cette signature sont moins bons qu'avec les couleurs quantifiées sur la fenêtre (cf partie 5.4.3). Ceci s'expliquant par le fait que les contours correspondent à des zones de transition entre deux couleurs où celles-ci sont moins bien définies.

5.4.5 Histogramme de la distribution des couleurs normalisées autour des contours

Afin de pouvoir profiter de l'information contour, il nous faut donc pouvoir utiliser des informations couleurs dans des régions locales autour des contours et non sur les contours. Certains auteurs, comme [122], ont calculé des distances entre des histogrammes de distributions de couleurs quantifiées autour des points de contours orientés, afin de détecter des objects dans les images. D'autres [123] ont travaillé, avec la distribution de texture, ou encore, [124] ont utilisé des cooccurrences entre histogrammes des contours couleurs.

Plutôt que de prendre les couleurs juste au niveau du contour comme dans la partie 5.4.4, nous les avons donc prises jusqu'à une certaine distance du contour fixée à l'avance, dans notre cas à cinq pixels.

Ainsi, comme le montre la Fig. 5.11, pour chaque pixel appartenant au contour, il est défini cinq zones correspondant à quatres pas d'éloignement du contour en pixel. Il est à noter que la première zone est le contour lui même, on retrouve alors dans la première ligne de l'histogramme, la signature qui a été établie dans la partie 5.4.4. L'histogramme 2D généré contient le nombre de pixels de chaque couleur normalisée quantifiée dans chaque zone. Puis, cet histogramme, est d'abord normalisé par rapport au nombre de pixels dans chaque zone autour du contour. Cela donne alors un histogramme qui représente le taux de chaque couleur quantifiée dans chaque zone. Afin de pouvoir être utilisé par le SVM, nous formatons cet histogramme sous la forme d'un vecteur ligne. D'après nos tests, nous avons des résultats satisfaisants en prenant en compte une distribution de 5 pixels autour du contour, comme le montre les Fig 5.12 et 5.13.

Cette signature est aussi peu sensible à la taille de la fenêtre et aux rotations.

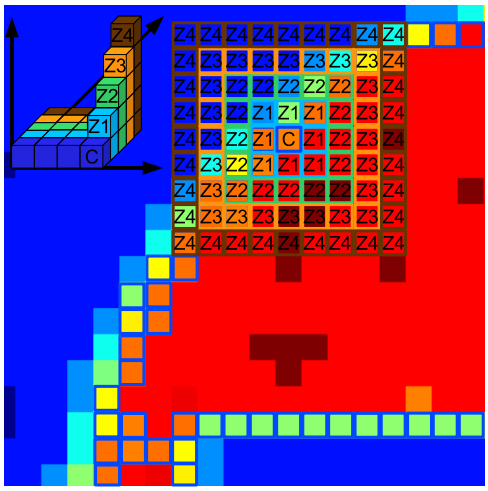


FIGURE 5.11 – Extraction de la distribution des couleurs normalisées autour des contours.

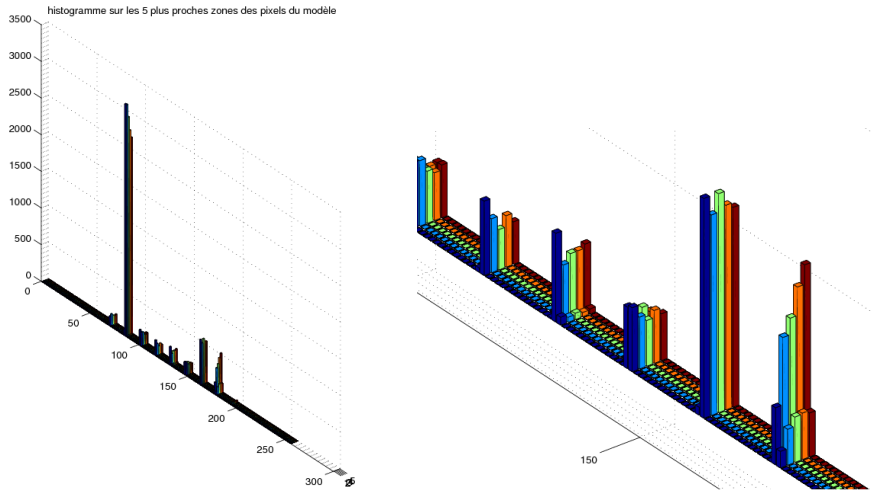


FIGURE 5.12 – Histogramme de la distribution des couleurs normalisées autour des contours dans la fenêtre du modèle de la Fig.5.3 (à gauche) et un zoom sur quelques valeurs (à droite).

Elle donne de bons résultats avec un bon taux de reconnaissance et un taux de fausses détection plus bas qu'avec la signature de la partie 5.4.3 (histogramme des couleurs normalisées) qui ne prend pas en compte les contours. En revanche,

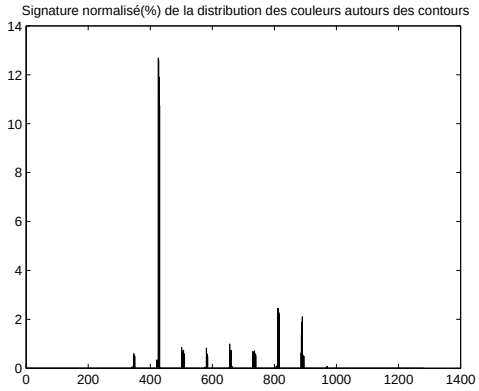


FIGURE 5.13 – Histogramme normalisé de la distribution des couleurs normalisées autour des contours dans la fenêtre du modèle de la Fig.5.3.

cette signature est la plus longue à calculer, car en effet, en prenant 5 zones, elle comporte alors 1280 éléments. De plus, sa création est, par elle même, très coûteuse en temps de calcul.

5.4.6 Histogramme des angles des contours OEH

La signalisation horizontale ayant des caractéristiques de formes particulières et définies, on s'est également intéressé à la détection de celle-ci en apprenant les formes de son contour. Nous avons donc utilisé comme signature un histogramme normalisé de l'orientation du contour OEH (Oriented Edge Histogram), déjà calculé en 5.4.1. Les meilleurs résultats obtenus sont avec une quantification de 12 bins. Or, cette signature n'étant pas assez discriminante, ses résultats restent très faibles, avec un taux de détection en dessous de 0,1 et un taux de fausses détections très élevés.

5.4.7 Histogramme de la concaténation OEH et couleurs

L'information de forme à elle seule n'étant pas suffisante, et l'information de couleurs quantifiées donnant des résultats corrects, nous avons essayé de fusionner ces deux signatures afin de tester si une amélioration des résultats était possible.

Nous avons donc concaténé les histogrammes d'orientation du contour sur 12 bins (5.4.6) et de couleurs quantifiées *rb* sur 12^2 bins (5.4.3). Or, cette signature n'améliore pas les résultats et au contraire dégrade les performances tout en augmentant considérablement le nombre de Supports Vecteurs, donc le temps d'apprentissage et de détection.

5.5 Etude sur les Noyaux

Outre le choix d'une signature adaptée à notre problème, nous devons aussi effectuer le choix du noyau $k(x,x')$ qui est une mesure de dissimilarité entre deux exemples x et x'.

5.5.1 Noyau Gaussien

Le noyau le plus classique est le noyau gaussien, qui s'exprime sous la forme :

$$k_{gaussien}(x,x') = \exp(-\frac{\|x-x'\|^2}{\sigma^2}) \tag{5.4}$$

Le défaut de ce noyau est la nécessité de sélectionner une échelle σ correcte par rapport aux données. Comme le montre la Fig.5.14, ce paramètre a une grande influence sur le résultat de la classification. Nous avons trouvé que $\sigma = 100$ est une valeur adéquate pour nos données.

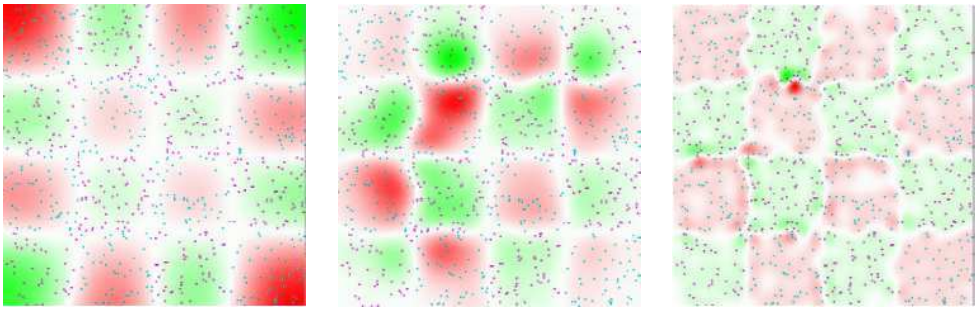


FIGURE 5.14 – Classification de positifs et négatifs disposés selon un jeu d'échec, par le noyau gaussien, avec de gauche à droite une échelle σ du noyau de 1, 10 et 100 respectivement. Extrait de [121].

5.5.2 Noyau Polynomial

Un autre noyau classique est le noyau polynomial défini par :

$$k_{poly}(x,x') = (1 + \frac{x \cdot x'}{\sigma^2})^{\alpha} \tag{5.5}$$

Ce noyau comporte un facteur d'échelle σ et un facteur de degré entier α, il est donc plus compliqué à régler que le noyau gaussien. En essayant plusieurs valeurs de α et de σ lors de nos tests, nous n'avons pas obtenu de meilleurs résultats qu'avec le noyau Gaussien. Le noyau polynomial ayant un paramètre de plus à régler, nous lui préférerons donc le noyau Gaussien.

5.5.3 Noyau Laplacien

Utilisant des histogrammes comme signature, il semble a priori intéressant d'utiliser le noyau Laplacien. Comme le présente [125], en plus de respecter la condition de Mercer (il est donc défini positif), il est plus adapté pour la classification d'histogrammes que celui Gaussien, et il améliore généralement les résultats par rapport à ce dernier.

Ce noyau est de la forme :

$$k_{Laplace}(x,x') = \exp(-\frac{\|x - x'\|}{\sigma}) \tag{5.6}$$

En utilisant ce noyau, nous avons de bons résultats, en prenant un facteur d'échelle de $\sigma = 50$. Ceux-ci sont sensiblement meilleurs que les résultats obtenus avec le noyau gaussien.

5.5.4 Noyau Triangulaire

Le noyau triangulaire, aussi appelé noyau puissance est intéressant car, comme montré dans [126], il permet de s'affranchir du choix du paramètre d'échelle :

$$k_{power}(x,x') = -\|x - x'\|^{\alpha} \tag{5.7}$$

Comme il a été montré par [126], [127] et [128], ce noyau n'est certes pas défini positif, mais il reste utilisable car il est conditionnellement défini positif

(cf. partie **??**) pour $0 < \alpha \leq 2$. Les propriétés intéressantes de ce noyau est qu'il n'a pas de paramètre d'échelle, et donc il est invariant à la densité des données. Ainsi, le seul paramètre qui influence sur la forme de la frontière de classification qu'il reste à régler est le paramètre C de régularisation du SVM, pour un α fixé. En effet, comme le montre la Fig. 5.15 qui peut être comparée à la Fig. 5.14, le noyau triangulaire a un comportement qui reste identique aux différentes échelles des données, tandis que le noyau gaussien est très sensible à ces changements.

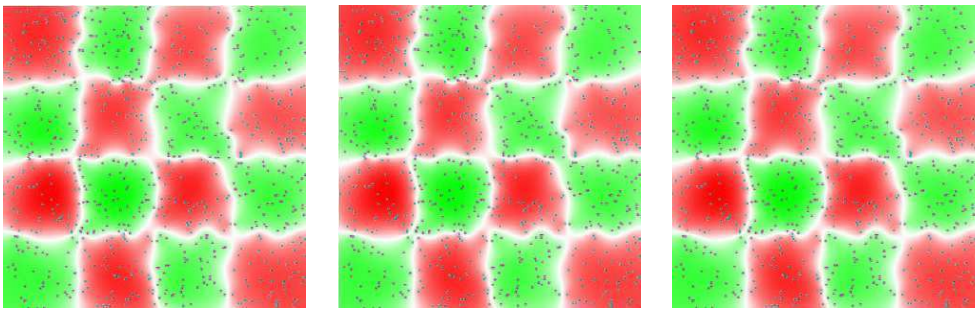


FIGURE 5.15 – Classification des positifs et des négatifs disposés selon un jeu d'échec, par le noyau triangulaire, avec de gauche à droite une échelle de donnée de 1,10 et 100 respectivement. Extrait de [121].

Cette invariance à la densité des données est importante lorsque le SVM doit séparer des classes à différentes échelles dans les données, comme c'est le cas dans la Fig. 5.16. On voit que selon l'échelle choisie, le noyau gaussien fait du sous-apprentissage ou du sur-apprentissage, tandis que le noyau triangulaire s'adapte à la densité de la distribution des données. Cela peut aussi se comprendre par la forme que prennent ces deux noyaux en fonction de l'échelle des données, comme le montre la Fig. 5.17.

Ce noyau est donc bien adapté à notre problème car la densité de nos données est peu prévisible et variable entre les deux classes. Nos deux classes n'auront pas la même distribution de données : les exemples de la classe négative sont beaucoup plus nombreux que ceux de la classe positive.

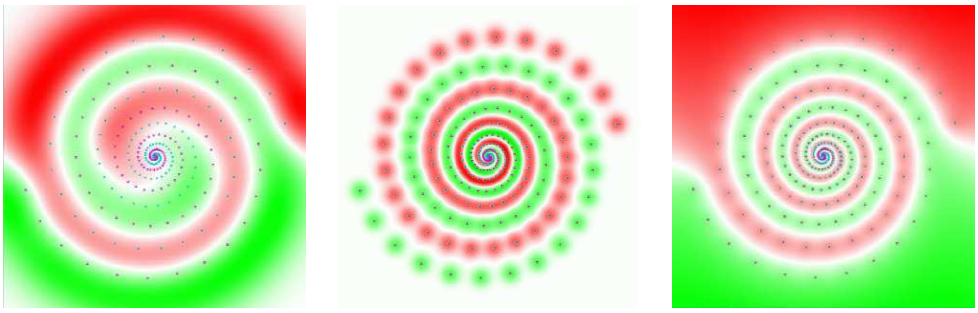


FIGURE 5.16 – Classification des positifs et des négatifs disposés selon des spirales, par le noyau gaussien (les 2 images de gauche à deux échelles) et par le noyau triangulaire (image de droite). Extrait de [121].

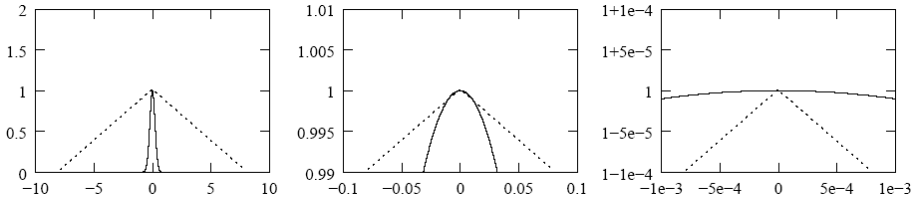


FIGURE 5.17 – Le noyau gaussien (ligne continue) et le noyau triangulaire (ligne discontinue) vus à différentes échelles. Extrait de [126]..

5.6 Choix des couples noyau/signature de meilleures performances

A l'issu de l'étude précédente, nous avons vu les avantages et les inconvénients de quelques signatures et noyaux. Nous avons sélectionnés certains d'entre eux, les plus prometteurs, et réalisés des tests afin de choisir le meilleur couple noyau/signature adéquat à notre problème. Nous rappelons ici que notre algorithme se doit d'avoir de très bonnes performances en termes de détection de panneaux, c'est à dire ne pas en rater, afin d'évaluer la saillance de l'intégralité de la signalisation de police présente dans la scène routière analysée. De plus, il doit être corrélé aux détections et fausses détections réalisées par un conducteur dans la même scène routière, ceci afin de prendre en compte les distracteurs visuels ressemblant à des panneaux. Nous devons donc évaluer les performances de l'al-

gorithme selon deux vérités :

- La vérité terrain, c'est à dire les zones de l'image qui contiennent des panneaux.
- et la « vérité humaine » : issue des données comportementales visuelles, en condition la plus écologique au vu de nos moyen expérimentaux disponibles. Cette vérité correspond aux performances humaines lors d'une tâche de recherche visuelle de panneaux de police dans une scène routière.

Nous présentons donc ici, après la description de la vérité terrain, une première expérimentation qui nous a permis de récolter des données oculaires, puis leur analyse. Ces premiers résultats expérimentaux nous ont permis de guider et d'évaluer nos choix.

5.6.1 Vérité Terrain, la signalisation verticale

Afin de tester notre détecteur, il est nécessaire de disposer de la vérité terrain qui est, pour chaque image, les coordonnées des pixels qui appartiennent à la signalisation verticale. Afin de créer ces calques de présence de panneaux, pour chaque image, nous effectuons l'accumulation des masques de chaque panneau contenu dans l'image. Les masques individuels sont obtenus par l'utilisation du logiciel interactif Safor, comme nous l'avons présenté dans la partie 5.2. Nous obtenons alors une carte de la vérité terrain comme présentée dans la Fig. 5.18

5.6.2 Données de comportement humain, expérimentation exploratoire sur les fixations du regard dans un contexte routier

L'algorithme développé, outre le fait d'être corrélé à la vérité terrain, doit également être corrélé au comportement humain en conduite, ou plus exactement à la sous-tâche de recherche de la signalisation de police durant la conduite. Pour autant, les moyens expérimentaux à notre disposition ne nous ont pas permis l'utilisation d'un véhicule instrumenté ou d'un simulateur de conduite. Ainsi, nous avons pris le parti de prioriser le contexte routier, les panneaux de police

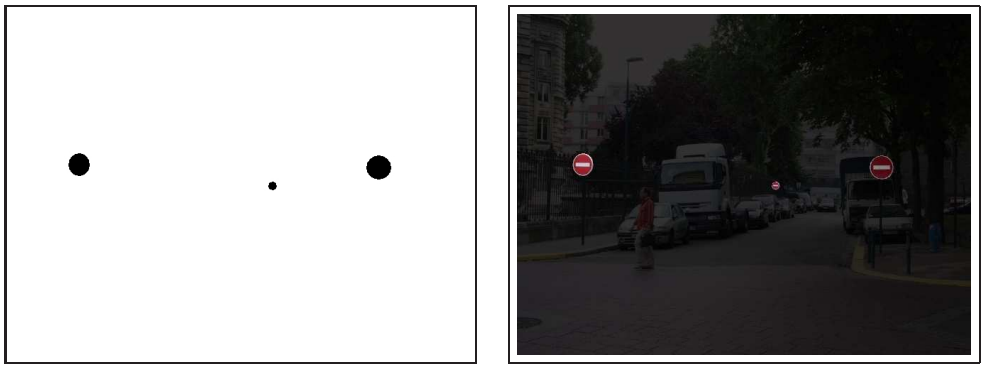


FIGURE 5.18 – A gauche, masque de vérité terrain correspondant aux panneaux de « sens interdit » dans l'image de la Fig.5.5. A droite, parties correspondantes de l'image.

faisant partie de l'environnement du conducteur. Pour cela, nous présentons des images de scènes routières aux participants de l'expérimentation exploratoire. La consigne proposée permettait d'induire un comportement de recherche attentionnelle chez les sujets. L'algorithme peut donc, lui aussi faire des fausses détections, à condition que celles-ci correspondent à des zones de l'images qui contiennent des distracteurs dont les caractéristiques de saillance sont proches de celle des panneaux recherchés. Pour autant, le manque de connaissance des zones prises en compte dans le relevé d'informations par le conducteur, lors de cette sous tâche de recherche, explique la mise en oeuvre de cette expérimentation exploratoire, dont l'objectif était le recueil de données oculaires sur des images routières avec des observateurs.

Protocole

Participants : Huit sujets, ayant le permis de conduire et une vision normale ou corrigé, ont passé l'expérimentation.

Matériel : Un système oculomètrique [94] et son ordinateur pilote sont interfacés en ethernet avec un autre ordinateur qui permet d'afficher les images de scènes routières. Ce deuxième ordinateur utilise un programme d'expérimentation psycho-visuelle, le logiciel *Presentation*^{TM}. Les chemins visuels

des sujets, dans les images, sont enregistrés par l'oculomètre. Les précisions sur le matériel utilisé se trouvent dans le chapitre 7.

Images : Les 20 images présentées aux sujets sont des scènes routières contenant ou non des panneaux de sens interdit (jusqu'à quatre). Ces images sont affichées, avec une résolution de 640 x 480, au centre d'un écran d'ordinateur de 19 pouces, avec une résolution de 1280 x 1024. Le contour de l'image est constitué d'un niveau de gris correspondant au niveau de gris moyen de l'image, comme le montre la Fig.5.19. Cela permet de faire une adaptation de la luminosité et de réduire le contraste entre l'image et le fond.

Procédure : Les sujets sont placés à une distance de 63 cm de l'écran, afin que la scène routière soit présentée avec un angle visuel de 20°, pour éviter les effets de la vision périphérique et les mouvements de tête réflexes. En effet, le matériel utilisé ne permet pas de mouvement de la part du sujet, puisqu'ils occasionnent une perte du suivi du regard et donc une dégradation des données enregistrées.

Les sujets ont pour consigne de compter le nombre de panneaux de sens interdit dans l'image qui leur est présentée, ce qui induit une tâche de saillance de recherche [8]. La réponse des sujets sur le nombre total de panneaux de sens interdit présent dans chaque image est enregistrée. Avant l'apparition de l'image contenant la scène routière, les sujets doivent fixer une croix noire sur un fond gris uniforme, afin que toutes les explorations commencent du même point.

Variables : Les variables dépendantes analysées sont l'emplacement, l'ordre et le temps des fixations des sujets. Les variables indépendantes sont la variation de la taille des panneaux, leur position, et leur aspect colorimétrique ainsi que la présence de distracteurs naturels ayant les couleurs du panneau d'intérêt de notre étude.

Hypothèse : L'hypothèse est qu'il sera trouvé une corrélation positive entre les fixations faites par les sujets dans la scène et les zones prédites par l'algorithme, comme étant susceptible de contenir un panneau de sens interdit.



FIGURE 5.19 – Exemples d'images présentée lors de l'expérimentation de recueil des données

Analyse des données et création de la carte de détection humaine

Sur les huit sujets, l'analyse des données de trois participants ont été retenues. Les données des cinq autres sujets ont été supprimé des analyses du fait des erreurs trop nombreuses dans les données oculométriques. Cette pertes de données est dû aux limites du système utilisé et aux conditions d'expérimentation (Mouvements de têtes des sujets, temps de réponse illimité). Ces points, pris en compte, sont discutés et modifiés dans l'expérimentation de validation présentée dans le chapitre 7.

La Fig. 5.20 montre le chemin oculaire effectué par les trois sujets lors de leur exploration pendant la recherche de panneaux de sens interdit, sur la même image. Ce chemin est constitué d'une suite de fixations, séparées par des saccades cf. parties 3.1 et 3.4. Les fixations sont déterminées en fonction des critères spatiaux (taille maximale de la zone dans laquelle les données oculaires doivent se trouver pour être considérer comme appartenant à la fixation) et temporels (temps minimum, c'est à dire nombre minimum de données dans la zone pour la considérer comme une fixation).

Si l'on ne considère que les fixations, comme par exemple sur la Fig. 5.21, un problème lié à la technique utilisée en oculométrie se pose (cf. partie 3.4.1). En effet, l'oculomètre ne récupère comme donnée que la succession échantillonée des points de passage du centre du regard, ce qui correspond à la vision fovéale. Ainsi, quand un objet est notifié, par le biais de la vision para-fovéale ou péri-fovéale, ou si le sujet ne reste pas fixé assez de temps sur celui-ci afin d'être pris comme fixation, on ne dispose alors pas d'observation sur la détection de cet objet par l'oculomètre.

Nous récupérons, à l'issu de la présentation de l'image, la réponse des sujets sur le nombre de panneaux de sens interdit qu'ils ont comptés dans la scène routière. Ceci nous permet d'inférer si un panneaux a été « vu mais pas regardé ». Dans ce cas, on peut alors supposer une forte saillance de ce dernier, comme cela est visible sur la Fig. 5.21. Sur les images de cette figure, le sujet répond correctement à la question posée : il compte respectivement de gauche à droite 3, 4, et 3 panneaux, alors qu'il ne fixe respectivement que 2, 3, et 2 d'entre eux. Par ailleurs, les sujets ont des fixations sur des objets qui ne sont pas des sens interdit,



FIGURE 5.20 – Chemins oculaires des trois sujets pendant la recherche de panneaux de sens interdits. Chaque cercle représente une fixation. La durée est indiquée en ms. Les fixations commencent au milieu de l'image.

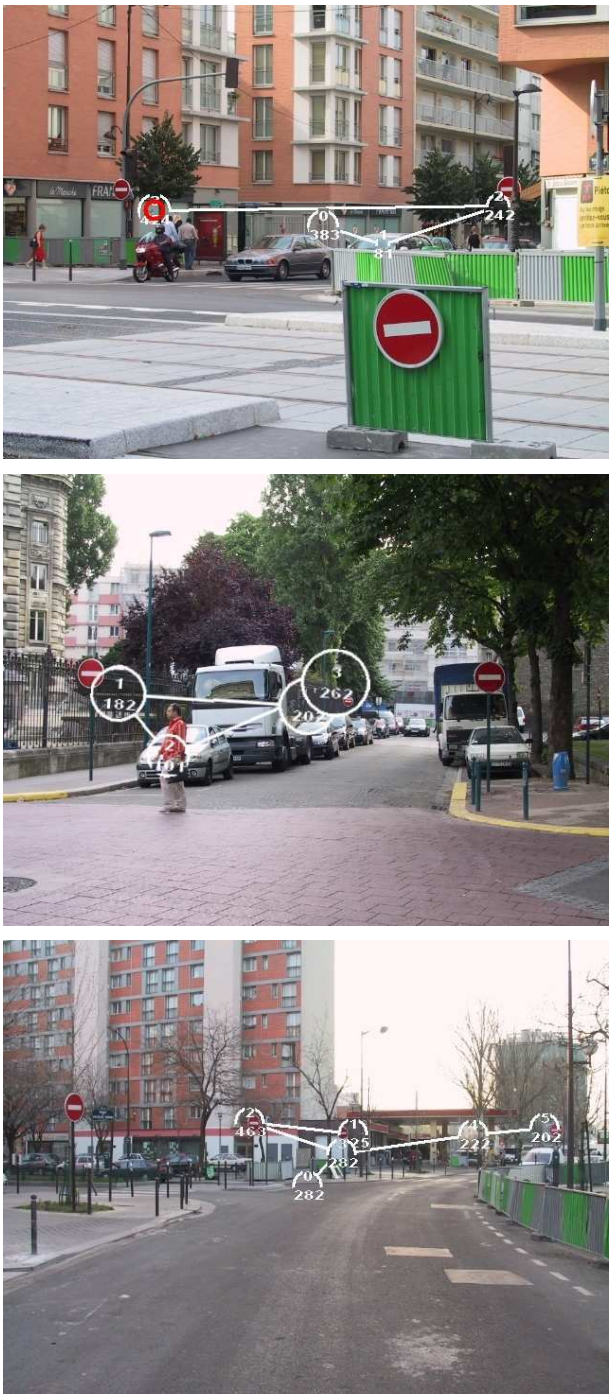


FIGURE 5.21 – Chemins oculaires d'un sujet sur trois images. Mise en évidence des effets de vision en proche périphérie et des distracteurs.

nous supposons que cela est du à l'aspect colorimétrique de ces objets (chemise à rayure large blanche et rouge) qui prennent le statut de distracteurs.

Afin de concevoir une *carte binaire de vérité de la « saillance de recherche »*, comme il est montré dans la Fig. 5.22, nous prenons en compte les réponses des sujets afin de déterminer quels panneaux de sens interdit ont été catégorisés et identifiés, donc détectés. La vision fovéale est définie sur deux degrés d'angle total, nous prenons donc en considération, dans la scène, les zones correspondantes aux panneaux notifiés (via la réponse) par le sujet et fixés (données oculométrique). Afin de prendre en compte les distracteurs dans la carte, nous prenons aussi en considération l'ensemble des points de fixation des sujets. La précision de l'oculomètre est introduite par un disque correspondant à 2° de vision (32 pixels dans notre cas) appliqué autour de ces points.

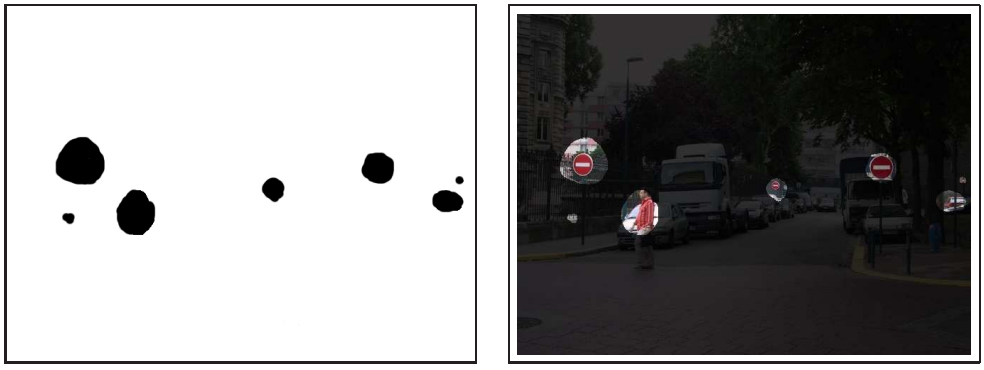


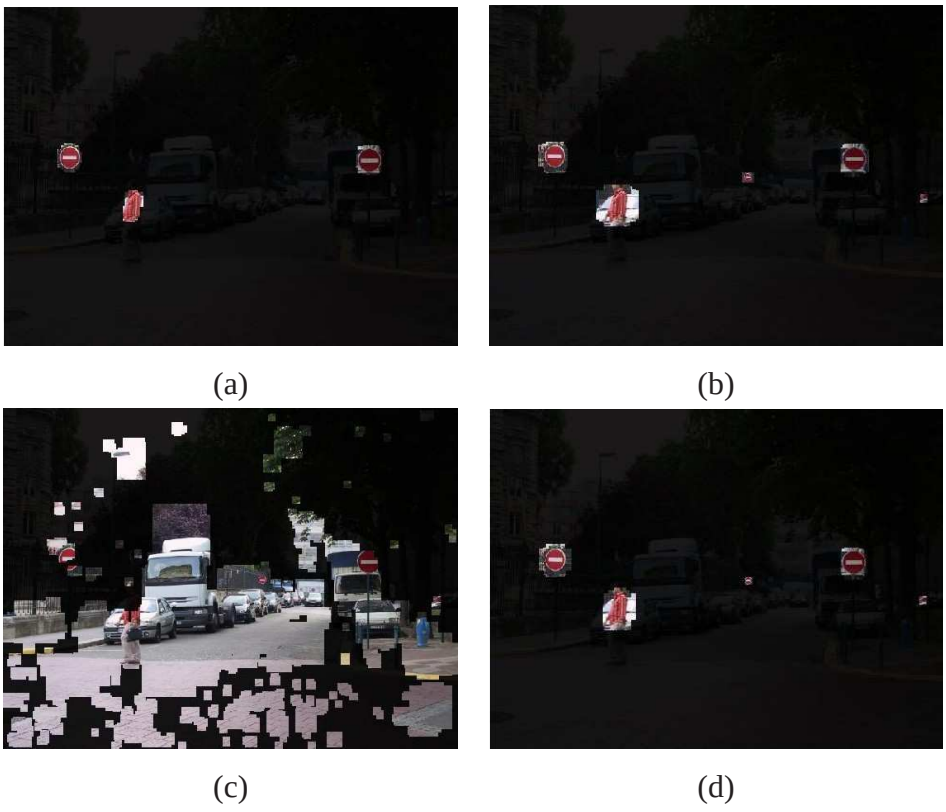
FIGURE 5.22 – A gauche, masque de vérité de détection en tâche de recherche correspond à l'image de la Fig.5.5. A droite, parties correspondantes de l'image.

5.6.3 Choix du couple en fonction des deux vérités

Au vu des tests préliminaires et de nos remarques décrites dans les parties 5.4 et 5.5, nous avons sélectionné comme signatures :

- L'histogramme des couleurs *RGB*, à 6 bins par plan, cf. 5.4.2 ;
- L'histogramme des couleurs normalisées *rb* à 12 bins par plan, cf. 5.4.3 ;
- L'histogramme de l'orientation des contours *OEH*, cf. 5.4.6 ;
- La concaténation des deux dernières *rb* et *OEH*, cf. 5.4.7.

La Fig. 5.23, montre les résultats obtenus pour ces quatres signatures. On voit d'emblée que la signature utilisant le contour n'est pas assez discriminante, mais elle peut être utilisée en complément avec des signatures couleurs (Fig. 5.23(c) et (d)).



(a) (b)

(c) (d)

FIGURE 5.23 – Carte des décisions du SVM (seuillage à la valeur 1) obtenue à partir de l'image originale de la figure 5.5(a), en utilisant le noyau triangulaire et différentes signatures : (a) Histogramme des couleurs *RGB* (6^3-bins) dans la fenêtre, (b) Histogramme des couleurs normalisées *rb* (12^2-bins) dans la fenêtre, (c) Histogramme des orientations du contour (12-bins) dans la fenêtre, (d) Histogramme de la concaténation des couleurs *rb* (12^2-bins) et des orientations du contour (12-bins) dans la fenêtre.

D'autre part, il est intéressant de comparer les capacités de classification de deux noyaux :

– Le noyau Laplacien, couramment utilisé pour la classification d'histo-

grammes cf. 5.5.3 ;

– Le noyau triangulaire, dont les propriétés sont bien adaptées à notre problème, cf. 5.5.4.

Pour plus de justesse et de précision dans nos comparaisons, nous construisons deux types de courbes ROC à partir des deux références de détections définies précédemment. Le paramètre de variation des courbes ROC est la valeur seuil appliquée sur la carte de confiance pour chaque couple noyau/signature. Les résultats sont affichés dans la Fig.5.24.

Dans tous les cas, l'histogramme des couleurs normalisées *rb* (12^2-bins) dans la fenêtre donne les meilleurs résultats en terme de taux de détections correctes par rapport aux taux de fausses détections. Comme prévu, l'orientation des contours seule donne de très mauvais résultats. Concatener ces deux signatures n'améliore pas les capacités de classification voire même les diminue un peu. Au prix de l'augmentation de la taille de la signature, du nombre de supports vecteurs, et donc du temps de classification ; utiliser cette signature n'est donc pas un choix judicieux.

Si l'histogramme des couleurs *RGB* (6^3-bins) dans la fenêtre et celui des couleurs normalisées *rb* (12^2-bins) ont des capacités de classification proches lorsque la vérité terrain utilisée est celle correspondante au « sens interdit », la supériorité des couleurs normalisées est perceptible lorsque l'on prend comme référence la carte des détections humaines issues de l'expérimentation oculométrique. Ceci s'explique par le fait que l'histogramme des couleurs normalisées *rb* (12^2-bins) utilisé comme signature, occasionne des fausses détection (au sens de la vérité « sens interdit ») qui correspondent en fait aux distracteurs présents dans la scène routière et détectés par le conducteur (exemple de la chemise bi-colore des Fig. 5.20, 5.22 et 5.23).

En utilisant l'histogramme des couleurs normalisées *rb* (12^2-bins) comme signature, la comparaison des performances du SVM en fonction du noyau utilisé montre que l'utilisation du noyau triangulaire, permet, pour un taux de détection semblable, de réduire le taux de fausses détections de 50% par rapport à l'utilisation du noyau Laplacien, tout en restant corrélé au comportement oculaire lors

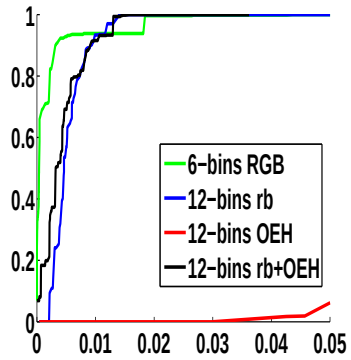
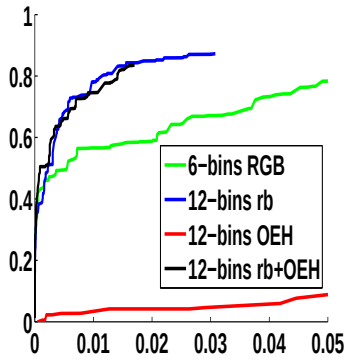
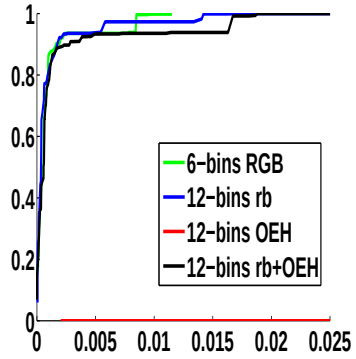
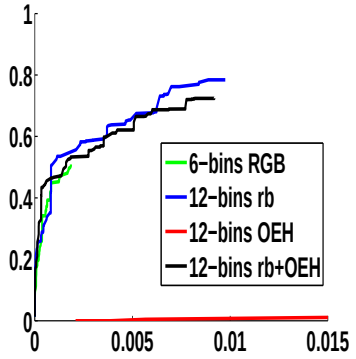
	Vérité « Terrain »	Vérité « Humaine »	Nombre de Support Vecteurs
Noyau Laplacien			28 SV positifs 11483 SV négatifs
Noyau Triangulaire			74 SV positifs 939 SV négatifs

FIGURE 5.24 – Comparaison des courbes ROC obtenues sur l'image originale de la Fig. 5.5, avec différents types de signatures : Histogramme des couleurs *RGB* (6^3-bins), Histogramme des couleurs normalisées *rb* (12^2-bins), Histogramme des orientations du contour (12-bins), Histogramme de la concaténation des couleurs *rb* (12^2-bins) et des orientations du contour (12-bins). Dans la colonne de gauche, la carte de vérité est celle des panneaux de « sens interdit » tandis que à droite, c'est celle des détections en tâche de recherche issue de l'expérimentation. Sur la première ligne, le noyau Laplacien est utilisé, tandis que dans la seconde, c'est le noyau triangulaire.

d'une tâche de recherche.

Par ailleurs, lors de cette étude, l'apprentissage s'est déroulé sur 28 images comprenant 31 panneaux de sens interdit. Après l'extraction des exemples fournis à l'apprentissage, un total de 177 exemples pour la classe positive et 106139 exemples pour la classe négative sont générés. A l'issu de l'apprentissage effectué avec le noyau Laplacien, le nombre de Supports Vecteurs (SV) dans chaque classe est de 28 positifs et de 11483 négatifs. A l'issu de l'apprentissage effectué avec le noyau triangulaire, il y a 74 Supports Vecteurs pour la classe positive, et plus que 939 pour la classe négative. Cette diminution globale du nombre de SV permet au noyau triangulaire d'être 20 fois plus rapide en classification que le noyau Laplacien. D'autre part, la prise en compte de plus de SV positifs et moins de SV négatifs par rapport au Laplacien montre que le noyau triangulaire a mieux adapté l'hyperplan de séparation à la densité des données et profité au mieux de la régularisation dissymétrique.

Nos tests nous prouvent qu'il est inutile de rajouter l'information de forme à la signature et qu'une signature courte, de type histogramme quantifié des couleurs normalisées suffit à avoir de très bons résultats et une corrélation avec les détections des sujets de l'expérimentation. Enfin, le noyau triangulaire, adapté à notre problème, a montré sa supériorité sur le noyau gaussien, dans ce cas précis.

Nous prendrons donc, pour la suite de cette étude, le couple noyau/signature formé du noyau triangulaire et de l'histogramme des couleurs normalisées *rb* comprenant 12^2-bins.

Chapitre 6

Un modèle d'estimation de la saillance de recherche basé sur les SVMs

Les résultats de l'expérimentation présentés dans la partie 5.6.3 montrent que notre algorithme de détection des zones saillantes est corrélé avec les focalisations observées sur le comportement visuel humain. Notre approche cherche à relier la confiance d'un détecteur SVM de panneaux avec la saillance de recherche de la signalisation de police.

6.1 Utilisation des valeurs de confiance du SVM comme estimateur de la saillance intrinsèque

Comme expliqué dans la partie 5.3.3, nous effectuons la détection des panneaux grâce au déplacement d'une fenêtre selon six tailles différentes. Chaque échelle utilisée fournie une carte de décision. Les six cartes sont fusionnées en prenant le maximum de la valeur de classification (décision) du SVM pour chaque pixel sur les six cartes, comme le montre la Fig. 5.5. Nous avons montré que cet algorithme est corrélé au performances visuelles des sujets en termes de détection de zones d'intérêts. Par analogie au comportement visuel des sujets de notre ex-

périmentation, notre algorithme trouve les panneaux de sens interdit, mais aussi certains distracteurs (par exemple, le piéton portant une chemise, à rayure large, rouge et blanche).

Notre idée est de caractériser la saillance à travers une fonction de la confiance du SVM en la détection. Ainsi, nous avons confronté les cartes fusionnées de valeurs de décisions avec les données oculométriques recueillies lors de l'expérimentation exploratoire.

Comme on le montre dans [90], il faut faire attention à quelques caractéristiques du comportement visuel humain. Ainsi, certains panneaux, très saillants, ne sont pas regardés alors qu'ils sont vus (réponse correcte à la question), ce qui montre l'existence de détections en vision para-fovéale et péri-fovéale. Ainsi, comme le montre la Fig.6.1, si on utilise l'information fournie par la réponse du sujet sur le nombre de panneaux comptés dans l'image, on peut discriminer quels panneaux ont été détecté par le conducteur.

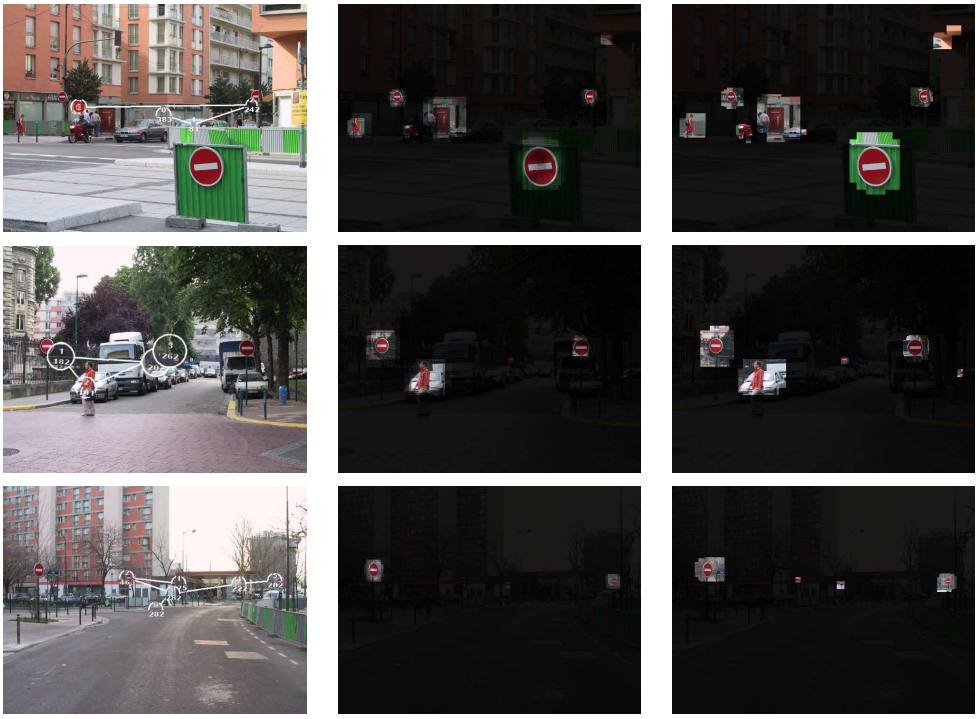


FIGURE 6.1 – A gauche, le chemin visuel des sujets. Au centre, les décisions du SVM sur l'image. A droite, les zone détectées comme saillante (seuillage à 1).

La saillance entretient une relation avec le taux de détection des objets recherchés par les sujets, ainsi qu'avec l'ordre et le temps de fixation. En effet, si on analyse le premier des trois niveaux de tâches de l'action de perception, cf. partie 3.2, dans la description cognitive de l'activité de conduite, moins le panneau est saillant, moins il y a de facilité à le prendre en compte dans la phase de détection. Les panneaux les plus saillants sont donc détectés (fixés) en premier. Or, un tel comportement n'a pas été observé, du fait que cette phase de détection prend moins de 150ms et est effectuée, en partie, par la vision en proche périphérie, qui se repose principalement sur des caractéristiques de couleur, cf. partie 3.1. Ceci ne peut pas être mis en évidence avec la technologie de l'oculomètre à notre disposition, comme expliqué dans la partie 3.4. La saillance de recherche, peut donc être mise en relation avec le taux de détection réalisé par les sujets de l'expérimentation. Malheureusement, lors de l'expérimentation exploratoire, les 38 panneaux contenus dans les 20 images, n'ont pas permis d'induire des taux de détections variables sur les trois sujets valides. En effet, les panneaux étaient, soit de bonne saillance ou de saillance mauvaise, ce qui a conduit a un mauvais échantillonnage des taux de détection. Toutefois, comme montré dans la partie 5.6.3, notre estimateur de saillance de recherche, si l'on seuille les valeurs de décision, est corrélé aux performances des sujets en termes de détections et de fausses détections, ce qui nous conforte dans la cohérence de notre démarche.

Comme on le remarque sur la Fig. 6.2, il y a une corrélation forte entre la durée moyenne de fixation sur chaque panneau et la valeur de décision moyenne calculée par le SVM sur ce panneau. Plus la saillance prédite est basse, plus la durée de fixation est longue. Il est intéressant de noter qu'un seuil au alentour de 280 ms apparaît. Ce seuil, est selon les cognitivistes [129], le temps minimum requis pour prendre une décision. Les valeurs de la Fig. 6.2 correspondent aux valeurs moyennes des durées de fixation des sujets sur les panneaux qui ont été détectés, donc comptés et/ou fixés par les sujets. Une fixation d'une durée nulle correspond à une prise en compte du panneau en vision périphérique par les sujets. De part les seuils fixés lors de la caractérisation des fixations à partir des données brutes d'oculométrie, celles-ci sont considérées nulles si leur durée est en dessous de 100 ms. Le comportement observé, correspond donc à la tâche d'identification (ou reconnaissance), deuxième phase de la description cognitive de l'activité de

conduite. Plus ces caractéristiques visuelles sont dégradées, plus la saillance du panneau est mauvaise, et plus le temps nécessaire à la prise de décision sur la reconnaissance augmente. Toutefois, pendant toute la durée de la fixation, le sujet est focalisé sur le panneau qu'il a, au préalable, détecté. Le sujet procède donc à ce moment là à l'analyse des caractéristiques visuelles de l'objet visé. La position du panneau dans la scène routière, ainsi que son fond environnant n'est alors plus pris en compte dans cette phase. On peut donc considérer que la durée de fixation et par leur relation, la valeur de décision du SVM, correspondent à la saillance intrinsèque du panneau. Nous définissons ici la saillance intrinsèque comme la capacité du panneau à être identifié indépendamment du fond l'environnant. Elle est directement relié à la facilité d'analyse du panneau pour l'usager l'ayant au préalable détecté et étant fixé dessus. La carte finale des décisions du SVM définit notre *Carte de Saillance Intrinsèque* (*CSI*).

La valeur de décision du SVM, est un premier estimateur de la saillance intrinsèque des panneaux (objets appris et recherchés). Or, notre objectif est de modéliser la saillance de recherche des panneaux, ce qui correspond à la facilité de détection de ce dernier pour un fond donné. Il nous faut donc aussi prendre en compte, dans le modèle la complexité de la scène, c'est à dire le fond environnant le panneau ainsi que sa taille afin de caractériser sa propension à être détecté.

6.2 Prise en compte du fond environnant

Comme expliqué dans la partie précédente, la carte des décisions du SVM représente la saillance intrinsèque de la signalisation verticale. Toutefois, la saillance de recherche d'un panneau peut être associé à la facilité de sa reconnaissance (sa saillance intrinsèque). Pour la saillance de recherche, il faut aussi prendre en considération le fond complexe de la scène routière, sur lequel le panneau a un niveau de propension à se détacher. En effet, pour un panneau donné, la valeur de saillance calculée ne doit pas être la même en fonction des fonds de complexités différents sur lequel le panneau est placé.

Nous proposons donc de soustraire la valeur moyenne des décisions sur toute l'image. Lorsque l'on compare les nouvelles valeurs de décisions obtenues, avec

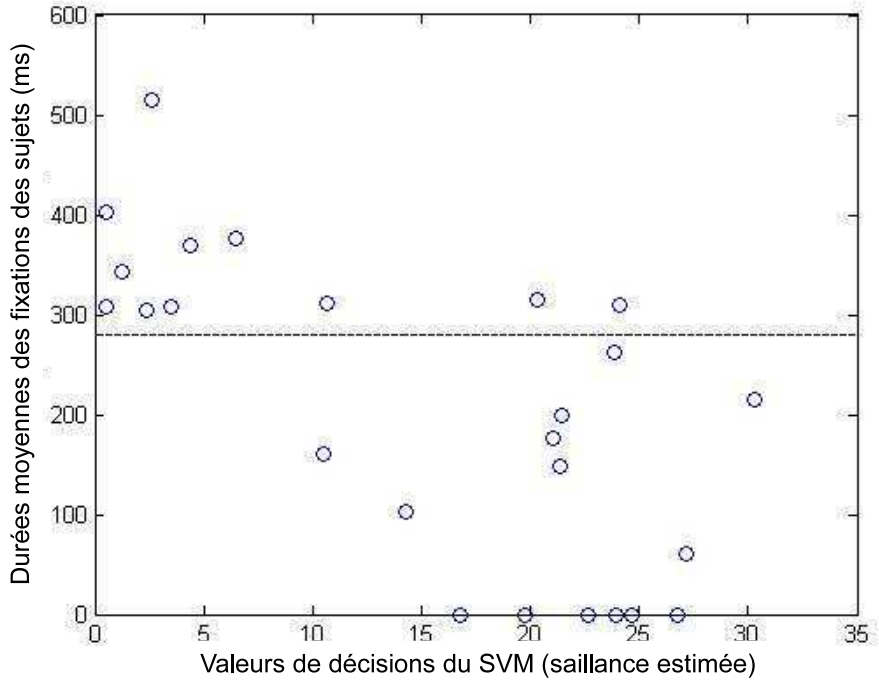


FIGURE 6.2 – Moyenne des durées de fixation des sujets en fonction des valeurs prédites de saillance sur les panneaux. Une durée nulle correspond à la prise ne compte du panneau en vision périphérique par le sujet.

les anciennes, sans la compensation par le fond, au taux de détections des panneaux par les sujets lors de l'expérimentation 7.1, on perçoit une légère augmentation de la corrélation des valeurs de saillance prédites avec la propension à être vue des panneaux. Mais ce résultat n'est pas optimal, du fait que lorsque l'on prend en compte la globalité de l'image pour calculer sa saillance moyenne, les autres panneaux et distracteurs de forte saillance, ont pour effet d'augmenter la moyenne et donc agissent en facteur négatif sur la saillance des autres objets, ainsi que sur leur propre saillance. De même, un fond qui à une valeur de décision négative (formellement reconnu comme n'étant pas un panneau), spatialement éloigné d'un panneau, augmente la valeur de saillance de ce dernier, alors qu'ils ne peuvent pas être perçus et analysés en même temps lorsque l'on considère le comportement humain.

Nous avons résolu le problème par une approche plus locale. Un premier essai, infructueux, prenant une taille de fond en fonction de la taille de la zone pré-détectée comme saillante, nous a prouvé que cette taille se devait d'être fixe. En effet, les petites zones, sont dans ce cas, comparées à un fond très proche et les grandes à un fond très large. Or dans le système visuel humain, la taille du fond pris en compte autour du point de fixation reste fixe car dépend de la physiologie de l'œil, cf. partie. 3.1.

Nous avons effectué des tests de corrélation de la saillance prédite avec le taux de détection des panneaux par les sujets, en faisant varier, comme paramètre, la taille du fond pris en compte autour des objets détectés comme potentiellement saillants (avec la valeur de décision du SVM seule). Nous avons obtenu le meilleur taux de corrélation pour une taille de fond pris en compte correspondant à 1° d'angle visuel (i.e. 32 pixels pour nos images) autour de l'objet pré-détecté, comme le montre la Fig. 6.3. La carte compensée par le fond environnant est définie comme la *Carte de Saillance relative au Fond* (*CSF*), c'est-à-dire une compensation de la *CSI* (*Carte de Saillance Intrinsèque*).

6.3 Prise en compte de la taille de la signalisation verticale

Il est connu qu'un des principaux facteurs influençant leur saillance est la taille des objets routiers [130]. La saillance de recherche est liée à la facilité de détection de celle-ci dans l'environnement. Lors des passations de l'expérimentation, décrite ci-après, pendant la phase d'évaluation de la visibilité des panneaux, nous nous sommes aperçus, que conformément aux constituants de la saillance, les sujets modulaient leurs réponses en fonction de la taille des panneaux. Par l'étude statistique de l'expérimentation présentée dans le chapitre 7, il est montré que la saillance estimée par les sujets est dépendante de la taille des panneaux, or notre CSF (*Carte de Saillance relative au Fond*) est, par nature, indépendante de la taille des objets. Ainsi, il nous faut aussi la prendre en compte afin de définir la *Carte de Saillance de Recherche* (*CSR*).

Pour ce faire, nous nous sommes inspirés de la loi de Ricco. Cette loi psycho-

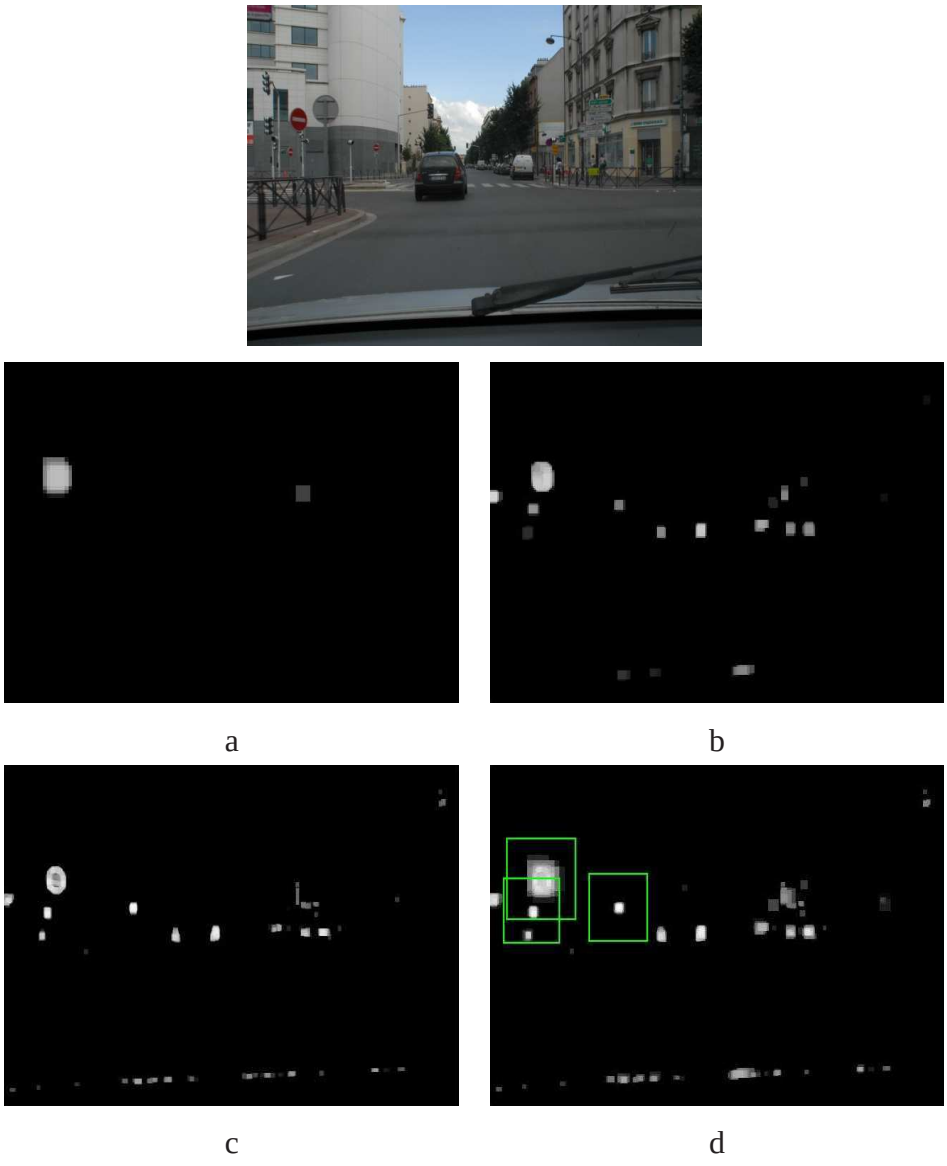


FIGURE 6.3 – Cartes de décisions obtenues sur l'image de la première ligne, à différentes échelles : (a) 40×40, (b) 20×20, (c) 10×10, et (d) carte finale de saillance des « sens interdit ». Les carrés verts montrent la fenêtre de fond de 2° d'angle visuel prise en compte pour la compensation.

physique de la sommation spatial décrit que le seuil L de contraste nécessaire à la détection d'une zone de lumière dépend de l'aire A du stimulus :

$$k = L \times A^n \tag{6.1}$$

où k est une valeur constante et n exprime la condition selon laquelle la sommation spatiale est complète ($n = 1$) ou si elle est partielle ($0 < n < 1$). Dans notre cas, $n = 1$ et nous proposons de remplacer L par la valeur de saillance issue de la *CSF* et ainsi de modéliser la *CSR* par une fonction de la *CSF* et de l'aire A de la zone pré-détectée comme saillante.

A l'issue de l'expérimentation, nous disposions des scores donnés par les sujets (*SSS*), pour chaque panneau i. Ces scores sont eux même corrélés au taux de détection des panneaux par les sujets, donc à la saillance de recherche. Une régression linéaire entre les $log(SSS(i))$, $log(CSF(i))$ et $log(A(i))$, nous a conduit au modèle suivant :

$$CSR(i) = \sqrt[4]{CSF(i) \times A(i)} \tag{6.2}$$

où $A(i)$ est l'aire du panneau i.

6.4 Calcul de saillance des objets routiers à partir du modèle proposé

Comme le montre la vue d'ensemble de la Fig. 6.4, la carte de saillance de recherche *CSR*, définie par l'équation 6.2, représente l'application de notre modèle sur l'ensemble de l'image de la scène routière. Afin de calculer la valeur de saillance d'un objet i donné, nous calculons la valeur moyenne de la carte sur la surface de l'objet. Nous appliquons cette méthode aux panneaux afin d'estimer leur saillance.

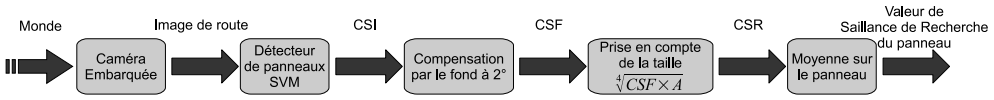


FIGURE 6.4 – Vue d'ensemble du modèle de saillance de recherche proposé.

Comme la *CSR* est une combinaison de l'aire $A(i)$ de l'objet détecté comme saillant et de la CFS, la taille de la signalisation est naturellement prise en compte dans le modèle. Rappelons que la CSF, carte de saillance relative au fond, est par construction, la compensation, par la valeur du fond qui entoure l'objet détecté, de la carte de saillance intrinsèque *CSI*. La CSF mesure la dissimilarité entre la saillance intrinsèque du fond et celle du panneau, c'est à dire la propension à différencier le panneau de son fond. L'utilisation combinée de la CSF et de la *CSI* peut permettre de caractériser, à l'intention du gestionnaire, si un éventuel défaut de saillance est dû au panneau en lui même ou à une trop grande similarité du fond environnant.

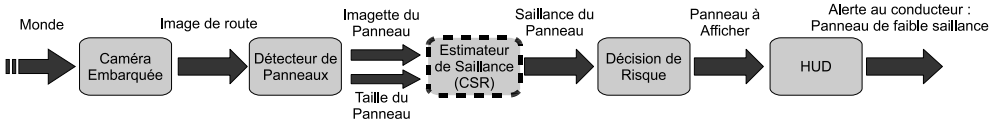


FIGURE 6.5 – Vue d'ensemble du système d'assistance à la conduite (ADAS) proposé autour de notre modèle de calcul de la saillance (en pointillé).

Comme nous l'avons montré dans [114], si on utilise l'algorithme afin de détecter la signalisation verticale et les autres points saillants en estimant dans le même temps leurs saillances sur l'intégralité de l'image, le résultat n'est pas donnée en temps réel. Le temps de calcul total (avec six tailles de fenêtres) sur une image de taille 1280 est de l'ordre d'une minute. Aussi, comme nous l'avons proposé dans [114], si on veut utiliser l'information de saillance dans un système avancé d'assistance à la conduite (ADAS), donc en temps réel, cela est possible en réduisant le calcul de la CSR au seuls panneaux détectés. Comme le montre la Fig. 6.5, il faut coopérer, en amont, avec un bon détecteur de signalisation verticale (taux de détection proche de 1) qui extrait de l'image de la scène routière

les panneaux et leurs fonds dans des imagettes, ainsi que la taille des panneaux. Une fois la saillance du panneau évaluée par notre algorithme, une décision de risque basée sur un seuillage de la valeur de saillance permet de classer si le panneau est considéré comme suffisamment saillant ou, si, au contraire, il a une faible saillance et donc risque de ne pas être pris en compte par le conducteur. Dans ce cas, le panneau est affiché au conducteur. A l'instar des systèmes d'assistance à la conduite, qui équipent certaines voitures actuelles (haut de gamme), capable d'afficher tous les panneaux de vitesses, cet affichage peut se trouver sur le tableau de bord, comme dans la nouvelle Opel insignia, voir même également sur un affichage tête haute (HUD), en option sur la dernière version de la BMW série 7, comme le montre la Fig. 6.6. Dans notre cas, le but est de n'afficher que les panneaux qui n'ont pas le niveau de saillance requit pour une bonne notification par l'usager. Il est en effet important de n'afficher que le strict nécessaire (et donc d'avoir un seuil d'alerte), car dans la cas contraire, vu le nombre élevé de signaux routier le long du réseau, il ne fait aucun doute que l'affichage trop fréquent d'alerte gêne le conducteur et ce dernier à terme ne les prend plus en compte du fait de l'accoutumance.



FIGURE 6.6 – ADAS indicateur de vitesse. A gauche sur l'opel insignia. A droite, sur la BMW série 7, en haut HUD, en bas affichage classique. Source http ://www.opel-insignia.fr/ et http ://www.bmw.fr/.

Chapitre 7

Expérimentation de validation par oculométrie cognitive

L'expérimentation exploratoire, présentée dans la partie 5.6.2, a montré la cohérence de notre démarche. Toutefois, certains points, non maîtrisés, ont été identifiés dans cette expérimentation :

- Le sujet n'avait pas de temps limite imposé pour effectuer la recherche des panneaux, ce qui conduisait à la détection de panneaux peu saillant et une augmentation de la durée du chemin visuel. De plus, ils pouvaient faire des aller-retours sur les objets saillants de l'image, ce qui rendait difficile le traitement des données oculométriques. ainsi, nous avons appris de cette expérimentation qu'il faut donc limiter le temps à disposition du participant lors de la recherche des panneaux et le prévenir de cette contrainte temporelle. Il est à noter que cette consigne reflète la nécessité d'une prise rapide de décision pendant l'activité de conduite.
- L'expérimentation exploratoire a été faite en conditions non optimales au regard des spécificités de notre matériel d'oculométrie. En effet, comme expliqué dans la partie 3.4.1, une fois la phase de calibrage effectuée, le sujet doit minimiser ses mouvements afin de ne pas dé-paramêtrer le système oculomètrique. Dans notre cas, les sujets étaient assis sur une chaise devant une table. Or, ils ont eu tendance à bouger au cours de la passation, du fait de l'inconfort de la chaise, ce qui a eu pour effet d'occasionner une dérive

dans les mesures. Cette dérive, non quantifiable, ne pouvait alors pas être corrigée.

– La consigne était de compter le nombre de panneaux de « sens interdit », tout en se mettant en condition de conducteur. Pourtant, les photos utilisées ont été prises par un piéton, ce qui pouvait occasionner un étonnement de la part du sujet, censé imiter un conducteur.
– Enfin, avec ce protocole, nous avons obtenu seulement les réponses de trois sujets concernant le nombre de panneaux comptés dans vingt images, ainsi que leurs fixations oculaires correspondantes. Afin que les comportements observés soit significatifs, il est nécessaire d'avoir plus de sujets et plus d'images.

Sur la base de ces différent points, nous avons mis au point un nouveau protocole [1], plus complet et mieux contrôlé.

7.1 Méthodologie

Dans le protocole exploratoire, l'unique tâche était de compter les panneaux de « sens interdit » dans les images, ce qui plaçait les sujets dans une activité de recherche que l'on peut raccordée aux processus attentionnel de type « top-down ». Ne connaissant pas *a priori* le poids relatif des processus attentionnels de type « top-down » et « bottom-up » qui jouent dans la prise en compte de la saillance de la signalisation de police chez les conducteurs, (cf. partie 3.3.3), nous avons mis en place un protocole qui, en plus de corriger les défauts exposés ci-dessus, nous permet d'explorer les deux types d'attention. De plus, il nous est également apparu nécessaire de récolter des mesures objectives et subjectives de la saillance de la signalisation de police.

Le sujet est placé devant un écran sur lequel les images sont affichées. Les images présentées aux sujets sont des scènes routières, avec des intersections de voies ou non, contenant ou non des panneaux de tous types.

1. Avec l'aide de Christian Bastien, maître de conférences - HDR en ergonomie cognitive, Université de Metz et Anaïs Mayeur, doctorante au LCPC en ergonomie cognitive.

Dans une première phase, la consigne propose au sujet de se placer dans une situation de conduite. Nous nous inspirons, pour cela, des questions formulées lors du passage à l'épreuve théorique générale du code de la route. Ainsi, la consigne de l'expérimentation amène le sujet à analyser la scène routière sans avoir d'*a priori* sur les éléments à étudier.

Dans une deuxième phase, le protocole de l'expérimentation exploratoire est repris et corrigé de ses limites. La consigne induit la mobilisation de la saillance de recherche par le sujet, c'est-à-dire de compter le nombre de panneaux de sens interdit présent dans la scène affichée. Le sujet doit également se placer dans une situation de conduite et effectuer ce comptage le plus rapidement possible. La réponse des sujets sur le nombre total de panneaux par image est enregistrée.

Le trajet visuel du sujet enregistré par l'oculomètre, après analyse, permet d'obtenir un descripteur objectif de la saillance.

Dans une troisième phase, il est demandé au sujet d'évaluer la saillance de chacun des panneaux indiqués. Deux notations, sur une échelle de 0 à 10, sont effectuées par les sujets. Pour la première, chaque panneau est présenté seul, à son emplacement dans la scène routière, mais sur un fond gris uniforme. Pour la seconde, le panneau est présenté, en même temps que les autres, en place dans la scène routière. Cette évaluation nous permet de définir un descripteur subjectif de la saillance perçue par les individus.

7.1.1 Matériel et dispositif expérimental

Les expérimentations sont réalisées dans une salle photométriquement contrôlée (salle CLOVIS du LCPC) équipée d'un écran LCD, d'un ordinateur qui exécute le protocole expérimental et d'une interface clavier-souris qui permet d'enregistrer les réponses des sujets, cf. Fig. 7.2 et 7.1. A cela, s'ajoute un système oculométrique binoculaire non-invasif, de marque S.M.I. iViewX^{in} RED, positionné sous l'écran de visualisation des images, et utilisé pour enregistrer les fixations oculaires des participants durant les deux premières phases de l'expérimentation. Les positions du centre du regard sont calculées avec une fréquence d'échantillonnage de 50 Hz. La précision sur la position du regard du système oculométrique est inférieure à 0.5°. L'utilisation des reflets cornéens, ainsi que le

suivi des pupilles, permet une mesure « stable » et réduit les erreurs causées par les petits mouvements de la tête. L'ordinateur de l'eye-tracker est synchronisé et commandé par celui qui exécute le protocole expérimental, via l'interface réseau. Un siège ergonomique permet de limiter les mouvements involontaires des sujets, ce qui optimise le recueil des données oculométriques et facilite leur exploitation.

Lieu de l'expérimentation : Avec son palindrome récursif, CLOVIS est un LOcal de VISualisation photométriquement contrôlé. Comme le montre la photo de la Fig.7.2, les murs sont entièrement noirs, ainsi l'attention du sujet ne peut se focaliser que sur l'écran de visualisation de l'expérimentation et la photométrie peut être contrôlée. Deux ordinateurs permettent, d'une part d'exécuter le logiciel *Presentation*^{TM}, et d'autre part de gérer l'oculomètre. Le premier (*Presentation*^{TM}) commande le deuxième (Oculomètre) via le port Ethernet.

Logiciels : Les images sont affichées sur un écran LCD via le logiciel *Presentation*^{TM}. Ce logiciel permet de réaliser des expérimentations psychovisuelles suivant un scénario déterminé, et d'enregistrer la suite des actions et des réponses. Il permet, par ailleurs, de commander l'oculomètre qui est géré par son propre ordinateur.

Les données de l'oculomètre sont par la suite traitées avec les logiciels Begaze (fournis avec l'oculomètre) et OGAMA (logiciel libre), qui permettent d'analyser les chemins visuels (scanpath), les points de fixations et leurs durées en fonction des paramètres de taille de la zone (en pixel) et de la durée minimum de point de regard dans celle-ci (en ms). Ces deux paramètres sont à déterminer pour le traitement des résultats. Ces logiciels permettent également de récolter des données statistiques en fonction de zones d'intérêts définies par l'utilisateur.

L'analyse statistique et comparative des indicateurs subjectifs et objectifs de saillance avec les valeurs prédites par notre modèle est réalisée sous STATA.

Dispositif d'affichage : Le sujet visualise les images sur un écran LCD de 17 pouces. Les images sont affichées avec une résolution de 1280 x 1024. La

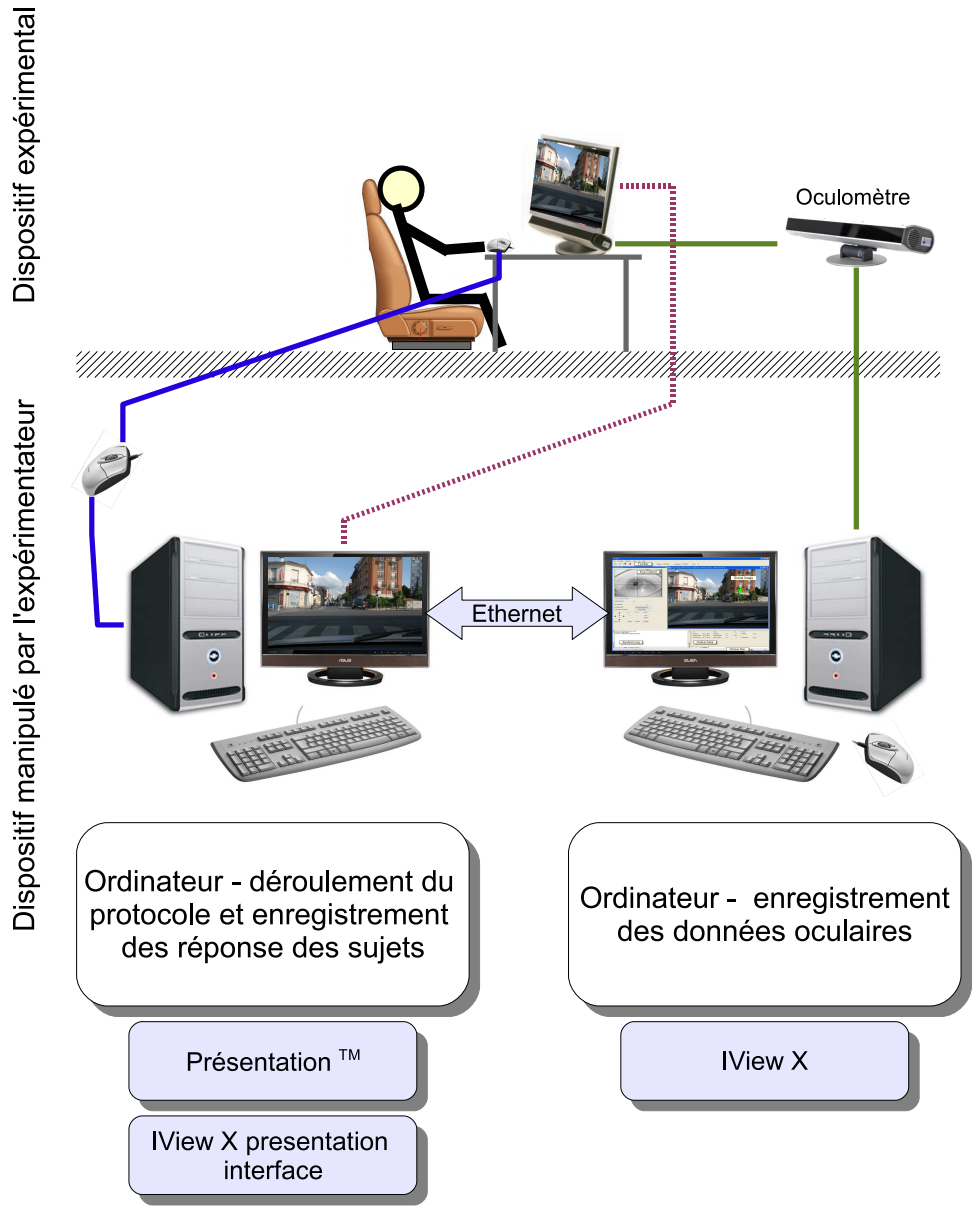


FIGURE 7.1 – Dispositif mis en place dans la salle CLOVIS.



FIGURE 7.2 – CLOVIS, local de visualisation photométriquement contrôlé.

scène routière est centrée et affichée sur 640 x 480. Le pourtour de l'image, situé entre le bord de l'image de la scène routière et le bord de l'écran est l'adaptation de luminance de la scène, correspondante au niveau de gris moyen de l'image centrale.

Oculomètre : Un oculomètre de chez Senso Motoric Instument, le Iview X RED a été utilisé. Celui-ci, comme le montre la photo de la Fig. 7.3 est positionné sous l'écran de visualisation grâce à un support dédié, ce qui permet une reproductibilité de l'expérimentation. Durant chaque expérimentation, la position du regard du sujet est suivie et enregistrée par l'oculomètre et le trajet visuel et les fixations sur l'image sont mesurés.

Les spécifications techniques de l'oculomètre sont :

- Fréquence d'échantillonnage de 50 Hz.
- Résolution de suivi de la pupille inférieur à 0,1 degré.
- Précision de la position du regard inférieur à 0,5 degré.

- Distance entre le sujet et la caméra de 50 à 70 cm.
- Suivi de la tête dans un cadre de 40x40 cm à 70 cm de distance.

Siège : Le siège utilisé est un siège ergonomique de voiture, Recaro ergomed, équipé de soutiens latéraux et appuie tête afin que le sujet soit dans une position agréable. Pour un confort et un maintien optimum, deux accoudoirs sont rajoutés sur ce siège. De ce fait, le sujet ne bouge pas pendant la durée de l'expérimentation, ce qui permet de recueillir des données oculométriques exploitables. Suite à des modifications, le siège est réglable en hauteur et équipé d'un système de blocage afin que le sujet ne puisse pas bouger sa position spatiale dans la pièce pendant l'expérimentation.

Position du sujet : La hauteur de l'écran et/ou du siège sont réglés de façon à ce que l'axe orthogonal au centre du plan de l'écran soit confondu avec l'axe visuel naturel du sujet. L'expérimentateur effectue le calibrage de l'oculomètre en proposant au sujet de fixer des croix idoines.

Le siège est placé à une distance de 60 cm devant l'écran, ce qui permet d'obtenir un champ angulaire de 20° couvrant la largeur de l'écran et de garantir les conditions optimales pour l'oculomètre.

Cette disposition permet de proposer les images de scènes routières d'une taille angulaire de 10°, c'est-à-dire en vision para-fovéale, ce qui évite les saccades oculaires de grandes amplitudes vers la périphérie.

Périphérique de saisie : Les différentes réponses possibles proposées aux sujet, et cela pour chaque phase, sont présentées sur l'écran. Le sujet n'a à sa disposition que la souris. Il clique sur la ou les réponse(s) voulue(s) pour répondre aux questions de l'expérimentateur. Cette façon de recueillir les données permet d'éviter que le sujet détourne le regard de l'écran de visualisation ou qu'il bouge la tête, ce qui occasionnerai une perte de précision, voir du suivi du regard par l'oculomètre.

Type d'images affichées : Les images affichées représentent des scènes routières

urbaines prises du point de vue du conducteur dans une voiture, par un appareil photo professionnel (Nikon D200) en haute résolution, comme le montre la Fig. 7.4.

Outre les scènes routières spécifiques à la ville, les images contiennent tous les types de panneaux de polices et sont scindées en quatre catégories selon leur contenu, cf. Fig. 7.4 :

– des panneaux d'interdiction seuls (21 images),
– tous types de panneaux (15 images),
– différents panneaux, sauf des panneaux de sens interdit (3 images),
– aucun panneau (1 image).



FIGURE 7.3 – L'oculomètre Iview X RED de marque SMI, monté sous son écran d'affichage dédié.

7.1.2 Participants

Cette expérimentation fait appel à 40 sujets. Afin de sélectionner les sujets participants à l'expérimentation, nous avons retenu comme critères d'inclusion, qu'il pouvaient être de tout genres, de tout ages, avec une vision correcte, même corrigée. Le premier critère d'exclusion est le fait que les sujets devaient avoir plus de

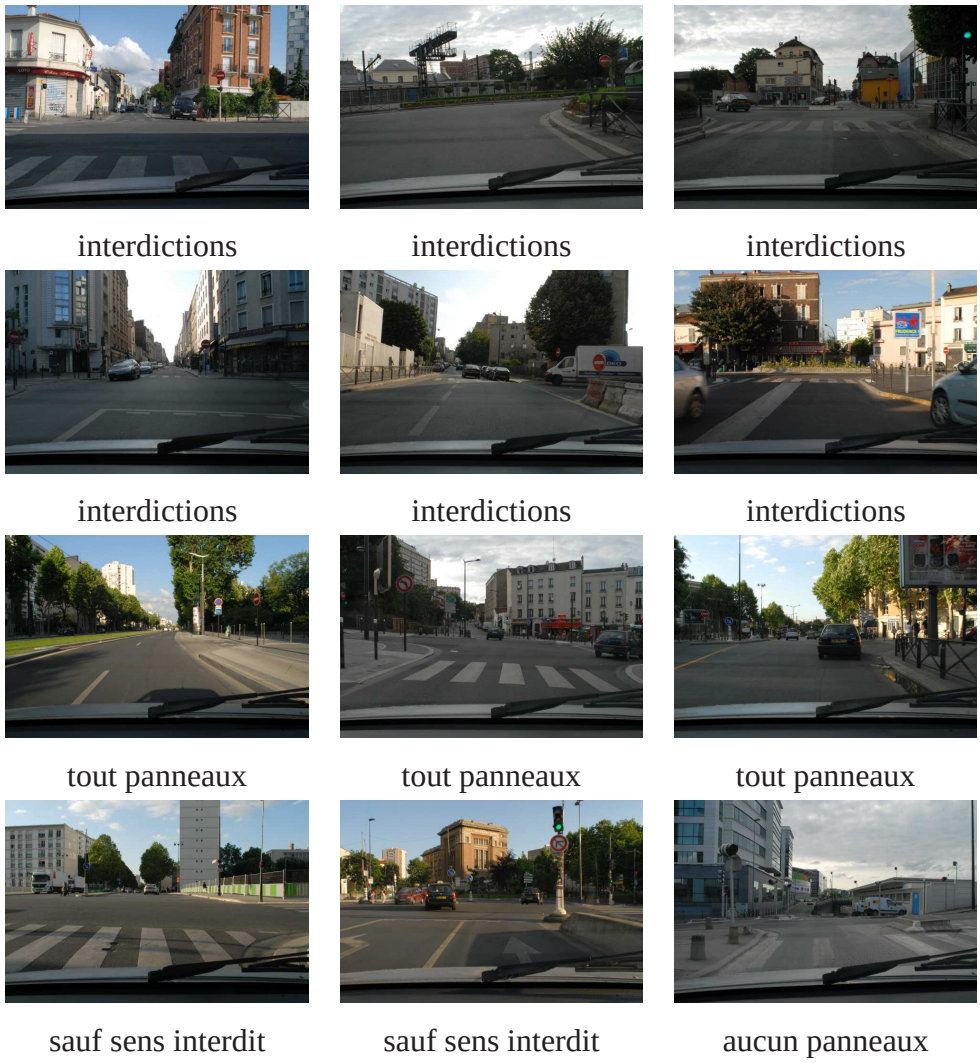


FIGURE 7.4 – Exemples de photos utilisées lors de l'expérimentation oculométrique selon leurs classes.

cinq ans de permis. En effet, l'expérimentation devant répondre au comportement d'un automobiliste moyen, on s'assure d'une homogénéité des sujets.

Le deuxième critère d'exclusion est le fait que les sujets ayant des problèmes oculaires de vergence ne peuvent pas être participants, du fait des spécificités technologiques de l'eye-tracker utilisé, qui ne permet pas le suivi du regard de ces sujets. Les données des sujets qui portent des verres correcteurs progressifs ou double foyer (i.e. atteints de presbitie), du fait de la modification du chemin optique qui fausse le reflet cornéen, ne peuvent pas être analysées. Du fait de l'approche couleur, des pathologies oculaires telles que le daltonisme, faisaient également partie des critères d'exclusion.

7.1.3 Procédure

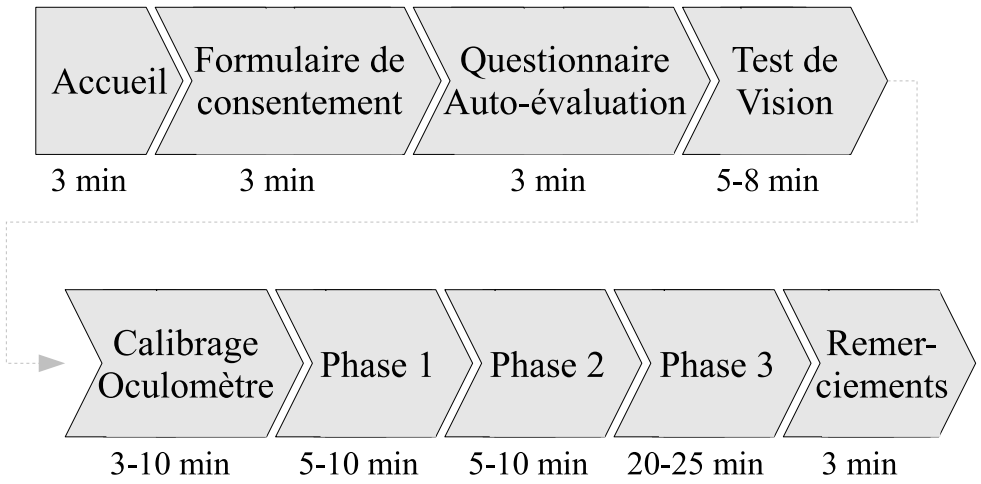


FIGURE 7.5 – Déroulement d'une passation de l'expérimentation. La durée totale varie de 50 à 75 minutes.

Le déroulement de la procédure de passation de l'expérimentation est présenté dans la Fig. 7.5. Selon la co-opérabilité du sujet, la durée totale de sa passation varie entre 50 min et 75 min. Le sujet est d'abord accueilli par l'expérimentateur qui lui présente le lieu de l'expérimentation. Il est convié à prendre connaissance du formulaire de *Consentement de participation à une recherche scientifique* (Fig.

7.6), à le remplir en deux exemplaires, puis à le signer. Le formulaire est ensuite contre-signé par l'expérimentateur.

Le sujet est ensuite invité à remplir le questionnaire d'auto-évaluation (Fig.7.7) qui permet de contrôler d'éventuels problèmes au niveau des résultats. L'ensemble des données du sujet sont inscrites, par lui même, sur des fiches comportant un numéro unique par sujet. Le numéro est attribué dans l'ordre de passation des sujets.

Le sujet répond aux questions suivantes :

- Nom
- Prénom
- Adresse email
- Age
- Sexe (M / F)
- Type(s) de permis de conduire obtenu(s)
- Date d'obtention du permis
- Possession d'une voiture ?
- Modèle de la voiture
- Fréquence d'utilisation (Rarement / Souvent)
- Période d'utilisation (Semaine / Week-end / Quotidiennement)
- Nombre moyen de Km par semaine et/ou par an
- Lieux de conduite fréquent (Urbain / Péri-Urbain / Rurale / Mixte)
- Véhicule conduit le plus souvent (Camion / Auto / Moto / Vélo)
- Vue (Normale / Corrigée)
- Raison de la correction
- Dispositif(s) de correction (Lunette / lentille)
- Vision corrigée en conduite (Oui / Non)

Ensuite, le sujet passe un test de vision effectué au moyen d'un appareil, Visio-Test, couramment utilisé par les médecins. Cet appareil permet de déterminer l'œil directeur du sujet, de mesurer l'acuité et de caractériser d'éventuels problèmes de vision non connus du sujet (vergence, daltonisme ou autre). Cela évite de faire passer l'expérimentation à un sujet dont on sait que l'on ne peut pas exploiter les données.

LCPC Laboratoire Central des Ponts et Chaussées

DESE-EV

Consentement de participation à une recherche scientifique

Juin/juillet 2008

Ludovic Simon,
Jean-Philippe Tarel

Je soussigné(e) :

.....

déclare accepter de participer à la recherche intitulée :

Oculométrie cognitive et perception de l'environnement routier par les automobilistes, organisée par L. Simon & J.-P. Tarel dans les conditions précisées ci-dessous.

J'ai reçu les informations détaillées oralement sur le déroulement de l'expérience, qui se résume ainsi : il s'agit de tester un modèle informatique de perception.
L'ensemble de l'expérience dure environ 1h15.

Il m'a été précisé que :
- Je suis libre d'accepter, de refuser, ou bien d'arrêter à tout moment de ma participation.
- Les données qui me concernent resteront strictement confidentielles. Je n'autorise leur consultation que par des personnes qui collaborent avec L. Simon & J.-P. Tarel.
- Je pourrai à tout moment demander des informations à L. Simon (01 40 43 65 56) ou J.-P. Tarel (01 40 43 53 76).
- La publication des résultats de la recherche ne comportera aucun résultat individuel identifiant.

Mon consentement ne décharge pas les organisateurs de la recherche de leurs responsabilités. Je conserve tous mes droits garantis par la loi.

Fait à ..... le .....

Nom et signature de l'expérimenteur Signature du participant

(un exemplaire co-signé doit être remis à la personne qui participe)

L'esprit de recherche au cœur des réseaux

FIGURE 7.6 – Consentement de participation à une recherche scientifique signé par les sujets.

LCPC Laboratoire Central des Ponts et Chaussées

DESE-EV

Questionnaire

Juin/juillet 2008

Ludovic Simon,
Jean-Philippe Tarel

Madame, Monsieur,

Vous participez à l'expérimentation dont l'objectif est de récolter des données oculométriques et des réponses comportementales sur la perception humaine en conduite, afin de tester des modèles informatiques.

Pour compléter la phase de test que vous allez effectuer, nous vous proposons maintenant de répondre aux quelques questions qui suivent.

N° participant : _____

Nom : .....

Prénom : .....

E-mail : .....

Age : .....

Sexe : M / F

Type de permis de conduire obtenus : A1 / A / B / C / D / Eb / Ec / Ed

Date d'obtention : .....

Vous possédez une ou des voiture(s) : oui / non

Quel(s)modèle(s) ? .....

Fréquence d'utilisation : rarement / souvent

Période d'utilisation : semaine / week-end / quotidiennement

Nombre moyen de km par semaine : ..... et / ou par an : .....

Lieux de conduite fréquent : Urbain / Péri-Urbain / Rurale / Mixte

Véhicule conduit le plus souvent : camion / auto / moto / vélo / autre

Votre vue est : normale / corrigée

Si vision corrigée, pourquoi ? .....

Dispositifs de correction : Lunettes / Lentilles

Votre vision est elle corrigée en conduite : oui / non

L'esprit de recherche au cœur des réseaux

FIGURE 7.7 – Questionnaire rempli par le sujet avant le début de l'expérimentation.

Enfin, le sujet est invité à prendre place dans le siège, pendant que l'expérimentateur lui explique le déroulement de l'intégralité de l'expérimentation, à savoir, le calibrage de l'oculomètre, puis deux phases successives pendant lesquelles il doit limiter au maximum ses mouvements et enfin une troisième et dernière phase pendant laquelle il est autorisé à reprendre un peu de mobilité.

Première phase : recueil de données comportementales sur la saillance attentionnelle

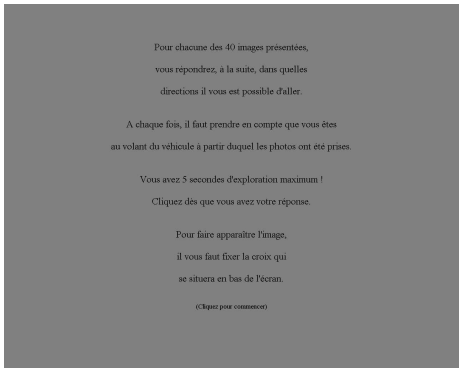
Comme le montre la Fig.7.1, le sujet est assis devant l'écran avec sa main posée sur la souris, en position d'attente. Une fois le calibrage de l'oculomètre effectué, l'expérimentateur lance le scénario *Presentation*^{TM}. En parallèle, l'oculomètre reçoit automatiquement l'ordre de se préparer à l'enregistrement des positions du regard du sujet et de passer en mode pause. Il n'enregistre les données oculaires que le temps de l'affichage des images, ce qui permet de délimiter automatiquement les données correspondantes à chaque image et de limiter le volume des données, ceci afin de faciliter l'analyse future des données.

L'expérimentateur explique alors au sujet que l'expérimentation n'est en aucun cas un test d'aptitude et que ses résultats ne seront pas divulgués. De plus, le sujet est aussi avisé du fait qu'il peut, s'il le désire, à tout moment effectuer une pause entre deux images. Dans tous les cas, même s'il effectue une pause, il lui est demandé de ne pas effectuer de mouvements pendant cette pause, ni de grand mouvements de tête.

Une première consigne, décrivant l'expérimentation s'affiche (Fig. 7.8), tandis que l'expérimentateur explique la procédure. Lorsqu'il est assuré de la bonne compréhension de celle-ci par le sujet, il l'invite à cliquer pour débuter par un exemple :

Pour chacune des 40 images présentées,
vous répondrez, à la suite, dans quelles
directions il vous est possible d'aller.

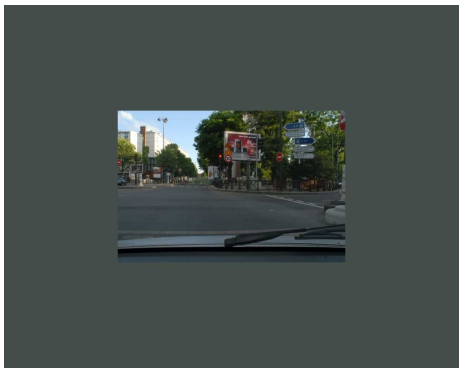
A chaque fois, il faut prendre en compte que vous êtes
au volant du véhicule à partir duquel les photos ont été prises.



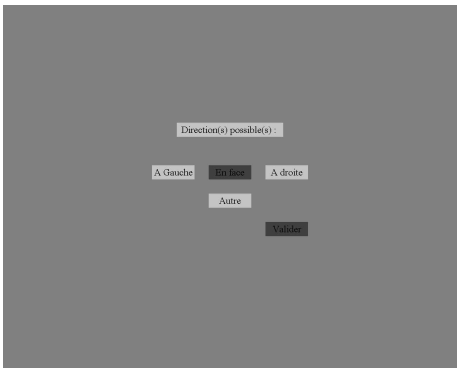
Consigne



Croix de fixation



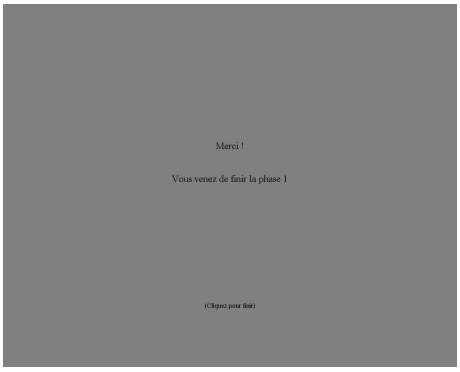
Scène routière



Écran de réponse



Écran de transition



Écran de fin

FIGURE 7.8 – Déroulement de la première phase de l'expérimentation.

Vous avez 5 secondes d'exploration maximum !
Cliquez dès que vous avez votre réponse.

Pour faire apparaître l'image,
il vous faut fixer la croix qui
se situera en bas de l'écran.

(Cliquez pour commencer l'expérimentation)

Pour commencer l'expérimentation, le sujet clique sur l'écran de consigne. Il commence par une image d'exemple afin que l'expérimentateur puissent s'assurer de la bonne compréhension de la consigne par le sujet. La consigne est à nouveau affichée, puis le sujet clique pour commencer l'expérimentation. Une série de 40 images de scènes routières lui sera présentée.

Le premier écran qui apparaît est constituée d'une croix blanche centrée en bas de l'écran sur un fond gris (Fig.7.8). Le niveau de gris est le niveau de gris moyen de l'image de la scène routière qui va être affichée juste après, ceci afin de s'assurer de l'adaptation du sujet à la luminance. La taille de la croix est de 16 pixels de large et de hauteur, ce qui représente $0,5$ degrés d'angle visuel. (précision de l'eye tracker $0,5^{\circ}$). Cette croix de fixation est placé à cet endroit afin de restituer le mouvement oculaire naturel d'un automobiliste qui quitte des yeux son tableaux de bord (compteur de vitesse) et analyse la scène routière dans laquelle il évolue. La croix est affichée jusqu'à ce que le sujet effectue dessus une fixation consciente d'au moins 500 ms. La fixation sur la croix, ou une fixation sur une zone de quatre degrés autour de celle-ci, déclenche l'apparition de l'image de scène routière, ce qui nous assure que la position de départ du chemin visuel sera la même pour toutes les images et pour tous les sujets.

Puis, une image de la scène routière, entourée de son fond gris (exemple Fig. 7.8) apparaît. Par décision du sujet ou après un temps d'apparition maximal de cinq secondes l'image disparaît. Le sujet peut décider de faire disparaître l'image

avant les cinq secondes, pour donner sa réponse, en cliquant sur la souris.

L'écran de réponse, présenté dans la Fig. 7.8 apparaît, et propose au sujet de sélectionner, avec la souris, sa ou ses réponse(s). Le sujet doit alors sélectionner la ou les direction(s) possible(s) que la scène routière lui permettait de prendre (A gauche ; En face ; A droite ; Autre). Une fois la réponse validée, un écran apparaît, (Fig. 7.8) qui indique au sujet qu'il doit cliquer pour accéder à l'image suivante. Ainsi, si le sujet désire effectuer une pause, il le fait à ce moment précis, et clique lorsqu'il veut reprendre.

Dès que le sujet effectue le clic de souris pour continuer, une nouvelle croix de préparation pour une autre image apparaît, sur laquelle le sujet doit fixer son regard afin de passer à l'image de la scène routière.

40 images dans un ordre aléatoire sont présentées au sujet de la manière décrite ci-dessus. A la fin de la passation des 40 images, un écran, (cf. Fig. 7.8) annonce au sujet, en le remerciant, qu'il vient de finir la première phase de l'expérimentation.

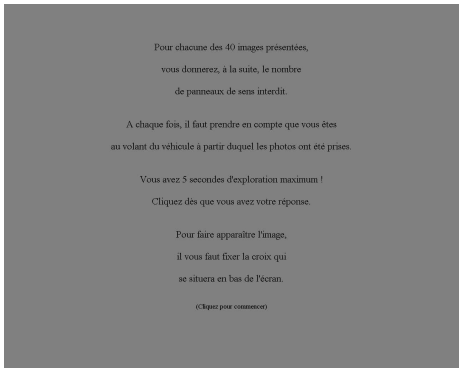
L'expérimentateur intervient alors pour rappeler au sujet qu'il ne doit pas bouger entre les deux phases et lance le scénario de la deuxième phase.

Deuxième phase : recueil de données comportementales sur la saillance de recherche

Cette phase reprend le protocole de l'expérimentation exploratoire à quelques nuances près. Les images présentées sont les mêmes images que dans la première phase.

Le sujet est assis devant l'écran avec sa main posée sur la souris, en position d'attente. Une fois le calibrage de l'oculomètre vérifié, et/ou ré-effectué si le sujet à bougé, l'expérimentateur lance le scénario *Presentation*^{TM}. En parallèle, l'oculomètre reçoit automatiquement l'ordre de se préparer à l'enregistrement des positions du regard du sujet et de passer en mode pause. Comme pour la première phase, il n'enregistre les données oculaires que le temps de l'affichage des images.

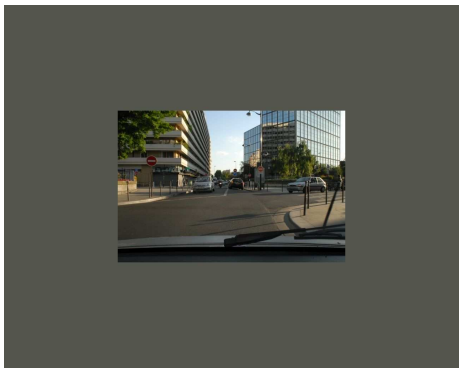
L'expérimentateur explique à nouveau au sujet que l'expérimentation n'est en aucun cas un test d'aptitude et que ses résultats ne seront pas divulgués. Il est également rappelé au sujet qu'il peut, s'il le désire, à tout moment effectuer une pause entre deux images. Dans tous les cas, même s'il effectue une pause, il lui est rappelé de ne pas effectuer de mouvements pendant cette pause, ainsi que de



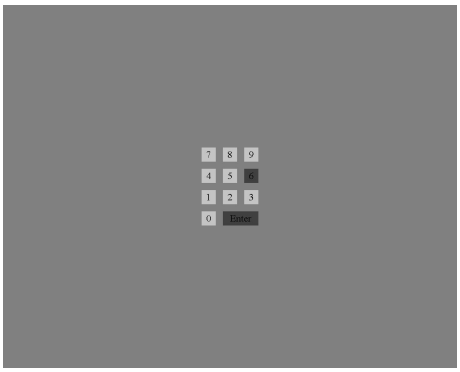
Consigne



Croix de fixation



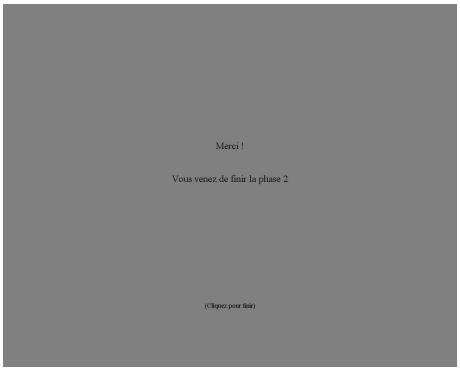
Scène routière



Écran de réponse



Écran de transition



Écran de fin

FIGURE 7.9 – Déroulement de la deuxième phase de l'expérimentation.

ne pas effectuer de grand mouvements de la tête.

La consigne, (cf. 7.9). décrivant la phase de l'expérimentation s'affiche, tandis que l'expérimentateur explique la procédure. Lorsqu'il est assuré de la bonne compréhension de celle-ci par le sujet, il l'invite à cliquer pour débuter par un exemple :

Pour chacune des 40 images présentées,
vous répondrez, à la suite, le nombre
de panneaux de sens interdit.

A chaque fois, il faut prendre en compte que vous êtes
au volant du véhicule à partir duquel les photos ont été prises.

Vous avez 5 secondes d'exploration maximum !
Cliquez dès que vous avez votre réponse.

Pour faire apparaître l'image,
il vous faut fixer la croix qui
se situera en bas de l'écran.

(Cliquez pour commencer l'expérimentation)

Le sujet clique pour commencer l'expérimentation sur une image choisie afin que l'expérimentateur puissent s'assurer de la bonne compréhension de la consigne par le sujet. La consigne est ré-affichée et le sujet clique pour commencer l'expérimentation sur une série de 40 images. Le déroulement est le même que pour la première phase. Par contre la consigne est différente. L'écran avec une croix blanche centrée en bas est toujours affichée entre chaque image de scène routière (cf. Fig.7.9).

Lorsque l'image de la scène routière apparaît, le sujet doit compter le plus rapidement possible les panneaux de sens interdit et donner sa réponse en cliquant

sur la souris ou après le temps d'apparition maximal de cinq secondes. Un pavé numérique, (cf. Fig. 7.9) apparaît, le sujet doit alors sélectionner le chiffre (de 0 à 9) qui correspond à son décompte de panneaux de sens-interdit.

Une fois la réponse validée, un écran apparaît, (cf.7.9), et indique au sujet qu'il doit cliquer pour acceder à l'image suivante. Comme pour la première phase, si le sujet désire effectuer une pause, il le fait à ce moment précis, et clique lorsqu'il veut reprendre.

Dès que le sujet effectue le clic de souris pour continuer, une nouvelle croix de préparation pour une autre image apparaît, sur laquelle le sujet doit fixer son regard afin de passer à l'image de la scène routière.

Les 40 images sont présentées au sujet de la manière décrite ci-dessus, dans un ordre aléatoire. Après l'affichage et la réponse du sujet sur la dernière image, un écran, cf.Fig7.9 ; annonce au sujet, en le remerciant, qu'il vient de finir la deuxième phase de l'expérimentation.

L'expérimentateur intervient alors pour avertir le sujet qu'à partir de ce moment il est libre de tout mouvement et lance le scénario de la troisième phase.

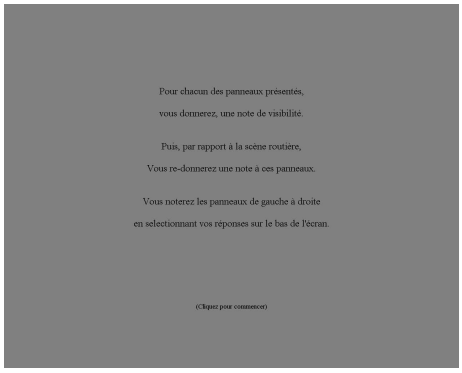
Troisième phase : recueil de données subjectives sur la saillance

L'objectif des deux premières phases est de nous procurer une mesure objective, de la part du sujet, de la saillance des panneaux. Dans cette troisième phase, nous avons recueilli une mesure subjective de la saillance des panneaux en jouant sur la consigne. Ainsi, nous demandons au sujet d'évaluer la saillance de chaque panneau de la phase précédente.

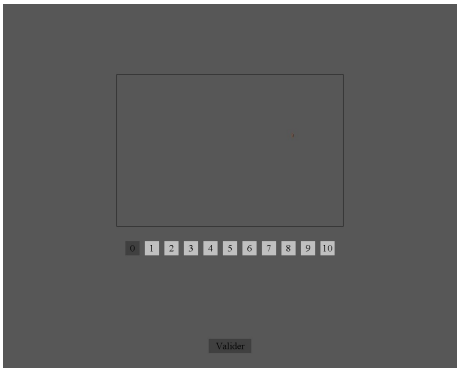
L'expérimentateur explique au sujet qu'il va devoir noter les panneaux qui lui sont proposés sur échelle allant de 0 à 10. Le critère de notation étant la saillance, afin de ne pas perturber le sujet avec un vocabulaire compliqué, il est fait mention dans la consigne de « facilité à voir le panneau dans la scène » et de « facilité à être perçu ».

Les images présentées dans cette dernière phase sont les mêmes images que celles présentées dans les deux phases précédentes.

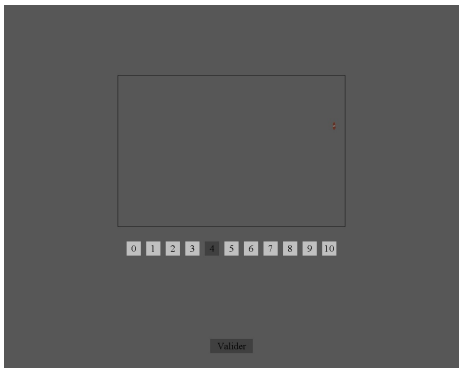
Le sujet est assis devant l'écran avec sa main posée sur la souris, en position d'attente. Il est notifié au sujet que lors de cette dernière phase, le temps ne compte pas et qu'il peut faire une pause à tout moment.



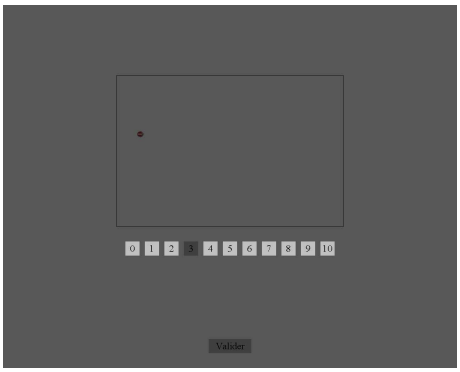
Consigne



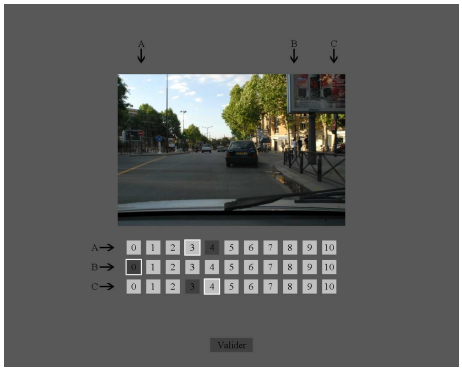
1er panneau seul



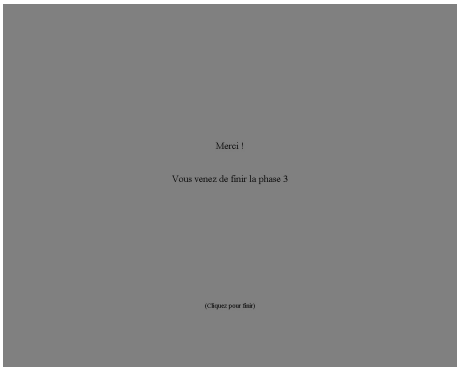
2ème panneau seul



3ème panneau seul



Evaluation des panneaux dans la scène



Écran de fin

FIGURE 7.10 – Déroulement de la troisième phase de l'expérimentation.

La consigne (cf. Fig. 7.10) décrivant la troisième phase de l'expérimentation s'affiche, tandis que l'expérimentateur explique la procédure. Lorsqu'il est assuré de la bonne compréhension de celle-ci par le sujet, il l'invite à cliquer pour débuter par un exemple :

Pour chacun des panneaux présentés,
vous donnerez, une note de visibilité

Puis, par rapport à la scène routière,
vous re-donnerez une note à ces panneaux.

Vous noterez les panneaux de gauche à droite
en sélectionnant vos réponses sur le bas de l'écran.

A chaque fois, il faut prendre en compte que vous êtes
au volant du véhicule à partir duquel les photos ont été prises.

(Cliquez pour commencer l'expérimentation)

Le sujet clique pour commencer l'expérimentation sur une image d'exemple afin que l'expérimentateur puissent s'assurer de la bonne compréhension de la consigne par le sujet. La consigne est ré-affichée et le sujet clique pour commencer l'expérimentation sur la série de 40 images.

Comme le montre la Fig. 7.10, pour chacune des images, chaque panneau de l'image, est tout d'abord affiché seul, à son emplacement dans la scène routière mais l'environnement du panneau n'est pas affiché et est remplacé par un fond gris. Il est demandé au sujet de donner une note entre 0 et 10 sur la « visibilité » du panneau, en cliquant sur la note idoine parmi celles proposées sous l'image et en validant celle-ci par l'icône « enter » à droite. Ce processus est répété pour tous les panneaux de l'image, dans un ordre aléatoire. Dans un deuxième temps, l'image de la scène routière comprenant tous les panneaux notés précédemment,

est affichée. Cette fois ci, l'environnement routier est présent. Il est demandé au sujet de donner une note à chaque panneau, en effectuant l'analyse de gauche à droite. Il y a autant de lignes de notes disponibles en bas de l'écran qu'il y a de panneaux à noter dans l'image. La note donné individuellement au panneau lors de l'étape précédente est rappelée au sujet par un cadre blanc autour de celle-ci. Les lignes sont identifiées par des lettres qui font référence aux panneaux correspondants, comme le montre la Fig. 7.10.

Une fois tous les panneaux de l'image notés et les réponses validées, le sujet passe automatiquement à l'image suivante, tirée aléatoirement parmi les 40 images de l'expérimentation.

Après l'évaluation de la dernière image, un écran annonce au sujet, en le remerciant, qu'il vient de finir la troisième et dernière phase de l'expérimentation.

Ainsi, s'achève l'expérimentation pour le sujet. Ce dernier est remercié, reçoit un cadeaux (clé usb ou porte-clé du LCPC) et l'expérimentateur lui fait comprendre que pour l'intérêt de la manipulation, afin d'éviter d'influencer d'autres sujets, il ne doit pas parler du contenu, ni de l'objectif de cette expérimentation. L'expérimentateur lui rappelle qu'il dispose d'un droit d'accès à ses données personnelles et qu'en aucun cas ses données ne seront divulguées nommément.

Il est à noté que les trois phases de l'expérimentation ne sont pas contrebalancées car, celles-ci se focalisant de plus en plus sur l'objet de l'étude, l'effet de rescence serait alors trop important et fausserait l'estimation de la saillance attentionnelle en première phase, si la deuxième phase était effectuée avant.

7.1.4 Variables étudiées

Variables indépendantes : Les variables indépendantes sont la saillance de chaque panneau, leur taille et leur emplacement.

Variables dépendantes : Les mesures effectuées sont :

- Les réponses objectives des sujets aux questions sur l'image pour les deux premières phases.
- Les réponses verbales des sujets apportant des précision sur leurs réponses lors de la troisième phase.
- Les notes subjectives de saillance pour la troisième phase.

- Les chemins visuels des sujets et l'analyse des fixations en fonction des paramètres de taille et de durée, plus précisément :
  - Emplacement ;
  - Durée ;
  - Nombre total de Fixations ;
  - Ordre des fixations ;
  - Temps passé de l'affichage de l'image jusqu'à la fixation.

7.2 Analyse des données et modèle statistique

L'exploration des données à l'issu de l'expérimentation, a montré plusieurs pertes de suivi du regard ou des pertes du calibrage pour certains sujets. Les raisons étaient multiples (lunettes à verres progressifs, formes de l'œil incompatible avec la technique de reflet cornéen, sujets « remuants » ou non coopératifs). Ainsi, les données de 32 sujets ont été gardés pour l'analyse. Le protocole de l'expérimentation permet de faire ressortir deux types d'indicateurs de la saillance :

- Des indicateurs objectifs, par l'analyse du mouvement des yeux et des réponses lors des deux premières phases.
- Des indicateurs subjectifs, par les notations effectuées par les sujets grâces aux échelles lors de la troisième phase.

7.2.1 Evaluation objective

Dans la première et la deuxième phase, pour chaque image et pour chaque sujet j, les données oculométriques nous permettent de mesurer l'emplacement et la durée des fixations visuelles [131]. Une fixation est considérée quand le nombre de données oculaires (à la fréquence de $50Hz$, i.e. toutes les $20ms$) incluse dans une fenêtre carré est au dessus du seuil choisi. Ce seuil et la taille de la fenêtre dépendent de la distance par rapport à l'écran ($70cm$), de la précision de l'oculomètre (0.5°) et de la résolution de l'image (640×480). En pratique, nous prenons donc un seuil de $100ms$ et une taille de fenêtre de 32×32, ce qui correspond à 1° d'angle visuel.

Comme nous l'avons discuté précédemment, on ne peut affirmer qu'un objet n'a pas été vu par la vision péri-fovéale, avec les données oculaires seules. En revanche, si on couple ces données avec les réponses données par les sujets dans chaque phase, l'analyse nous permet de savoir si le sujet a prêter attention à un panneau donné. Dans la première phase, les réponses sur les directions possible à prendre nous permettent d'inférer si le sujet à pris en compte les panneaux de sens interdit. Dans la deuxième phase, c'est la réponse des sujets sur le nombre de sens interdit contenus dans chaque image qui nous permet d'obtenir cet indicateur.

Ainsi, on peut décrire pour chaque phase une valeur de détection du panneau i par le sujet j. Cette valeur binaire $D(i,j)$ indique que le panneau i a été pris en compte par le sujet j, durant la recherche d'indice visuel pour la première phase et durant la recherche de sens interdit dans la deuxième phase.

Pour un panneau i donné, la moyenne des détections sur l'ensemble des sujets nous donne un Taux de Détection en Recherche (TDR) pour ce panneau :

$$TDR(i) \quad = \quad \frac{\sum_{j=1}^{j=32} D(i,j)}{32} \tag{7.1}$$

L'autre indicateur objectif de saillance que nous prenons en compte est la durée moyenne des fixations effectuées sur chaque panneau. A la nuance près qu'il n'y a de durée que si il y a fixation, or un panneau très saillant n'est pas forcément fixé du fait de la détection et de la reconnaissance en vision proche périphérique, comme le montre le trajet oculaire de la Fig. 7.11 sur une des images de l'expérimentation.

7.2.2 Evaluation subjective

Lors de la troisième phase, le sujet évalue la saillance des panneaux de sens interdit, ou plutôt, leur visibilité ou facilité à être perçus en conduite. Pour chaque panneau, les sujets donnent deux notes, la première au panneau seul, sur un fond gris et la deuxième, au même panneau mais au sein de la scène routière. Notre objectif est de modéliser la saillance de la signalisation verticale en fonction de la scène routière, nous n'avons donc exploré, dans le cadre de ce manuscrit de thèse, que la notation des panneaux dans leur contextes routiers.

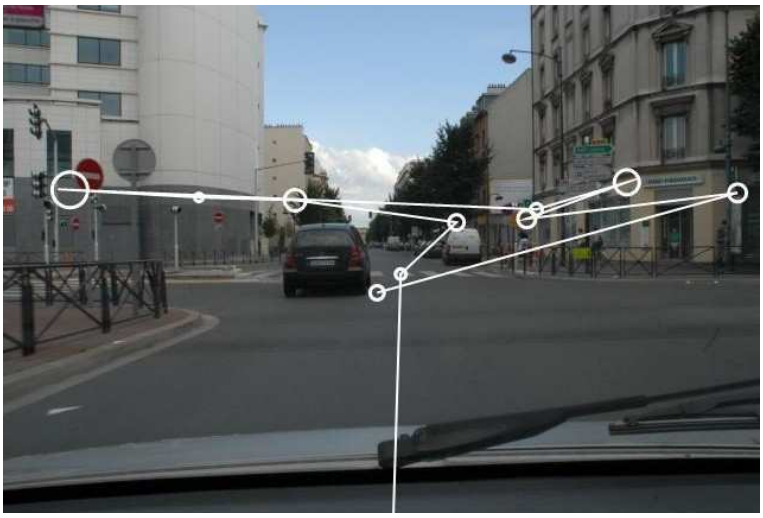


FIGURE 7.11 – Trajet Oculaire d'un sujet qui cherche les panneaux de « sens interdit ». Chaque cercle représente une fixation. Plus le cercle est large, plus la fixation est longue. Le regard commence en bas de l'écran, au milieu.

La réponse des sujets sur l'évaluation de la saillance des panneaux de sens interdit dans la scène routière nous défini un $score(i,j)$ qui évalue chaque panneau i par chaque sujet j. Du fait de la variabilité des sujets dans l'utilisation de l'échelle de notation, les scores des sujets sont standardisés en suivant une loi gaussienne. Le Score de Saillance Standardisé (SSS), relatif à l'évaluation subjective de la saillance est ainsi défini :

$$SSS(i,j) = \frac{score(i,j) - E_i(score(i,j))}{\sigma_i(score(i,j))} + 5 \tag{7.2}$$

ou $E_i(score(i,j))$ est la moyenne des scores sur les panneaux pour chaque sujet j et $\sigma_i(score(i,j))$ est la déviation standardisée du score sur les panneaux pour chaque sujet j.

7.3 Résultats et Interprétation : Une corrélation positive entre le modèle et le comportement humain

Lorsque l'on compare le taux de détection des panneaux entre les deux premières phases, comme le montre la Fig. 7.12, on en retire l'analyse que les panneaux ont des taux de détection plus élevés lorsqu'il sont activement recherchés lors de la deuxième phase. Or, lors de la première phase, le sujet n'avait pas pour consigne de se focaliser sur les panneaux, cette phase est donc plus écologique par rapport à une situation de conduite que la deuxième phase. On peut donc inférer que la baisse de performance de détection des panneaux sera d'autant plus grande que des éléments de la tache de conduite seront introduits. Ce résultat va dans le sens qu'un panneau se doit d'être très saillant pour être pris en compte par le conducteur et donc d'avoir un taux de détection élevé, lors de la tâche de recherche. Par la suite, nous nous sommes focalisés sur l'exploitation des résultats pour la tâche de recherche, i.e. pour la deuxième phase.

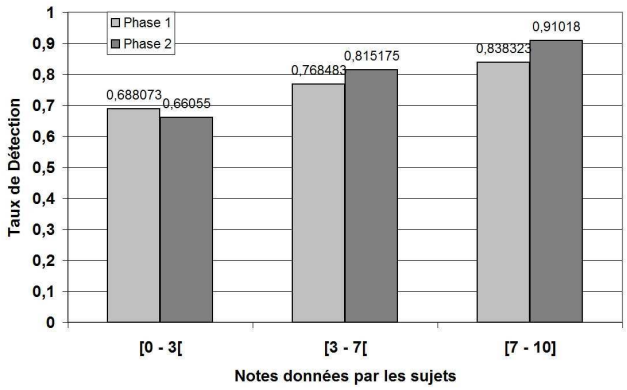


FIGURE 7.12 – Taux de détections des panneaux de sens interdit en fonction de la note de visibilité donnée par les sujets.

Nous avons procédés à une analyse statistique [2] avec le logiciel Stata.

2. Cette analyse à été réalisé avec Henri Panjo, lors de son stage de 6 mois qui s'inscrivait dans

Dans une première étape, nous avons étudié le lien entre les performances de détection des panneaux par les sujets (*TDR*) et leur évaluation subjective (*SSS*). Comme le montre la Fig. 7.13, plus le score du panneau (*SSS*) augmente, plus le taux de détection (*TDR*) augmente. Si on seuille le score à la valeur de 4, on peut couper le graphique en deux classes principales, les mauvais scores et les bons scores. Les mauvais scores correspondent aux panneaux de sens interdit qui ne sont pas correctement pris en compte par les sujets, alors que la tâche est spécifiquement de les chercher. Assurément, dans un vrai contexte de conduite, ce seuil serait plus haut, car conduire implique le déroulement simultané de plusieurs tâches [132].

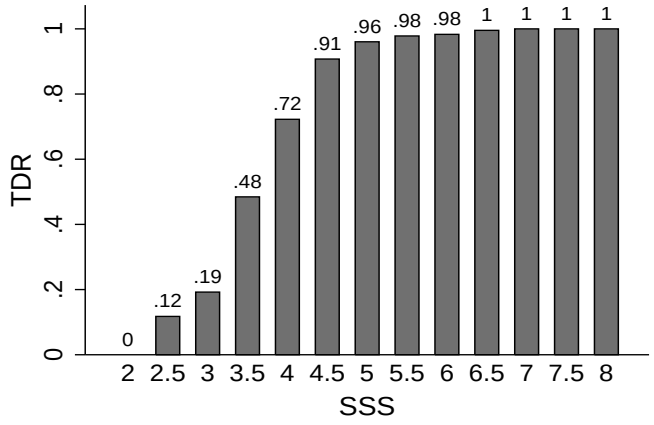


FIGURE 7.13 – Taux de Détection en Recherche (*TDR*) en fonction du Score de Saillance Standardisé (*SSS*).

Lorsque l'on compare pour chacun des panneaux leurs valeurs de saillance extraites de la Carte de Saillance de Recherche (*CSR*), définie dans l'équation 6.3 à leurs valeurs subjectives de saillance évaluées par les sujets (*SSS*), on observe une relation linéaire entre elles, comme le montre la Fig. 7.14.

D'autre part, connaissant la relation entre le taux de détection des panneaux (*TDR*) et la *SSS* de la Fig. 7.13, cela illustre que la *CSR* est aussi corrélée à la

le cadre de sa deuxième année de Master Ingénierie Mathématique, Science du Vivant ; parcours Statistiques Appliquées.

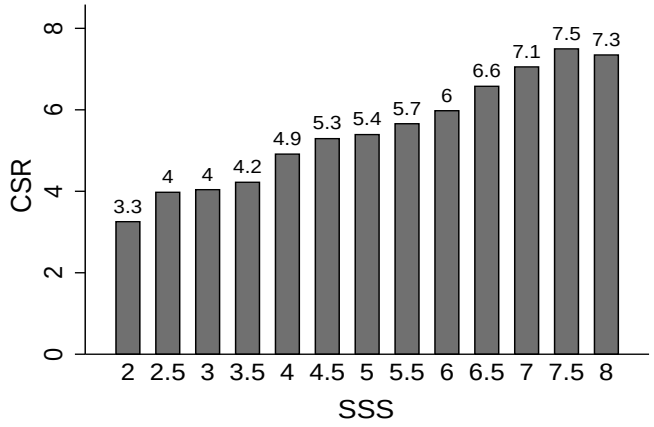


FIGURE 7.14 – Valeur de la Carte de Saillance de Recherche (*CSR*) en fonction du Score de Saillance Standardisé (*SSS*).

facilité de détection des panneaux (*TDR*), qui définie la visibilité ou plutôt la saillance du panneau vis à vis de son environnement routier (cf. partie 3.2).

Par ailleurs, l'étude statistique prouve que le model de saillance de recherche proposé dans l'équation 6.3, i.e. la *CSR*, explique 56% de la variance entre les panneaux contenus dans les images de l'expérimentation. Le même test en prenant comme modèle la seule taille des panneaux ($CSR(i) = \sqrt[4]{A(i)}$) montre que celui-ci explique 46% de la variance entre les panneaux. Donc, l'estimateur de saillance de recherche proposé améliore le model basé sur la taille et augmente l'explication de la variance de 18%

Lorsque l'on restreint l'étude aux panneaux pour lesquels nous avons des fixations de la part des sujets, on montre Fig. 7.15 que plus le modèle de saillance de recherche proposé (*CSR*) estime la saillance du panneau, moins le temps de fixation moyen sur celui-ci est important. On rappelle que le temps de fixation est directement relié à la tâche d'identification, qui correspond à la reconnaissance de l'objet détecté, donc à la saillance intrinsèque de l'objet, l'observateur étant fixé sur cet objet (cf partie. 3.2).

Comme le montre la Fig. 7.16, la même relation est observée entre l'évaluation subjective de la saillance des panneaux par les sujets (*SSS*) et la moyenne des fixations effectuées par les sujets sur les mêmes panneaux. On retrouve clairement

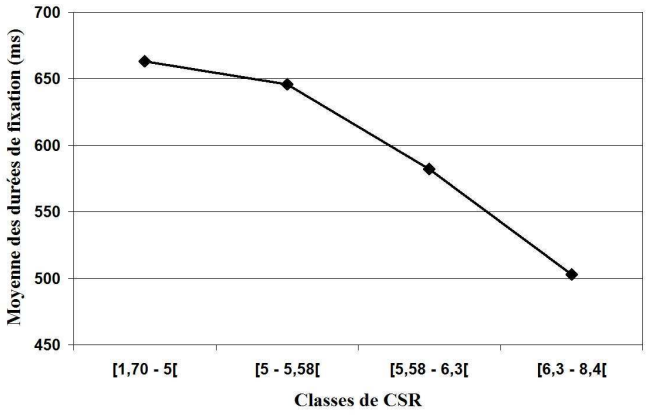


FIGURE 7.15 – Moyenne des durées de fixation des sujets sur les panneaux de sens interdit en fonction de la valeur estimée par le modèle de saillance de recherche (*CSR*).

le seuil de note de 4.5, délimitant la frontière entre la classe de panneau de faible saillance et celle des panneaux de forte saillance. Ce seuil, ainsi que la forme de la relation semblable aux résultats trouvés précédemment entre le taux de détection TDR et le *SSS* (cf.Fig. 7.13), prouve qu'il y a une corrélation forte entre la facilité de détection des panneaux TDR (i.e. saillance par rapport au fond), la durée moyenne des fixations (i.e. saillance intrinsèque) et notre modèle de saillance de la signalisation.

Nous prouvons donc que notre modèle d'estimation de la saillance de recherche des panneaux est corrélé à la capacité de détection des sujets et aux durées de fixation dans une tâche de recherche de l'information de la signalisation, qui est une sous tâche de la conduite. Ce constat induit une corrélation forte du modèle au comportement visuel humain en situation de conduite, car avec l'introduction des autres sous-tâches de conduite qui se dérouleront en parallèle, il y aura une diminution globale des performances des sujets. Ceci ne changera pas la corrélation mais aura pour effet de réhausser le seuil de séparation entre panneaux de mauvaise saillance et de bonne saillance.

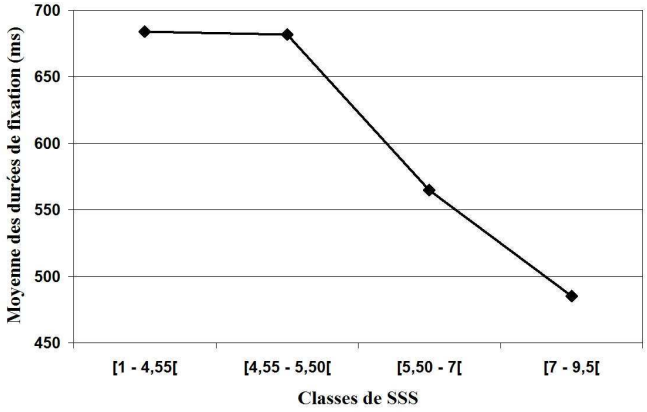


FIGURE 7.16 – Moyenne des durées de fixation des sujets sur les panneaux de sens interdit en fonction du Score de Saillance Standardisé (*SSS*).

Chapitre 8

Conclusion et perspectives

Ce travail de thèse portait sur l'étude de la faisabilité d'un outil d'aide au diagnostic de la saillance de la signalisation verticale à partir d'une image routière. La définition de la saillance que nous proposons est la facilité de détection et d'identification des panneaux, par l'usager lorsqu'il en effectue sa recherche, cf. partie 3.3.3 et [90, 91]. Afin d'évaluer la saillance de recherche, comme nous l'avons montré dans le chapitre 6 , nous avons fondé notre estimateur sur l'utilisation des valeurs de confiance issues d'un classifieur SVM que l'on utilise, par déplacement sur l'image de la scène routière, pour détecter les zones potentiellement saillantes au regard du panneau recherché, cf. partie 6.1 et [133]. Nous avons pris également en compte dans le modèle final l'environnement proche du panneau (cf. partie 6.2) et sa taille (cf. partie 6.3) dans la scène routière [114, 134]. In fine, cette décomposition du modèle nous permet de proposer trois mesures différentes :

- Une *Carte de Saillance Intrinsèque*, qui permet de calculer un indice de facilité d'identification du panneau par un usager lorsqu'il l'a au préalable détecté.
- Une *Carte de Saillance relative au Fond*, qui permet de calculer un indice de détachement du panneau par rapport à son fond environnant.
- Une *Carte de Saillance de Recherche*, qui permet de calculer, pour un panneau, un indice de sa propension à être détecté par un usager en situation de recherche de l'information nécessaire à la tâche de conduite.

L'utilisation du dernier des trois indicateurs permet de déterminer si un panneau donné a le niveau de saillance requis pour assurer la qualité de service de l'infrastructure. Dans le cas où ce panneau ne serait pas suffisamment saillant, l'utilisation des deux premières cartes, en comparaison avec la dernière, permet de caractériser si ce défaut est dû à la taille du panneau, à une similarité de son fond proche ou à un défaut inhérent au panneau lui même (causé par la vétusté, un défaut colorimétrique, etc.). Ces trois indicateurs peuvent donc être utiles aux gestionnaires pour évaluer la saillance de leur infrastructure, cf. partie 6.4.

Le dernier indicateur peut également permettre une application de sécurité active dans les véhicules intelligents. Comme présenté dans la partie 6.4 et [114], il peut intervenir dans une assistance à la conduite qui prévient le conducteur de la présence d'un panneaux uniquement si celui-ci risque de ne pas être perçu du fait d'une faible saillance.

Le cœur de notre algorithme, la méthode de classification par SVM, est optimisé afin de répondre aux spécificités de notre objet d'étude et à son application, comme présenté dans le chapitre 5 et [133]. Une régularisation dissymétrique, un noyau spécifique (triangulaire) et une signature basée sur une approche couleur sont utilisées.

Le modèle a été testé par rapport à la performance visuelle de sujets dans des tâches impliquées en conduite. Nous avons conduit une expérimentation d'oculométrie cognitive, présentée dans le chapitre 7, afin de recueillir des données de comportement visuel humain en situation écologique par rapport à une prise d'information visuelle en conduite. Les résultats de l'étude statistique, présentés dans la partie 7.3 et [134, 135], ont montré qu'il y a une corrélation positive entre l'estimation issue de notre *Carte de Saillance de Recherche* et la saillance évaluée objectivement et subjectivement par les sujets.

Toutefois, le modèle proposé ne prend pas en compte tous les éléments constituants la saillance. En effet, il est nécessaire de considérer également la position du panneau par rapport au champ visuel du conducteur qui est déterminé par la géométrie de la route. Ce paramètre spatial pourrait être pris en compte dans le modèle grâce à une pondération de la *Carte de Saillance de Recherche* à travers une

carte de probabilité de présence de la signalisation verticale. Cette dernière pourra être obtenue par la représentation des normes d'implantation de la signalisation et par l'accumulation des points de fixation visuelle effectués par des conducteurs, récoltés au moyen d'un oculomètre embarqué dans un véhicule. L'approche de vision focalisée, présentée dans [82, 83, 84] pourrait également contribuer à prendre ce facteur en compte. Les panneaux ayant une taille normée selon la catégorie (vitesse autorisée) de la route sur laquelle ils sont implantés, ce paramètre devra également être introduit dans le futur afin d'avoir une notion de distance de vue du panneau dans la saillance.

Notre modèle a été testé et validé pour un type de panneaux de police particulier, les panneaux de sens interdit, du fait de la prépondérance de ceux-ci dans notre base d'image et de la nécessité d'adapter la consigne de l'expérimentation pour induire un comportement de recherche spécifique. Il est donc primordial, dans un premier temps, d'appliquer l'algorithme d'apprentissage à d'autres types de panneaux de police. Ceci peut être réalisé par différentes approches, comme l'utilisation d'une méta-classe de panneaux, comprenant tous les types recherchés, ou par la fusion de classifieurs, chacun spécialisé dans un type de panneaux donné. Il est également intéressant d'appliquer ce modèle à la signalisation directionnelle, pour laquelle la saillance de recherche est majoritairement impliquée. Il faudra dans ce cas adapter la signature utilisée dans le SVM, pour représenter ce type de panneaux, car dans ce cas, même si la détection se fait par la couleur, le traitement de l'information fait intervenir de la reconnaissance d'écriture. De même, de nuit, en adaptant la signature des panneaux à l'utilisation de leur capacité de rétro-réflexion (donc d'intensité lumineuse), il devient alors possible de caractériser leur saillance dans ces conditions de visibilité nocturne. Enfin, de part l'utilisation d'un algorithme de classification, il apparaît aussi possible de modifier les caractéristiques de ce dernier afin d'évaluer la saillance de tous autres objets appris, qu'ils soient routiers ou non.

Par ailleurs, dans ce premier estimateur de saillance, seules des informations de couleur ont été prises en compte pour effectuer la recherche des panneaux, ce que l'on peut comparer au comportement de la vision pré-attentive, effectué en vision proche périphérie (vison para-fovéale et péri-fovéale) et périphérique. Le

traitement de l'information, pendant les fixations, se fait en vision fovéale où les contours sont pleinement définis (acuité maximum). Il serait donc intéressant d'en tenir compte lors du calcul de la saillance intrinsèque, ce qui implique d'ajouter des informations de forme à la signature utilisée dans le SVM. Par ailleurs, nous avons réalisé une étude sur l'utilisation de la distance de Hausdorff dans un SVM, présenté dans [118]. Nous avons transformé cette distance de similarité entre deux contours en un noyau prenant comme signature les coordonnées des points de contours dans l'image à classifier. Des tests ont été réalisés sur des données synthétiques et nécessitent d'être prolongés dans des conditions réelles de classification de formes (panneaux) dans des images complexes (scènes routières).

Concernant les caractéristiques prises en compte dans la signature qui représente les données pour le SVM, nous avons choisi celle qui donnait les meilleurs résultats parmi celles que nous avons testées. Cette représentation est constituée de deux plans quantifiés de l'espace des couleurs normalisées. Il serait intéressant de tester d'autres espaces de définition des couleurs, tels que le système *TSL* (Teinte Saturation Luminosité, en anglais « *HSL* ») basé sur la perception humaine de la couleur ; le système de couleur $L*a*b*$, dans lequel deux couleurs sont proches lorsqu'elles le sont également pour la perception humaine, car dans ce modèle les relations non-linéaires pour $L*$, $a*$ et $b*$ ont pour but d'imiter la réponse logarithmique de l'œil ; ou bien encore l'espace des couleurs de Krauskopf, utilisé dans le cadre du calcul de la saillance attentionnelle par [66, 67]. Une caractéristique intéressante de ces modèles luminance-chrominance est qu'ils permettent de s'affranchir des changements d'illumination de notre objet d'étude si on ne considère pas le plan de luminosité. De même, une exploration du système *TSV* (Teinte Saturation Valeur, en anglais « *HSV* ») couramment utilisé dans la reconnaissance d'image pourrait être un apport. Toutefois, il est à rappelé que l'algorithme de détection, en plus d'avoir un bon taux de détection, doit aussi être corrélé au comportement visuel humain.

Le noyau triangulaire, que nous avons choisi par nos tests, convient bien aux caractéristiques de notre base d'exemples. Mais, du fait que nos signatures soient fondées sur des histogrammes, il serait intéressant, de plus, de tester des noyaux spécifiques aux histogrammes tels que le noyau χ^2, proposé par Chapelle et

al. [125], le noyau intersection d'histogrammes *HI*, introduit par Swain & Ballard [120] et sa version généralisée, appelée noyau *GHI* proposé par Boughorbel et al. [136]. Toutefois, ces noyaux seront probablement moins à même de fonctionner à cause de la différence de densité entre nos deux classes d'exemples. Ceci peut être corrigé par une procédure d'apprentissage par les erreurs que nous avons commencé à tester. Cette procédure permet d'apprendre itérativement sur les images de la base d'apprentissage, uniquement ceux qui sont nécessaires. Ainsi, la taille de la base d'exemples (négatifs, i.e. non-panneau, dans notre cas) et le nombre de Support Vecteur sont fortement réduits, ce qui conduit à une rapidité accrue du SVM dans les phases d'apprentissages et de classification. L'influence de cette procédure sur les résultats reste à étudier expérimentalement.

En ce qui concerne la validation du modèle, de par le matériel mis à notre disposition, nous avons évalué le modèle proposé par rapport à des images de scènes routières présentées aux sujets tandis que les différentes consignes induisaient une saillance attentionnelle ou de recherche. Lors de l'étude statistique, nous avons pris comme indicateurs objectifs de saillance, des indicateurs oculométriques tels que l'emplacement et le temps de fixation du regard. Une étude d'autres indicateurs tels que les matrices de transition entre les zones d'intérêts fixées, le nombre, l'amplitude et la taille des saccades, ainsi que les ratios des temps de fixation - saccade, pourra amener des précisions ou des éléments nouveaux. Il serait intéressant, également, de réaliser d'autres expérimentations, avec l'utilisation d'un nouvel oculomètre permettant d'être dans un environnement dynamique et donc de recueillir des données du comportement visuel humain en conduite réelle ou simulée.

En tout état de cause, les perspectives de cette thèse sont nombreuses et nécessitent une coopération complète entre les domaines des sciences de l'ingénieur en informatique (traitement d'image et vision par ordinateur) et les domaines des sciences humaines (psychologie et ergonomie cognitive).

Bibliographie

[1] R.L. Gregory. *Eye and Brain : The psychology of seeing*. Oxford University Press Oxford, 1997.

[2] Ministère de l'Ecologie de l'Energie, du Développement Durable et de la Mer (MEEDDM). Instruction interministérielle sur la signalisation routière. In *IISR*. MEEDDM, November 2008.

[3] Direction de la Sécurité et de la Circulation Routières DSCR. Les signaux routiers. In *Les signaux routiers*. Direction de la Sécurité et de la Circulation Routières, Ministère de l'Equipement, des transports et du Logement, DSCR, November 2002.

[4] Michel BRY. La perception visuelle de la route et de ses équipements. *Revue Générale des Routes et Aérodromes*, 865, Mars 2008.

[5] D. Shinar, N. Meir, and I. Ben-Shohan. How automatic is manual gear shifting ? *Human Factors*, 40 :647–654, 1998.

[6] D. Shinar, T.H. Rockwell, and J. Maleck. The effects of changes in driver perception on rural curve negociation. *Ergonomics*, 23 :263–275, 1980.

[7] Hills. Vision, visibility and perception in driving. *Perception*, 9 :183–216, 1980.

[8] P.K. Hughes and B.L. Cole. Search and attention conspicuity of road traffic control devices. *Australian Road Research,* 14 :1–9, 1984.

[9] B.L. Cole and S.E. Jenkins. The nature and measurement of conspicuity. *10th Conference Australian Research Board*, 10(4) :99–107, 1980.

[10] Michel BRY and Michèle COLOMB. Visibilité de la signalisation : les besoins des usagers et les technologies disponibles. *Revue générale des routes et aérodromes*, 658, 1988.

[11] CIE137. The conspicuity of traffic signs in complex backgrounds. In *CIE137, Technical report of the Commision Internationale de L'Eclairage (CIE)*, 2000.

[12] CIE Commision Internationale de l'Eclairage. Roadsigns. In *Publication CIE N° 74*, 1988.

[13] T. Bellet, B. Bailly-Asuni, P. Mayenobe, and A. Banet. A theoretical and methodological framework for studying and modelling drivers' mental representations. *Safety Science*, 47(9) :1205–1221, 2009.

[14] TM Allen, H. Lunenfeld, and GJ Alexander. Driver information needs. In *Highway Research Board*, volume 366, pages 102–115, 1971.

[15] Direction de la Sécurité et de la Circulation Routières DSCR. Norme xp p 98-501 « signalisation routière verticale - généralités ». In *Norme XP P 98-501 « Signalisation routière verticale - Généralités »*. Direction de la Sécurité et de la Circulation Routières, Ministère de l'Equipement, des transports et du Logement, DSCR, Mars 2002.

[16] Nicolas Hautière. *Détection des conditions de visibilité et estimation de la distance de visibilité par vision embarquée*. PhD thesis, Université Jean Monnet de Saint-Étienne, September 2005.

[17] J.J. GIBSON and L.E. CROOKS. A theorical field-analysis of automobile-driving. *The American Journal of Psychology*, 3 :453–471, 1938.

[18] Michel NEBOIT. Perception, anticipation et conduite automobile. In *Le Travail Humain*, volume 37, pages 53 – 72. Laboratoire de Psychologie de l'O.N.S.E.R., 1974.

[19] Gilles Malaterre. L'acte de conduite. In *CAHIERS Risques Collectifs et Situations de Crise, Recherche et sécurité routière Pour une action publique renouvelée*, volume 6. Publications de la MSH-ALPES, November 2006.

[20] JC Lecas. *L'attention visuelle. De la conscience aux neurosciences*. Mardaga, Wavre, Belgique, 1995.

[21] P. Poucet. Vers une conception intégrative des processus de sélection pour la saccade et la perception. *L'année psychologique*, 105(4) :697–739, 2005.

[22] Y. Legrand. *Optique physiologique. Tome 2 : lumière et couleurs. 2ème édition*. Masson, Paris, 1972.

[23] Y. Legrand. *Handbook of Sensory Physiology : Visual Psychophysics.* Springer-Verlag, New York, 1972.

[24] J.H. Goldberg and X.P. Kotval. Computer interface evaluation using eye movements : Methods and constructs. *International Journal of Industrial Ergonomics*, 24(6) :631–645, 1999.

[25] Robert.W. Rodieck. *La vision*. De Boeck Diffusion, Februar 2003.

[26] P.M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology*, 47(6) :381–391, 1954.

[27] P. Viviani. *In Kowler, E., Eye movements and their role in visual and cognitive processes, Ch. 8*. Elsevier Publishing Company, 1990.

[28] AJ McKnight and AG Adams. Driver Education Task Analysis. *Human Resources Research Organization*, 1970.

[29] AJ McKnight and AG Hundt. Driver Education Task Analysis : Instructional Objectives. *Human Resources Research Organization*, 1971.

[30] J. Leplat. Relations between task and activity : Elements for elaborating a framework for error analysis. *Ergonomics*, 33(10) :1389–1402, 1990.

[31] F. Saad. Strucuration de la tâche et activités perceptices dans la conduite automobile. In *Document interne de l'ONSER*, 1975.

[32] F. Saad. Contribution des techniques d'observation et de verbalisation à l'analyse psychologique de l'activité : étude de cas dans le domaine de la conduite automobile. *Performances Humaines et Techniques*, HS :16–23, 1996.

[33] C. Mundutéguy. *Reconnaissance d'intention et prédiction d'action pour la gestion des interactions en environnement dynamique*. PhD thesis, Thèse de doctorat d'ergonomie, Conservatoire National des Arts et Métiers, Paris, 2001.

[34] T. Vaa. Cognition and emotion in driver behaviour models : some critical viewpoints. In *ICTCT Workshop*, pages 48–59, Melbourne, 2008.

[35] M. Sivak. The information that drivers use : is it indeed 90% visual ? *Perception*, 25 :1081–1090, 1996.

[36] Joseph.E. Barton, Theodore.E. Cohn, and Masayoshi Tomizuka. Towards a complete human driver model : The effect of vision on driving performance. In *Proceedings of the 2006 American Control Conference (ACC'06)*, Minneapolis, Minnessota, USA, 2006.

[37] J.-M Cellier. Exigences et gestion temporelle dans les environnements dynamiques. In *Cellier, V. De Keyser, Valot, C., La gestion du temps dans les environnements*

dynamiques. Coll, Le Travail Humain, pages 19–48. Paris : Presse universitaire de France, 1996.

[38] Christophe Munduteguy and Françoise Darses. Perception et anticipation du comportement d'autrui en situation simulée de conduite automobile. *Le Travail Humain*, 70 :1–32, 2007.

[39] O. Laya. De la situation réelle à la simulation : exemples d'analyses de l'exploration visuelle de l'automobile. In A. Pottier and M. Neboit, editors, *L'analyse ergonomique du travail par l'étude de l'exploration visuelle*, pages 69–94. Octarès, Toulouse, 1995.

[40] A. Bertrand and P. H. Garnier. *Psychologie cognitive*. Jeunes Editions, Levallois-Perret, 2005.

[41] W. P. Tanner and J. A. Swets. A decision-making theory of visual detection. *Psychological Review*, 61 :401–409, 1954.

[42] D. M. Green and J. A. Swets. *Signal detection theory and psychophysics*. John Wiley and Sons, New York, 1966.

[43] R. Ghiglione and J.-F. Richard. *Cours de psychologie. T. 1. Origines et bases*. Dunod, Paris, 1999.

[44] G. A. Rousselet and M. Fabre-Thorpe. Les mécanismes de l'attention visuelle. *Psychologie Française*, 48(1) :29–44, 2003.

[45] M. I. Posner. Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32 :3–25, 1980.

[46] W. Schneider and R.M. Schiffrin. Controlled and automatic human information processing : I. detection, search, and attention. *Psychological Review*, 84 :1–66, 1977.

[47] A.M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive psychology*, 12(1) :97–136, 1980.

[48] A. Treisman and J. Souther. Search asymmetry : a diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology*, 114 :285–310, 1985.

[49] J. Wolfe. Visual attention. In J. Wolfe, editor, *Seeing*, pages 335–386. Academic Press, London, 2000.

[50] P.K. Hughes and B.L. Cole. The effect of attentional demand on eye movement behaviour when driving. In A.G. Gale, editor, *Vision in vehicles II*. Elsevier Science Publishers B.V., Amsterdam, 1988.

[51] P.K. Hughes and B.L. Cole. What attracts attention when driving ? *Ergonomics*, 29 :377–391, 1986.

[52] R. Brémond, , J.-P. Tarel, H. Choukour, and M. Deugnier. La saillance visuelle des objets routiers, un indicateur de la visibilité routière. In *Proceedings of Journées des Sciences de l'Ingénieur (JSI'06)*, Marne la Vallée, France, 2006. http ://perso.lcpc.fr/tarel.jean-philippe/publis/jsi06b.html.

[53] J.-P. Tarel and H. Choukour. Expérimentation pour tester un modèle de calcul de la saillance attentionnelle. Technical report, Laboratoire Central des Ponts et Chaussées, DESE - EV, 2006.

[54] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20 :1254–1259, 1998.

[55] C. Koch and S. Ullman. Shifts in selective visual attention : towards the underlying neural circuitry. *Human neurobiology*, 4(4) :219, 1985.

[56] L. Itti and C. Koch. Feature combination strategies for saliency-based visual attention systems. *Journal of Electronic Imaging*, 10 :161, 2001.

[57] B. Julesz and J.R. Bergen. Textons, the fundamental elements in preattentive vision and perception of textures. *The Bell system technical journal*, 62(6) :1619–1645, 1983.

[58] S. Ahmad. VISIT : An efficient computational model of human visual attention. *University of Illinois at Urbana-Champaign, Champaign, IL*, 1992.

[59] R. Milanese. *Detecting salient regions in an image : from biological evidence to computer implementation*. PhD thesis, University of Geneva, 1993.

[60] J.K. Tsotsos, S.M. Culhane, W.Y. Kei Wai, Y. Lai, N. Davis, and F. Nuflo. Modeling visual attention via selective tuning. *Artificial intelligence*, 78(1-2) :507–545, 1995.

[61] G. Backer, B. Mertsching, and M. Bollmann. Data-and model-driven gaze control for an active-vision system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1415–1429, 2001.

[62] N. Ouerhani, R. von Wartburg, H. Hugli, and R. Muri. Empirical validation of the saliency-based model of visual attention. *Electronic Letters on Computer Vision and Image Analysis*, 3(1) :13–24, 2004.

[63] W. Einhäuser, T.N. Mundhenk, P. Baldi, C. Koch, and L. Itti. A bottom–up model of spatial attention predicts human error patterns in rapid scene recognition. *Journal of Vision*, 7(10) :6, 2007.

[64] O. Le Meur, D. Thoreau, P. Le Callet, and D. Barba. A spatio-temporal model of the selective human visual attention. In *IEEE International Conference on Image Processing*, volume 3, pages 1188–91, 2005.

[65] O. Le Meur, P. Le Callet, D. Barba, D. Thoreau, and F. France. Modélisation spatio-temporelle de l'attention visuelle. In *COmpression et REpresentation des Signaux Audiovisuels (CORESA'05)*, november 2005.

[66] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau. A coherent computational approach to model bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5) :802–817, 2006.

[67] O. Le Meur, P. Le Callet, and D. Barba. Predicting visual fixations on video based on low-level visual features. *Vision Research*, 47(19) :2483–2498, 2007.

[68] G. Underwood and T. Foulsham. Visual saliency and semantic incongruency influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology*, 59(11) :1931–1949, 2006.

[69] J.M. Henderson, PA Weeks, and A. Hollingworth. The effects of semantic consistency on eye movements during complex scene viewing. *Journal of experimental psychology human perception and performance*, 25 :210–228, 1999.

[70] Dana H. BALLARD. Modeling the role of task in the controle of gaze. *Perception supplement*, 38 :4, 2009.

[71] Dana H. BALLARD. Modeling the role of task in the controle of gaze. In $32^n d$ *European Conference on Visual Perception 2009 (ECVP'09)*, august 2009.

[72] Constantin A. Rothkopf, Dana H. Barla, and Hayoe Marie M. Task and context determine where you look. *Journal of Vision*, 7(14) :1–20, 2007.

[73] AL Yarbus. Eye movements and vision. 1967. *New York*, 1967.

[74] V. Navalpakkam and L. Itti. Modeling the influence of task on attention. *Vision Research*, 45(2) :205–231, 2005.

[75] A. Torralba, A. Oliva, M.S. Castelhano, and J.M. Henderson. Contextual guidance of eye movements and attention in real-world scenes : The role of global features in object search. *Psychological Review*, 113(4) :766–786, 2006.

[76] V. Sundstedt, K. Debattista, P. Longhurst, A. Chalmers, and T. Troscianko. Visual attention for efficient high-fidelity graphics. In *Spring Conference on Computer Graphics (SCCG 2005)*, pages 162–168. Citeseer, 2005.

[77] D. Walther and C. Koch. Modeling attention to salient proto-objects. *Neural Networks*, 19(9) :1395–1407, 2006.

[78] D. Gao and N. Vasconcelos. Discriminant saliency for visual recognition from cluttered scenes. In *Proc. NIPS*, pages 481–488. Citeseer, 2004.

[79] D. Gao and N. Vasconcelos. Bottom-up saliency is a discriminant process. In *IEEE International Conference on Computer Vision*. Citeseer, 2007.

[80] G. Witus and R.D. Ellis. Computational modeling of foveal target detection. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, 45(1) :47, 2003.

[81] J.M. Wolfe. Guided Search 4.0 : Current Progress with a model of visual search. *Integrated models of cognitive systems*, pages 99–119, 2007.

[82] R. Aufrere, R. Chapuis, and F. Chausse. A model-driven approach for real-time road recognition. *Machine Vision and Applications*, 13(2) :95–107, 2001.

[83] N. Trujillo, R. Chapuis, F. Chausse, and M. Naranjo. Object Recognition : A Focused Vision Based Approach. *Lecture Notes in Computer Science*, 4842 :631, 2007.

[84] R. Chapuis, F. Chausse, and N. Trujillo. Progressive Focusing : A Top Down Attentional Vision System. In *Proceedings of the 4th International Symposium on Advances in Visual Computing*, page 477. Springer, 2008.

[85] L. Karsenty and B. Pavard. Différents niveaux d'analyse du contexte dans l'étude ergonomique du travail collectif. *Réseaux*, 85 :73–100, 1997.

[86] H. Shinoda, M. M. Hayhoe, and A. Shrivastava. What controls attention in natural environments ? *Vision Research*, 41 :3535–3545, 2001.

[87] M. H. Martens and M.R.J. Fox. Do familiarity and expectations change perception ? drivers' glances and response to changes. *Transportation Research Part F : Traffic Psychology and Behaviour*, 10 :476–492, 2007.

[88] Masamitsu Möri and Mohamed Hani Abdel-Halim. Road sign recognition and non-recognition. *Accident Analysis & Prevention*, 13(2) :101–115, 1981.

[89] Luoma J. Drivers' eye fixations and perceptions. In A.G. Gale, editor, *Vision in vehicles II*, pages 231– 237. Elsevier Science Publishers B.V., New York, 1988.

[90] Ludovic SIMON, Jean-Philippe TAREL, and Roland BREMOND. A new paradigm for the computation of conspicuity of traffic signs in road images. In *Commission Internationale de l'Eclairage (CIE'07)*, volume II, pages 161 – 164. Laboratoire Central des Ponts et Chaussées, DESE - EV, July 2007.

[91] Ludovic SIMON, Jean-Philippe TAREL, and Roland BREMOND. A model for automatic diagnostic of road signs saliency. In *Transport Research Arena - Europe 2010 (TRA'10)*. Laboratoire Exploitation Perception Simulation Simulateur, LEPSiS UMR LCPC-INRETS, june 2010.

[92] N. P. Bichot. Attention, eye movements, and neurons : linking physiology and bahavior. In M. Jenkin and L.R. Harris, editors, *Vision and attention*, pages 209–234. Springer-Verlag, New York, 2001.

[93] N. Mandran, B. Meillon, and A. Pellegrin. Novembre 2007, IHM, Paris. *Interface Hommes Machines*, 2007.

[94] SMI. iview x^{TM} red, sensomotoric instruments, 2006.

[95] Pertech. Pertech : eye tracking, conception de eye tracker et tests utilisateurs, 2009.

[96] P.M. Fitts and R.E. Jones. Psychological Aspects of Instrument Display. 1. Analysis of 270'Pilot-Error'Experiences in Reading and Interpreting Aircraft Instruments, 1947.

[97] A.L. Yarbus. Eye movements during perception of complex objects. *Eye movements and vision*, 7 :171–196, 1967.

[98] L.R. Young and D. Sheena. Eye-movement measurement techniques. *American Psychologist*, 30(3) :315–330, 1975.

[99] T. Baccino and T. Colombi. L'analyse des mouvements des yeux sur le Web. *Revue d'Intelligence Artificielle*, 14 :127–148, 2001.

[100] T. Baccino. *La lecture électronique*. Presses universitaires de Grenoble, 2004.

[101] A. MASSONNEAU and N. BIARD. Etat de l'art des différents systèmes de pointages à l'œil. *Dossier de la Plate-Forme Nouvelles Technologies, Assistance Publique Hôpitaux de Paris*, 2008.

[102] A.T. Duchowski. *Eye tracking methodology : Theory and practice*. Springer-Verlag New York Inc, 2007.

[103] A. Poole and L.J. Ball. Eye tracking in human-computer interaction and usability research : current status and future prospects. *Encyclopedia of human computer interaction*, pages 211–219, 2005.

[104] D.D. Salvucci and J.H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 71–78. ACM New York, NY, USA, 2000.

[105] T. Baccino, C. Bellino, and T. Colombi. *Mesure de l'utilisabilité des interfaces*. Hermès Science : Lavoisiser, 2005.

[106] Alefs B., Eschemann G., Ramoser H., and Beleznai C. Road Sign Detection from Edge Orientation Histograms. In *Intelligent Vehicles' 07 Symposium, Proceedings of the*, pages 993–998, 2007.

[107] C. Bahlmann, Y. Zhu, R. Visvanathan, M. Pellkofer, and T. Koehler. A system for traffic sign detection, tracking, and recognition using color, shape, and motion information. In *Intelligent Vehicles' 05 Symposium, Proceedings of the*, pages 255–260, 2005.

[108] S. Estable, J. Schick, F. Stein, R. Janssen, R. Ott, W. Ritter, and Y.-J Zheng. A real-time traffic sign recognition system. In *Intelligent Vehicles' 94 Symposium, Proceedings of the*, pages 213–218, 1994.

[109] R. Janssen, W. Ritter, F. Stein, S. Ott, and A.G. Daimler-Benz. Hybrid approach for traffic sign recognition. In *Intelligent Vehicles' 93 Symposium., Proceedings of the*, pages 390–395, 1993.

[110] P. Parodi and G. Piccioli. A feature-based recognition scheme for traffic scenes. In *Intelligent Vehicles' 95 Symposium., Proceedings of the*, pages 229–234, 1995.

[111] G. Piccioli, E. De Micheli, P. Parodi, and M. Campani. Robust road sign detection and recognition from image sequences. In *Intelligent Vehicles' 94 Symposium, Proceedings of the*, pages 278–283, 1994.

[112] L. Priese, J. Klieber, R. Lakmann, V. Rehrmann, and R. Schian. New results on traffic sign recognition. In *Intelligent Vehicles' 94 Symposium, Proceedings of the*, pages 249–254, 1994.

[113] L. Priese, R. Lakmann, and V. Rehrmann. Ideogram identification in a realtime traffic sign recognitionsystem. In *Intelligent Vehicles' 95 Symposium., Proceedings of the*, pages 310–314, september 1995.

[114] Ludovic SIMON, Jean-Philippe TAREL, and Roland BREMOND. Alerting the drivers about road signs with poor visual saliency. In *Proceedings of IEEE Intelligent Vehicles Symposium 2009 (IV'09)*, pages 48 – 53. Laboratoire Exploitation Perception Simulation Simulateur, LEPSiS UMR LCPC-INRETS, June 2009.

[115] A. de la Escalera, J.M. Armingol, and M. Mata. Traffic sign recognition and analysis for intelligent vehicles. *Image and Vision Computing, 2003*, 21(3) :247–258, 2003.

[116] C.Y. Fang, C.S. Fuh, P.S. Yen, S. Cherng, and S.W. Chen. An automatic road sign recognition system based on a computational model of human recognition processing. *Computer Vision and Image Understanding, 2004*, 96(2) :237–268, 2004.

[117] Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, March 2000.

[118] Ludovic SIMON and Jean-Philippe TAREL. Utilisation de la Distance de Hausdorff dans un SVM ? In *Quatrième édition du congrés francophone de doctorants en STIC (MajecSTIC'06)*. Laboratoire Central des Ponts et Chaussées, DESE - EV, November 2006.

[119] JM Hoc. Towards ecological validity of research in cognitive ergonomics. *Theoretical issues in ergonomics science*, 2(3) :278–288, 2001.

[120] Michael J. Swain and Dana H. Ballard. Color indexing. *Int. J. Comput. Vision*, 7(1) :11–32, 1991.

[121] Marin Ferecatu. *Image retrieval with active relevance feedback using both visual and keyword-based descriptors*. PhD thesis, UNIVERSITY OF VERSAILLES SAINT-QUENTIN-EN-YVELINES, July 2006.

[122] Seong-O Shim and Tae-Sun Choi. Edge color histogram for image retrieval. *Image Processing. 2002. Proceedings. 2002 International Conference on*, 3 :957–960 vol.3, 2002.

[123] J. Song, M. Cai, and M.R. Lyu. Edge color distribution transform : an efficient tool for object detection in images. In *ICPR '02 Proceedings of the 16th International Conference on Pattern Recognition*, pages 608–611 vol.1, 2002.

[124] J. Luo and D. Crandall. Color object detection using spatial-color joint probability functions. *Image Processing, IEEE Transactions on*, 15(6) :1443–1453, 2006.

[125] O. Chapelle, P. Haffner, and V.N. Vapnik. Support vector machines for histogram-based image classification. *Neural Networks, IEEE Transactions on*, 10(5) :1055–1064, 1999.

[126] F. Fleuret and H. Sahbi. Scale invariance of support vector machines based on the triangular kernel. In *Proceedings of the workshop on Statistical and Computational Theories of Vision of the IEEE International Conference on Computer Vision (ICCV/SCTV)*, 2003.

[127] S. Boughorbel, J.-P. Tarel, and N. Boujemaa. Conditionally positive definite kernels for svm based image recognition. In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME'05)*, Amsterdam, The Netherlands, 2005. http ://www-rocq.inria.fr/ tarel/icme05.html.

[128] S. Boughorbel, J.-P. Tarel, and F. Fleuret. Non-mercer kernels for svm object recognition. In *Proceedings of British Machine Vision Conference (BMVC'04)*, pages 137 – 146, London, England, 2004. http ://www-rocq.inria.fr/ tarel/bmvc04.html.

[129] Jacob R.J.K. and Karn K.S. Eye tracking in human-computer interaction and usability research : Ready to deliver the promises,. In *HYONA, J., The Mind's Eyes : Cognitive and Applied Aspects of Eye Movements*. ed. : RADACH, R & DEUBEL, H. Elsevier Science, 2003.

[130] HW McGee and DJ Mace. Retroreflectivity of Roadway Signs for Adequate Visibility : A Guide. *Report No. FHWA/DF-88/001, US Department of Transportation, Federal Highway Administration, Washington, DC (November 1987)*, 1987.

[131] A. Voßküehler, V. Nordmeier, L. Kuchinke, and A.M. Jacobs. Ogama (opengazeandmouseanalyzer) : Open-source software designed to analyze eye and mouse movements in slideshow study designs. *Behavior Research Methods*, 40(4), 2008.

[132] E. Hollnagel. A function-centered approach to joint driver-vehicle system design. *Cognition, Technology & Work*, 8 :169–173, 2006.

[133] Ludovic SIMON, Jean-Philippe TAREL, and Roland BREMOND. Towards the estimation of conspicuity with visual priors. In *Third International Conference*

on Computer Vision and Application (VISAPP'08), volume 2, pages 323 – 328. Laboratoire Central des Ponts et Chaussées, DESE - EV, January 2008.

[134] Ludovic SIMON, Jean-Philippe TAREL, Roland BREMOND, and Henri PANJO. A computational model of visual search saliency for road signs diagnostic. In $32^n d$ *European Conference on Visual Perception 2009 (ECVP'09)*, page 20. Laboratoire Exploitation Perception Simulation Simulateur, LEPSiS UMR LCPC-INRETS, august 2009.

[135] Ludovic SIMON, Jean-Philippe TAREL, Roland BREMOND, and Henri PANJO. A computational model of visual search saliency for road signs diagnostic. *Perception supplement*, 38 :46, 2009.

[136] S. Boughorbel, J-P. Tarel, F. Fleuret, and N. Boujemaa. The GCS kernel for SVM based image recognition. In *Proceedings of the International Conference on Artificial Neural Networks (ICANN)*, volume 2, pages 595–600, 2005.