

Détection et reconnaissance de la signalisation verticale par analyse d'images

Ouvrage collectif coordonné par Philippe Foucher

Juillet 2010



Laboratoire Central
des Ponts et Chaussées

Laboratoire central des ponts et chaussées
58, boulevard Lefebvre, F 75732 Paris Cedex 15

Conformément à la note du 04/07/2014 de la direction générale de l'Ifsttar précisant la politique de diffusion des ouvrages parus dans les collections éditées par l'Institut, la reproduction de cet ouvrage est autorisée selon les termes de la licence CC BY-NC-ND. Cette licence autorise la redistribution non commerciale de copies identiques à l'original. Dans ce cadre, cet ouvrage peut être copié, distribué et communiqué par tous moyens et sous tous formats.



(CC BY-NC-ND 4.0)



Attribution — Vous devez créditer l'Oeuvre et intégrer un lien vers la licence. Vous devez indiquer ces informations par tous les moyens possibles mais vous ne pouvez pas suggérer que l'Ifsttar vous soutient ou soutient la façon dont vous avez utilisé son Oeuvre.



Pas d'Utilisation Commerciale — Vous n'êtes pas autorisé à faire un usage commercial de cette Oeuvre, tout ou partie du matériel la composant.



Pas de modifications — Dans le cas où vous effectuez une adaptation, que vous transformez, ou créez à partir du matériel composant l'Oeuvre originale (par exemple, une traduction, etc.), vous n'êtes pas autorisé à distribuer ou mettre à disposition l'Oeuvre modifiée.

Le patrimoine scientifique de l'Ifsttar

Le libre accès à l'information scientifique est aujourd'hui devenu essentiel pour favoriser la circulation du savoir et pour contribuer à l'innovation et au développement socio-économique. Pour que les résultats des recherches soient plus largement diffusés, lus et utilisés pour de nouveaux travaux, l'Ifsttar a entrepris la numérisation et la mise en ligne de son fonds documentaire. Ainsi, en complément des ouvrages disponibles à la vente, certaines références des collections de l'INRETS et du LCPC sont dès à présent mises à disposition en téléchargement gratuit selon les termes de la licence Creative Commons CC BY-NC-ND.

Le service Politique éditoriale scientifique et technique de l'Ifsttar diffuse différentes collections qui sont le reflet des recherches menées par l'institut :

- Les collections de l'INRETS, Actes
- Les collections de l'INRETS, Outils et Méthodes
- Les collections de l'INRETS, Recherches
- Les collections de l'INRETS, Synthèses
- Les collections du LCPC, Actes
- Les collections du LCPC, Etudes et recherches des laboratoires des ponts et chaussées
- Les collections du LCPC, Rapport de recherche des laboratoires des ponts et chaussées
- Les collections du LCPC, Techniques et méthodes des laboratoires des ponts et chaussées, Guide technique
- Les collections du LCPC, Techniques et méthodes des laboratoires des ponts et chaussées, Méthode d'essai



Institut Français des Sciences et Techniques des Réseaux,
de l'Aménagement et des Transports
14-20 Boulevard Newton, Cité Descartes, Champs sur Marne
F-77447 Marne la Vallée Cedex 2

Contact : diffusion-publications@ifsttar.fr

www.ifsttar.fr



Ouvrage collectif coordonné par **Philippe Foucher**
Chargé de recherche
ERA 27 LCPC
Laboratoire régional des ponts et chaussées de Strasbourg

Ce recueil présente les méthodes de détection et reconnaissance des panneaux de signalisation développées par quatre équipes du réseau scientifique et technique du ministère de l'Écologie, de l'Énergie, du Développement durable et de la Mer (MEEDDM) :

- le laboratoire LEPSiS (INRETS-LCPC),
- le laboratoire MATIS (IGN),
- les ERA 27 et 32 du LCPC (CETE de l'Est), en collaboration avec des laboratoires universitaires.

En couverture :
exemples de détection et de reconnaissance de la signalisation verticale
sur une image de scène routière (image originale prise par le LRPC Rouen)

Pour commander cet ouvrage :
Laboratoire central des ponts et chaussées
DISTC - Section Diffusion
58, boulevard Lefebvre, F 75732 Paris cedex 15
Téléphone : 01 40 43 50 20 - Télécopie : 01 40 43 54 95
Ou serveur Internet LCPC : www.lcpc.fr
Prix : 30 Euros HT

Ce document est propriété du Laboratoire central des ponts et chaussées
et ne peut être reproduit, même partiellement, sans l'autorisation de son Directeur général
(ou de ses représentants autorisés)

© 2010 - LCPC
ISSN 1160-9761
ISBN 978-2-7208-2578-1
DOI/Crossref 10.3829/erlpc.cr53-fr

Table des matières

| | | |
|----------|---|-----------|
| 1 | Introduction | 9 |
| 2 | État de l'art | 13 |
| 2.1 | Méthodes de Détection | 14 |
| 2.1.1 | Sélection colorimétrique | 14 |
| 2.1.2 | Caractéristiques géométriques | 14 |
| 2.1.3 | Combinaison de critères colorimétriques et géométriques | 15 |
| 2.1.4 | Méthodes avec apprentissage | 16 |
| 2.2 | Méthodes de reconnaissance | 16 |
| 2.3 | Évaluation des performances | 17 |
| 2.3.1 | Outils d'évaluation | 18 |
| 2.3.2 | Analyse des performances | 19 |
| 2.4 | Systèmes opérationnels | 19 |
| 2.5 | Conclusion | 20 |
| 3 | Détection par la couleur et par transformations univariées | 27 |
| 3.1 | Bases de données | 28 |
| 3.1.1 | Séquences d'images | 28 |
| 3.1.2 | Vérité-terrain | 28 |
| 3.1.3 | Métriques d'évaluation | 29 |
| 3.2 | Description de l'algorithme | 30 |
| 3.2.1 | Filtre colorimétrique | 30 |
| 3.2.2 | Utilisation des informations géométriques | 31 |
| 3.2.3 | Sélection des cercles par symétrie radiale | 31 |
| 3.2.4 | Sélection des polygones par transformée de Hough | 33 |
| 3.3 | Résultats expérimentaux | 34 |
| 3.3.1 | Réglages des paramètres couleur | 35 |
| 3.3.2 | Détermination des seuils de symétrie | 37 |
| 3.3.3 | Détection des panneaux polygonaux | 38 |
| 3.3.4 | Validation des résultats | 38 |
| 3.4 | Conclusion | 38 |
| 4 | Détection et reconnaissance par ajustement de modèles géométriques et radiométriques | 43 |
| 4.1 | Notre stratégie | 43 |

| | | |
|----------|---|-----------|
| 4.2 | Sélection des couleurs | 44 |
| 4.3 | Détection des formes des panneaux | 45 |
| 4.3.1 | Extraction de points de contour orientés | 45 |
| 4.3.2 | Détection d'ellipse | 47 |
| 4.3.3 | Détection de triangle et quadrilatère | 47 |
| 4.3.4 | Choix de la meilleure forme | 48 |
| 4.4 | Validation et identification d'hypothèses de panneaux | 48 |
| 4.4.1 | Rectification locale d'image | 48 |
| 4.4.2 | Appariement de texture avec les panneaux de référence | 50 |
| 4.5 | Résultats et évaluation | 51 |
| 4.5.1 | Résultats qualitatifs | 51 |
| 4.5.2 | Résultats quantitatifs | 51 |
| 4.6 | Conclusion | 54 |
| 5 | Détection des Panneaux de Signalisation Routière par Accumulation Bivariée | 57 |
| 5.1 | Détection de panneaux circulaires et rectangulaires | 57 |
| 5.1.1 | Transformation en Symétrie Généralisée TSG | 57 |
| 5.1.2 | Transformation Chinoise TC | 59 |
| 5.1.3 | Transformation Chinoise Bilatérale TCB | 59 |
| 5.2 | Détection de panneaux triangulaires | 62 |
| 5.2.1 | De la TSG à la détection de coins | 63 |
| 5.2.2 | Transformation en Sommet et Bissectrice TSB | 63 |
| 5.2.3 | Détection de Signalisation Triangulaire | 66 |
| 5.3 | Évaluations expérimentales | 67 |
| 5.3.1 | Les bases de test | 67 |
| 5.3.2 | Performances en détection de panneaux circulaires et rectangulaires | 68 |
| 5.3.3 | Détection de panneaux de signalisation routière triangulaires | 69 |
| 5.4 | Conclusion | 70 |
| 6 | Etude comparative de trois méthodes | 73 |
| 6.1 | Base de données | 73 |
| 6.2 | Comparaison des détecteurs de panneaux | 74 |
| 6.2.1 | Protocole d'évaluation | 74 |
| 6.2.2 | Métriques d'évaluation | 74 |
| 6.2.3 | Performances de détection | 75 |
| 6.2.4 | Cas des panneaux triangulaires | 77 |
| 6.3 | Conclusion | 78 |
| 7 | Ajustement de forme prototype par méta-heuristiques biologiques | 83 |
| 7.1 | Définition du problème d'estimation | 84 |
| 7.1.1 | Représentation de la forme : le prototype | 84 |
| 7.1.2 | Représentation des déformations | 84 |
| 7.1.3 | Fonctionnelle d'énergie | 85 |
| 7.2 | Optimisation par méta-heuristiques biologiques | 86 |
| 7.2.1 | La stratégie d'évolution : une métaphore darwinienne | 86 |

| | | |
|-----------|--|------------|
| 7.2.2 | La sélection clonale : une métaphore immunitaire | 87 |
| 7.2.3 | L'essaim de particules : une métaphore sociale | 87 |
| 7.2.4 | Mise en œuvre et validation | 89 |
| 7.3 | Conclusion | 89 |
| 8 | Modèles probabilistes d'apparence et estimation robuste | 95 |
| 8.1 | Détection d'objets par modèles probabilistes d'apparence | 96 |
| 8.1.1 | Décomposition sur espace propre | 96 |
| 8.1.2 | Le modèle d'observation | 97 |
| 8.1.3 | Détection et reconnaissance dans le cadre bayésien | 97 |
| 8.1.4 | Estimation robuste au sens du MAP | 99 |
| 8.2 | Modèles non paramétriques dans l'espace propre et optimisation | 103 |
| 8.2.1 | Le <i>Mean Shift</i> : un algorithme semi-quadratique | 103 |
| 8.2.2 | <i>Mean Shift</i> et estimation du MAP gaussien | 105 |
| 8.2.3 | Modélisation non gaussienne dans l'espace propre et estimation robuste de l'apparence | 105 |
| 8.3 | Conclusion | 106 |
| 9 | Estimation de la saillance visuelle | 111 |
| 9.1 | La saillance visuelle | 112 |
| 9.2 | Détection des zones saillantes | 113 |
| 9.3 | Estimation de la saillance des panneaux | 116 |
| 9.4 | Expérimentations | 117 |
| 9.5 | Conclusion | 120 |
| 10 | Conclusion générale | 123 |

Résumé

La détection et la reconnaissance de la signalisation verticale représentent un enjeu important en analyse de scènes routières. Leurs applications sont nombreuses. Citons par exemple les systèmes d'aide à la conduite, la gestion du patrimoine routier, la sécurité routière, ou encore la mise au point d'une nouvelle génération d'outils multimedia sur le web pour la navigation 3D géographique. Sur le plan méthodologique, elles recèlent d'importantes difficultés, liées à la fois au caractère non contrôlé des prises de vues employées et à la variabilité d'aspect des objets recherchés. Ces problématiques ont donc, naturellement, donné lieu à une grande variété de contributions au cours de ces deux dernières décennies. On trouve dans la littérature du domaine, des techniques (souvent motivées par le contexte temps-réel des applications d'aide à la conduite) exploitant judicieusement des informations de forme et de couleurs extraites des images par des algorithmes « bas-niveau », aussi bien que des méthodes plus sophistiquées de reconnaissance des formes, mettant en jeu des techniques d'apprentissage statistique ou des problèmes complexes d'optimisation. Ce recueil présente les méthodes de détection et reconnaissance des panneaux de signalisation développées par quatre équipes du Réseau Scientifique et Technique du Ministère de l'Écologie, de l'Énergie, du Développement Durable et de la Mer (MEEDDM) : le laboratoire LEPSiS (INRETS/LCPC), le laboratoire MATIS (IGN) et les ERA 27 et 32 du LCPC (CETE de l'Est), en collaboration avec des laboratoires universitaires. Il donne des éléments quantitatifs d'évaluation des algorithmes proposés et les situe par rapport à l'état de l'art du domaine.

Abstract

Traffic sign detection and recognition are an important issue in road scene analysis. Their applications are numerous. Examples include driver assistance systems, road asset management, road safety, or the development of a new generation of multimedia tools for 3D geographic web browsing. Methodologically speaking, they include significant difficulties related to both the uncontrolled nature of the images used and the variability of appearance of the objects sought. Therefore, these issues naturally lead to a wide variety of contributions during the past two decades. One may find in the literature of the field, techniques (often motivated by the real-time context of driving assistance) shrewdly exploiting the shape and color information extracted from the images using “low-level” algorithms, as well as more sophisticated pattern recognition methods, involving statistical learning techniques or complex optimization problems. This collective book presents the traffic sign detection and recognition methods that were developed by four teams of the Technical and Scientific Network of the French Ministry of Ecology, Energy, Sustainable Development and the Sea (MEEDDM), namely : the LEPSiS laboratory (INRETS/LCPC), the MATIS laboratory (IGN) and the ERA’s 27 and 32 LCPC (CETE de l’Est), in collaboration with academic laboratories. It provides quantitative assessment of the proposed algorithms and situates them in the state of the art of the field.

Chapitre 1

Introduction

La détection et la reconnaissance de la signalisation verticale est une problématique majeure dans l'analyse de scènes routières par traitement d'images. Les applications sont nombreuses, telles que le calcul d'itinéraire avec estimation des temps de parcours, le développement d'outils pour la gestion et la maintenance du patrimoine routier, les systèmes d'aide à la conduite en temps réel ou, en lien avec la robotique, l'automatisation des véhicules ou encore le mise au point d'une nouvelle génération d'outils multimedia sur le web pour la navigation 3D géographique.

Quelles que soient les applications, les méthodes de détection et de reconnaissance se heurtent aux difficultés classiquement rencontrées en analyse d'images en environnement non contrôlé. Même si la qualité des images s'est largement améliorée ces dernières années avec l'apparition de capteurs numériques toujours plus performants, les variations d'éclairement engendrent des modifications de couleur apparente, des ombres portées, des réflexions ou des contre-jours. Par ailleurs, des distorsions géométriques ou des rotations peuvent apparaître suivant l'angle d'observation et l'échelle des panneaux. Leur aspect peut également varier en fonction de leur état d'usure et d'éventuelles salissures ou dégradations. Enfin la scène routière en elle-même, de par sa complexité, est source de nombreux distracteurs et d'occultations partielles des objets. Les algorithmes développés pour la détection et l'identification de la signalisation verticale doivent répondre de manière robuste aux diverses perturbations observées et tenir compte de la variabilité d'aspect des panneaux.

Cependant, les panneaux de signalisation sont des objets manufacturés et normalisés. En tant que tels, ils présentent donc un certain nombre de caractéristiques géométriques, colorimétriques et pictographiques permettant de les distinguer de leur environnement. Dans cet ouvrage, nous nous focalisons sur les panneaux de danger, d'intersection, de priorité, de prescription et d'indication, ainsi que sur quelques balises, qui forment ce que l'on appelle parfois, en langage commun, la signalisation de police. Ces panneaux se caractérisent majoritairement par des couleurs vives, principalement le rouge et le bleu, par des formes géométriques simples (polygones, disques) et par un nombre relativement restreint de pictogrammes. Les méthodes proposées ne considèrent pas le cas des panneaux directionnels qui nécessitent la détection du panneau et de texte. De même, nous ne considérons pas les panonceaux, souvent achromatiques et de petites dimensions. Leur détection représente encore, à ce jour, un problème peu exploré.

Les difficultés inhérentes à l'analyse de la signalisation verticale, mais aussi les atouts présentés par le caractère particulier des objets considérés, expliquent la grande variété des mé-

thodes proposées dans la littérature au cours des deux dernières décennies. Leur évolution a accompagné l'essor des algorithmes modernes de détection et de reconnaissance des formes ainsi que les progrès des moyens informatiques disponibles. Elle se concrétise par l'apparition récente de systèmes embarqués, fonctionnant en temps réel à bord de certains véhicules automobiles du commerce, sur une gamme relativement réduite de panneaux.

Le but de cet ouvrage, proposé sous forme de recueil de contributions, est triple. Il s'agit en premier lieu de présenter les travaux de recherche relatifs à cette thématique menés par quatre équipes du *RST*¹, en collaboration avec des laboratoires universitaires. Tous les travaux décrits ici se focalisent sur l'analyse *hors-ligne* des images de scènes routières, dans le cadre d'applications de relevés des signalisations ou d'indexation de bases de données d'images. L'objectif est de contribuer à automatiser cette tâche fastidieuse pour l'opérateur, et propice aux oublis et aux erreurs. Dans ce contexte, on peut relâcher quelque peu la contrainte de temps réel, même si des temps de calculs réduits demeurent évidemment souhaités. Certaines solutions proposées exploitent les informations de forme et de couleur liées à la signalisation verticale, extraites des images par des algorithmes rapides, « bas-niveau ». D'autres mettent en œuvre des algorithmes plus sophistiqués de reconnaissance de forme, qui nécessitent un apprentissage statistique ou la résolution de problèmes complexes d'optimisation. En second lieu, l'accent est mis sur l'évaluation quantitative des performances individuelles de chacune des méthodes proposées, à partir de bases d'images indépendantes entre elles. De plus, les trois méthodes les plus récemment développées font l'objet d'une analyse comparative de performances sur une base de test commune, réalisée dans le cadre du projet iTOWNS². Enfin, le troisième objectif est de situer les approches proposées par rapport aux différents travaux de recherche menés dans ce domaine au niveau international ces vingt dernières années.

Ce document s'ouvre donc assez naturellement, au **chapitre 2**, par une présentation volontairement synthétique des différentes méthodologies rencontrées dans la littérature du domaine. Il ressort de ce chapitre que la grande majorité des méthodes existantes décompose le processus en deux phases : la détection des panneaux dans la scène routière et l'identification de ceux-ci. Dans ce contexte, chaque étape peut ainsi être évaluée indépendamment.

Les quatre chapitres suivants sont consacrés à la présentation des méthodes proposées indépendamment par les trois équipes collaborant dans le cadre du projet iTOWNS, et à leur évaluation comparative sur une base d'images commune. Le réglage des paramètres de chaque méthode est discuté dans le chapitre correspondant, ainsi que le choix entre différentes variantes.

Le **chapitre 3** décrit l'algorithme développé à l'ERA³ 27 du LRPC de Strasbourg. Cette méthode concerne uniquement la phase de détection. Elle est composée d'une étape de segmentation des pixels rouges et bleus, suivie par une sélection selon des critères géométriques en utilisant des techniques de vote univarié, c'est-à-dire pixel par pixel, de type Transformée de Hough (TH) ou Transformée de Symétrie Radiale (TSR).

La contribution du laboratoire MATIS de l'IGN fait l'objet du **chapitre 4**. La phase de détection des panneaux est traitée par un algorithme en deux étapes, qui débute par une segmenta-

¹Réseau Scientifique et Technique du Ministère de l'Écologie, de l'Énergie, du Développement Durable et de la Mer (MEEDDM).

²iTOWNS : image-based Town On-Line Web Navigation and Searchengine. Projet financé par l'Agence Nationale de la Recherche (ANR) sur la période 2008-2010, dans le cadre du thème Masse de Données et Connaissances (MDCO).

³Équipe de Recherche Associée

tion des pixels rouges et des pixels bleus, suivie de l'ajustement d'un modèle géométrique aux contours des panneaux, de forme elliptique, triangulaire ou rectangulaire. Dans un deuxième temps, un algorithme de reconnaissance par corrélation entre le panneau détecté et l'ensemble des objets d'une base de données de pictogrammes de panneaux est détaillé et évalué.

Le **chapitre 5** décrit la méthode développée par le laboratoire LEPSiS (INRETS/LCPC), qui exploite la somme des normes des gradients rouge et bleu, donc sans pousser la segmentation jusqu'à son terme. Cette méthode est fondée sur une technique de vote bivarié, c'est-à-dire impliquant des paires de pixels.

Dans le **chapitre 6**, ces trois algorithmes font l'objet d'une évaluation systématique comparative, sur la même base d'images, exploitant la même *vérité terrain* (référence construite par analyse manuelle des images), selon les mêmes critères objectifs. Les résultats obtenus sont comparés et discutés afin de dégager les principaux axes d'amélioration envisageables pour chaque algorithme.

Les deux chapitres suivants présentent des travaux antérieurs à ceux d'iTOWNS. La motivation de ces travaux était essentiellement d'ordre méthodologique, l'idée étant d'adapter à notre problème des méthodes éprouvées dans d'autres domaines et présentant un caractère novateur. Les outils décrits dans ces deux chapitres sont susceptibles d'être mis en œuvre de manière complémentaires avec les trois méthodes développées dans le cadre du projet iTOWNS. Les méthodes proposées ont en commun de mettre en œuvre des *modèles* des objets recherchés. Plus précisément, ces derniers prennent en compte des informations de plus haut niveau portant soit la silhouette, soit l'apparence des panneaux. Le problème de la détection et reconnaissance de la signalisation verticale est formulé dans le cadre statistique bayésien.

Le **chapitre 7** présente une méthode de détection de panneaux de danger fondée sur l'utilisation de modèles déformables. Développée au LRPC de Strasbourg (ERA 27 et 32), elle consiste à définir un modèle mathématique de la forme du panneau et à l'ajuster sur l'image. La détection devient alors un problème d'optimisation, qui consiste à minimiser une fonctionnelle d'énergie. Cette dernière, qui prend en compte la qualité de l'ajustement et le degré de déformation du modèle, est fortement non convexe et son optimisation est réalisée à l'aide d'algorithmes d'optimisation globale, inspirés de mécanismes biologiques. La méthode étant relativement coûteuse en temps de calcul, une étape de pré-détection par la couleur, préfigurant l'algorithme présenté au chapitre 3, a été mise au point dans le cadre de ces travaux.

Le **chapitre 8** décrit une approche fondée sur l'apprentissage statistique d'une représentation globale de l'apparence des panneaux. Ces travaux ont été menés conjointement par l'ERA 27 et le laboratoire LSIT (Université de Strasbourg/CNRS). Le problème de détection et de reconnaissance implique un problème de régression paramétrique, traité ici dans le cadre bayésien de l'optimisation. Ces recherches ont abouti sur un algorithme d'estimation robuste dont les performances, évaluées sur données synthétiques et sur données réelles acquises en conditions contrôlées, se sont avérées du niveau des meilleures méthodes alors disponibles, et dont le champ d'application déborde de celui de l'analyse de la signalisation.

L'approche proposée par laboratoire LEPSiS (INRETS/LCPC) dans le **chapitre 9** a un cadre applicatif particulier, puisqu'il concerne l'estimation de la saillance visuelle d'un panneau, c'est-à-dire son « efficacité » en termes de perception visuelle. La détection utilise une méthode d'apprentissage statistique nommée *Support Vector Machine* ou Séparateur à Vaste Marge (SVM). Une base d'apprentissage est construite, dans laquelle des exemples positifs (correspondant un aspect du panneau) et d'exemples négatifs (correspondant au fond) sont inté-

1. INTRODUCTION

grés. Chaque exemple de la base est défini par un vecteur de caractéristiques. Le SVM vise alors à déterminer la frontière qui divise au mieux l'espace de caractéristiques en deux catégories.

Enfin, le **chapitre 10** établit un récapitulatif des contributions présentées dans cet ouvrage et présente l'intégration de ces outils dans des logiciels développés au sein du *RST* en vue d'une utilisation opérationnelle. Plusieurs perspectives de travail, d'ordre méthodologique ou applicatif, sont également proposées.

Chapitre 2

État de l’art

*Philippe Foucher,
Pierre Charbonnier,
Jean-Philippe Tarel*

Cet état de l’art présente de manière synthétique les approches utilisées pour rechercher et identifier la signalisation verticale dans les images de scènes routières. De nombreuses équipes se sont intéressées à cette problématique (voir [34, 4, 44] pour différentes synthèses bibliographiques) et la grande majorité d’entre elles ont décomposé le processus en deux, voire trois étapes :

- détection d’objets dans l’image qui correspondent probablement à des panneaux de signalisation. Cette étape est considérée comme la plus difficile par certains auteurs [23],
- reconnaissance des candidats détectés par des méthodes de classification,
- suivi temporel de l’objet candidat dans la séquence d’images. Cette étape, parfois utilisée pour améliorer les performances de l’algorithme, ne sera pas présentée dans cet ouvrage. Nous renvoyons le lecteur à [3] pour un exemple d’utilisation dans un système d’aide à la conduite.

Les panneaux de signalisation sont des objets manufacturés et standardisés dont la forme, les dimensions, la couleur et la position sont fixées par des normes [35]. Les méthodologies de détection et de reconnaissance s’appuient sur une ou plusieurs de ces caractéristiques pour sélectionner les objets dans l’image. Ces approches doivent être robustes aux situations difficiles, classiquement rencontrées en environnement non contrôlé : variations des conditions lumineuses, distorsions géométriques, complexité de la scène, vétusté de l’objet...

Une présentation des différents travaux de recherche est proposée dans la section 2.1 pour la détection et la section 2.2 pour la reconnaissance. La section 2.3 aborde les performances des algorithmes, en présentant les outils d’évaluation. Enfin, la section 2.4 présente quelques exemples de systèmes opérationnels.

2.1 Méthodes de Détection

2.1.1 Sélection colorimétrique

Dans de nombreux cas, une première étape de détection consiste à identifier les pixels d'une image en couleurs correspondant aux caractéristiques colorimétriques du panneau. Il est alors nécessaire de fixer un ou des seuils au-delà duquel le pixel est classé comme appartenant au panneau ou non. De nombreux auteurs travaillent dans l'espace colorimétrique HSV et la segmentation est le plus souvent effectuée par seuillage de la composante de teinte (Hue), insensible aux changements de luminance. Cette approche est utilisée par [15] pour la sélection de panneaux rouges. Pour cette même catégorie de panneaux, une alternative, tout aussi efficace, consiste à considérer la composante normalisée $R/(R+V+B)$, également indépendante des variations lumineuses [18, 19] ou, de manière similaire, en prenant en compte les rapports R/B et V/B [17, 57]. Dans le cadre d'applications en temps réel, l'espace colorimétrique YUV, obtenu à partir de transformations linéaires de l'espace RVB est utilisé par [11]. Dans ces mêmes travaux, une correction chromatique est initialement effectuée sur toute l'image. Le coefficient de correction est obtenu en se basant sur les valeurs chromatiques des pixels de la chaussée, théoriquement grise (R , V et B égaux). Des espaces plus complexes comme CIE¹ Lab [59] ou CIECAM97 [54] sont aussi utilisés. Quelles que soient les méthodes, le choix des seuils est délicat. Ils sont le plus souvent fixés à partir des normes concernant les panneaux ou à partir de méthodes empiriques. On peut noter cependant les travaux de [48], proposant un outil de seuillage adaptatif fondé sur l'algorithme d'Otsu [46] et ceux de [61] proposant un seuillage dynamique dans l'espace HSV.

L'utilisation de la couleur permet de sélectionner rapidement des zones d'intérêt dans l'image. Cependant elle présente l'inconvénient de focaliser rapidement la recherche sur une catégorie de panneaux (rouge, bleu, vert ou jaune). D'autre part, pour certaines signalisations, comme les panneaux blancs (fin d'interdiction), la couleur n'est pas une information suffisamment discriminante pour être utilisée.

2.1.2 Caractéristiques géométriques

Certaines équipes se basent sur l'analyse des caractéristiques géométriques des contours extraits de l'image de luminance. Dans le cas de la détection de panneaux triangulaires, une méthode consiste à filtrer les segments, issus de la détection de contours, en fonction de leurs pentes et de leurs longueurs et à vérifier si les segments restants appartiennent à un même triangle équilatéral [49]. Une méthode intéressante proposée dans [52] permet de calculer pour un objet donné des mesures de ressemblances aux formes de types cercle, rectangle et triangle.

Quelques auteurs proposent d'utiliser les algorithmes de vote, comme la transformée de Hough [27], la transformée de Hough circulaire [23], la transformée chinoise [8, 7], la transformée en symétrie radiale [37, 4] et sa version étendue à la détection des polygones [36, 29]. Ces algorithmes de vote seront détaillés dans les chapitres 3 et 5.

¹Commission internationale de l'éclairage

2.1. MÉTHODES DE DÉTECTION

Une autre approche repose sur la recherche d'un modèle simplifié du panneau dans une image candidate. Les techniques de *template matching* peuvent alors être mises en œuvre sur une carte de distance aux contours calculée à partir de l'image en niveaux de gris [25]. Une autre possibilité revient à ajuster un modèle simplifié du panneau dans une image candidate. L'optimisation de cet ajustement est ensuite réalisée par des algorithmes génétiques [2].

En s'affranchissant de l'information couleur, les algorithmes proposés sont susceptibles de sélectionner tous types de panneaux, y compris les blancs. Ainsi [12] propose de détecter les panneaux de fin de limite de vitesse en recherchant les transitions "clair-sombre-clair" et en appliquant ensuite la transformée de Hough circulaire autour des zones d'intérêt. Cependant, l'efficacité de ces méthodes reste conditionnée à la qualité de l'extraction de contours et donc à l'amplitude et l'orientation des gradients. Par ailleurs, ces algorithmes conduisent généralement à des temps de calcul plus importants que lorsque la couleur est utilisée.

2.1.3 Combinaison de critères colorimétriques et géométriques

La stratégie la plus utilisée consiste à combiner les informations couleurs et géométriques. Les algorithmes de segmentation couleurs définis dans la section 2.1.1 peuvent ainsi être utilisés comme une première étape pour la détection de panneaux de signalisation. À l'issue de cette pré-détection, une analyse géométrique des composantes connexes permet d'affiner la sélection des objets. Il est ainsi possible d'éliminer rapidement les objets selon leurs tailles [39, 22] ou la valeur de la compacité, définie par le rapport entre l'aire de l'objet et son périmètre au carré [30]. Toutefois des descripteurs plus élaborés sont souvent nécessaires pour effectuer une sélection plus fine. Le choix s'oriente le plus souvent vers la caractérisation du contour de la forme en utilisant par exemple les descripteurs de Fourier obtenus par la transformée de Fourier normalisée de la signature de l'objet [26]. La circularité des objets rouge peut également être contrôlée selon l'approche définie dans [28]. La distance des bords du contour de l'objet aux bords de sa boîte englobante de l'objet est également une caractéristique intéressante qui peut permettre de discriminer cercles, triangles et rectangles [33, 32].

La sensibilité de ces approches face à la variabilité des situations (occultations, changement d'aspect...) représente une difficulté majeure.

À partir de l'image issue du filtre colorimétrique, les modèles d'ajustement de forme peuvent également s'appliquer en utilisant les algorithmes génétiques [15, 16], les algorithmes de recuit simulé [16] ou les algorithmes de sélection clonale ou d'essaims de particules [19]. Dans les chapitres 4 et 7, deux systèmes de détection de panneaux triangulaires par sélection colorimétrique et ajustement d'un modèle géométrique (RANSAC ou algorithme génétique) seront présentés.

L'efficacité de ces méthodes est liée aux performances de la segmentation colorimétrique. Le regroupement des pixels en composantes connexes peut en particulier s'avérer problématique puisqu'il conditionne la forme de l'objet à analyser par la suite. Une sur-segmentation entraîne une division de l'objet d'intérêt en différentes parties. Au contraire, une sous-segmentation présente un risque de fusion de plusieurs composantes. On peut cependant penser que les méthodes de détection robustes, notamment vis-à-vis des occultations, sont susceptibles d'être

robustes vis-à-vis d'une segmentation colorimétrique, que l'on pourrait qualifier de moyenne. Une alternative est d'introduire plus profondément l'information de couleur dans les méthodes géométriques par exemple en sélectionnant ou en pénalisant les contours en fonction des couleurs voisines. Une illustration de cette approche alternative est proposée dans [8, 7] (voir le chapitre 5).

2.1.4 Méthodes avec apprentissage

Certains travaux de détection sont basés sur les méthodes de classification qui consistent à comparer le vecteur de caractéristiques d'une observation à un vecteur de référence. Pour obtenir ces références, un apprentissage est classiquement réalisé sur un ensemble d'images représentatives. Différentes techniques d'apprentissage peuvent être utilisées, les plus connues étant Adaboost et SVM (*Support Vector Machine*). Moins classique, une méthode non supervisée, basée sur l'analyse des changements statistiques temporels des objets d'intérêt dans les séquences (apparition, grossissement, disparition), est proposée dans [14]. Les histogrammes multi-dimensionnels (de mesures de couleur ou de forme) entre deux images successives sont ainsi comparés et les objets sont localisés par rétro-projection d'histogramme. Dans ces travaux [14], une approche de détection par apprentissage est également présentée à partir de représentations globales de l'apparence de l'objet. En particulier, des détecteurs robustes, fondés sur l'hypothèse que l'image observée est une occurrence bruitée d'une image d'apprentissage, sont proposés.

Enfin, les algorithmes de type Adaboost, dans lesquels un modèle de classification dit « fort » est construit à partir d'une combinaison de classifieurs dits « faibles » peuvent être développés dans ce cadre. Le modèle Adaboost en cascade défini par [60], considère un jeu de caractéristiques basé sur les ondelettes de Haar. Il est utilisé comme détecteur de panneaux sur des images en niveaux de gris par [13, 5]. Notons cependant quelques différences entre ces deux approches. En effet, dans [13], le détecteur opère sur l'ensemble de l'image alors qu'une variante du jeu d'ondelettes de Haar, appelée dipôles dissociés est proposée par [5]. Une dernière approche consiste à prendre en compte l'information couleur en intégrant dans le détecteur sept images monochromes (les composantes R , V , B , les composantes R,V,B normalisées et l'image de luminance) [3]. Un autre exemple de détecteur fondé cette fois sur un apprentissage par SVM et utilisant des caractéristiques de forme et de couleurs est décrit par exemple dans [55]. Les performances de ces méthodes restent conditionnées à la qualité de l'apprentissage effectué au préalable.

2.2 Méthodes de reconnaissance

L'étape de détection permet de générer une liste d'objets représentant potentiellement des panneaux de signalisation. La phase de reconnaissance consiste à identifier l'objet candidat parmi un ensemble de référence comprenant les panneaux à répertorier, ou à rejeter l'objet candidat si aucun modèle ne correspond. Parmi ces méthodes, de nombreux auteurs font appel aux techniques classiques de corrélation en comparant les objets candidats à chaque modèle de référence. En particulier la fonction ZNCC (*Zero mean Normalized Cross-Correlation*) permet

2.3. ÉVALUATION DES PERFORMANCES

de rester invariant aux changements d'illumination. Les objets candidats sont ré-échantillonnés pour être remis à l'échelle des panneaux de référence et la corrélation est effectuée le plus souvent sur l'ensemble du panneau [49, 41, 16, 56, 57]. La corrélation peut également être réalisée en comparant le pictogramme extrait de l'objet candidat à une base de pictogrammes de référence [40] ou grâce à un système de reconnaissance de chiffres pour les panneaux de limite de vitesse [42].

Les objets sélectionnés par le détecteur présentent parfois une déformation géométrique due à l'effet de perspective. Cette distorsion peut compliquer l'appariement de l'objet à une référence de la base, construite à partir d'images fronto-parallèles. Pour pallier cette difficulté, il est possible de rectifier l'image candidate, correspondant à la boîte englobante de la composante connexe détectée, pour la remettre dans une géométrie correspondante à l'image de référence [57]. Des outils invariants comme les descripteurs SIFT peuvent également être proposés dans [20].

Les autres méthodes choisies pour la reconnaissance nécessitent un apprentissage. Elles sont souvent basées sur des réseaux de neurones, et en particulier, le perceptron multi-couches. Dans [2], les objets sont classés parmi trois catégories « panneau de limite de vitesse », « autres panneaux de signalisation » et « rejet ». Le même type de classifieur est également utilisé par [50, 17, 58, 1]. Plus rarement, d'autres types de réseaux de neurones ont été considérés dans le cadre de cette application. Citons les réseaux de fonction à base radiale [62, 24] et les réseaux ART (*Adaptive Resonance Theory*) [15]. Les méthodes de classification comme Adaboost et SVM peuvent également être envisagées dans l'étape de reconnaissance [53]. Cette phase d'identification peut être effectuée par les méthodes bayésiennes, soit par estimation d'une distribution gaussienne [3] en incluant une information de suivi temporel, soit à partir d'un modèle de type fenêtre de Parzen à base d'un noyau de Laplace [47] en utilisant les moments géométriques comme attributs. Enfin, de façon similaire aux outils de détection robuste définis dans la section précédente, un algorithme de reconnaissance robuste a été développé dans [14] et appliqué à l'identification de panneaux de signalisation. Cette dernière partie sera plus amplement détaillée dans le chapitre 8.

2.3 Évaluation des performances

Dans la littérature, les auteurs s'attachent le plus souvent à détailler l'aspect méthodologique du système de détection et de reconnaissance de panneaux. Les outils d'évaluation utilisés et l'analyse des performances y sont présentés de manière succincte. De plus, l'évaluation concerne le plus souvent l'algorithme dans son ensemble, c'est-à-dire à la fin de la dernière étape de traitement (détection et reconnaissance) et les performances des étapes intermédiaires (le détecteur seul par exemple) ne sont pas décrites.

2.3.1 Outils d'évaluation

Pour évaluer les performances d'un algorithme de détection et de reconnaissance de panneaux, les résultats en sortie de l'algorithme sont en général comparés à une vérité-terrain, dans laquelle les panneaux de signalisation sont relevés par un opérateur humain. On peut ainsi facilement calculer un taux de détections correctes (TDC) et un taux de fausses alarmes (TFA) correspondant aux objets détectés et/ou reconnus par l'algorithme ne correspondant pas à la vérité-terrain. Ces taux sont définis selon :

$$\text{TDC}(\%) = \frac{\text{VP}}{\text{N}} \quad \text{TFA}(\%) = \frac{\text{FP}}{\text{NbImg}}$$

Suivant les auteurs et les applications, les définitions de VP, N, FP et NbImg peuvent varier. Le plus généralement, dans le cadre de la détection de panneaux, on comptabilise le nombre d'objets correctement détectés : VP est ainsi le nombre de vrai positif. Cependant, lorsque les images sont acquises de manière successive dans une séquence par un véhicule en mouvement, un panneau apparaît sur plusieurs images et la possibilité de le détecter et de l'identifier est multiple. Ainsi, le nombre N peut être choisi comme le nombre de panneaux différents vus dans la séquence, ou alors comme le nombre total d'apparitions de panneaux dans une image de la séquence. Le deuxième cas est plus sévère car les meilleures performances sont obtenues seulement lorsque toutes les apparitions d'un panneau sont détectées. Dans [22], les résultats sont présentés en distinguant ces deux définitions. D'autres auteurs présentent et comparent les résultats en fonction du nombre de répétition des détections d'un même panneau pris dans des images successives [4]. FP correspond le plus souvent au nombre total d'objets détectés à tort comme panneau et NbImg est le nombre total d'images de la séquence. On remarque que ce taux de fausses alarmes peut ainsi être supérieur à 100% (dans ce cas, le nombre moyen de fausses alarmes par image est supérieur à 1). Cette possibilité est assez inhabituelle dans la littérature. On la retrouve cependant dans [38]. FP peut également correspondre au nombre d'images contenant au moins une fausse alarme [22]. Dans ce cas, le taux maximum est obligatoirement inférieur ou égal à 1.

A partir de ces résultats, deux outils sont classiquement utilisés pour comparer les méthodes et optimiser les paramètres :

- La courbe « Caractéristiques Opérationnelles du Récepteur (COR) » correspond aux variations du taux de détections correctes en ordonnée et du taux de fausses alarmes en abscisse en fonction des valeurs d'un paramètre. Dans les cas simples, une courbe située au-dessus d'une autre correspond à la méthode la plus efficace mais le plus souvent, les courbes se croisent et l'analyse est plus complexe [21]. Elle dépend en particulier du taux de détections correctes minimum acceptable ou du taux de fausses alarmes maximum toléré.
- le coefficient de similarité de Dice (*DSC*) également appelé F-mesure, est utilisé en complément de la courbe COR. Il est défini de la manière suivante :

$$\text{DSC} = \frac{2\text{VP}}{\text{VP} + \text{FP} + \text{N}}$$

Il permet notamment de déterminer la valeur optimale d'un paramètre, c'est-à-dire, celui offrant le meilleur « compromis » entre détections correctes et fausses alarmes. Par ailleurs, la largeur du pic de la courbe du *DSC* renseigne sur la sensibilité du détecteur par

2.4. SYSTÈMES OPÉRATIONNELS

rapport à la valeur du paramètre choisi. Plus le pic est étroit, plus la valeur du paramètre va influencer sur l'efficacité du détecteur.

2.3.2 Analyse des performances

Les travaux présentés ci-dessus donnent des résultats sur des bases comprenant un nombre relativement faible d'images. Quelques travaux de recherche proposent des résultats sur des bases de données de taille importante. Ainsi [51], dans le cadre d'un projet industriel avec *Daimler-Benz*[®], présente une évaluation systématique en termes de détections correctes et fausses alarmes sur une grande base de données, comprenant plus de 20000 images. Plus récemment, un algorithme de pré-détection de panneaux de danger et d'interdiction a été évalué sur une base de 26000 images, soit 175 kilomètres de route [22]. Citons également les travaux de [12] portant sur un système de détection de panneaux de fin d'interdiction. L'évaluation a été effectuée sur une base de 10000 images contenant 37 panneaux de fin de limite de vitesse différents. Dans [5], les auteurs proposent d'évaluer de façon distincte la phase de détection et d'évaluation sur une base de 9500 paires d'images stéréoscopiques. Enfin [43] propose une méthodologie d'analyse quantitative (à partir de courbes Precision-Recall, variante des courbes COR) et qualitatives (à partir de critères ergonomiques notamment) des performances de systèmes de reconnaissance de panneaux de signalisation.

Les performances des différents algorithmes peuvent difficilement être comparées entre elles. En effet, les outils d'évaluation, les bases de test et les objectifs diffèrent selon les publications. Par ailleurs, les auteurs étudient en général la signalisation verticale de leurs pays et on constate que les caractéristiques des panneaux diffèrent selon le pays. Cependant, dans [6], les performances de trois algorithmes de détection de panneaux sont comparées sur une même base d'images de scènes routières urbaines.

2.4 Systèmes opérationnels

D'une manière générale, les projets industriels sur la problématique des aides à la conduite sont peu nombreux (ou peu divulgués). Notons néanmoins les travaux de [51] au sein du consortium automobile allemand *Daimler-Benz*[®] portant sur les panneaux circulaires et triangulaires. Depuis 2008, deux autres constructeurs (*Opel*[®] et *BMW*[®]) commercialisent, sur un de leur véhicule, des systèmes de détection et reconnaissance de panneaux par vision. Ces systèmes se présentent sous forme d'alerte en temps réel au conducteur. Celle-ci porte uniquement sur les panneaux de limite de vitesse et d'interdiction de dépasser. Les méthodes sous-jacentes et les limites de ces systèmes restent confidentielles.

Les autres systèmes opérationnels concernent le relevé automatique de la signalisation verticale pour la gestion du patrimoine. Dans ce cas, la contrainte temps réel n'est plus indispensable. Les outils de détection et de reconnaissance sont intégrés dans la chaîne complète de traitement qui comprend l'acquisition des données, et l'analyse de l'image. Dans plusieurs cas, l'acquisition est effectuée par des véhicules équipés de caméras dans le visible et de capteur

LASER. Pour la société *Viametris*^{® 2}, la détection est réalisée à partir de l'image [45]. À l'inverse, la société *Geo3D*^{® 3} effectue une détection manuelle ou automatisée à partir de données LASER précédant une étape d'identification par analyse d'images, par comparaison du candidat détecté avec une base de modèles [31]. Les données LASER donnent également accès à une mesure de rétro-réflexion des panneaux. Dans le projet *RSR*⁴, l'équipe de l'*Institut National d'Optique (INO) du Canada* a développé un système à partir de caméras infrarouge exploitant les propriétés de rétro-réflexion des panneaux lors de la phase de détection [10]. Une approche similaire est utilisée dans le projet *VISUALISE* (*VISUAL* Inspection of Signs and panEls), porté par les sociétés Euroconsult^{® 5} et 3M-Spain⁶ et l'université d'Alcala (Espagne). Dans ce projet, on remarque que les relevés sont effectués de nuit.

La problématique est abordée de façon différente dans le projet *Mobvis*^{® 7}. En effet, l'acquisition de la signalisation verticale est effectuée par un PDA équipée d'un appareil photo et d'un GPS. L'image obtenue est ainsi fortement focalisée sur le panneau. L'automatisation du relevé consiste uniquement à identifier le panneau forcément présent dans l'image [9].

2.5 Conclusion

L'analyse de la signalisation verticale de police constitue un problème relativement classique de détection et reconnaissance des formes. Ce domaine, fortement relié à l'analyse de données, à la vision par ordinateur et à l'intelligence artificielle, se caractérise par une grande variété de méthodes. On retrouve donc naturellement cette variété dans la littérature spécialisée sur l'analyse de la signalisation. On peut néanmoins identifier deux tendances.

La première est celle des chaînes de traitement *bottom-up*, fondées sur une succession d'étapes : extraction rapide de primitives, puis filtrage des détections selon des critères géométriques ou photométriques, abstraction de la représentation et analyse des formes extraites. Les progrès récents de la qualité des images numériques et l'évolution de l'informatique ont permis de réelles avancées en termes de performances et remis au goût du jour certaines méthodes robustes, mais coûteuses. Ainsi, les techniques de vote ont connu un regain au début de la dernière décennie, notamment autour de la Transformée en Symétrie Radiale et de sa version adaptée aux polygones.

La seconde tendance a suivi les progrès des méthodes de détection et reconnaissance des formes. Non sans un certain effet de mode, des techniques telles que les réseaux de neurones, les modèles déformables, les modèles d'apparence, les Machines à Vecteurs Support (SVM) ou les méta-classifieurs de type Ada-Boost ont successivement été employées. Cette évolution des méthodes s'est également accompagnée d'un accroissement de la qualité et des performances.

Les techniques que nous proposons dans ce document sont représentatives de ces deux tendances, et nous semblent être au niveau de l'état de l'art actuel dans le domaine. On peut néanmoins nuancer cette affirmation en soulignant que certains systèmes commerciaux ne sont pas

²viametris.fr

³www.geo-3D.com

⁴www.ino.ca/en-ca/achievements/description/project-p/road-sign-detection-recognition.html

⁵www.euroconsult.es

⁶www.3m.com

⁷www.mobvis.org

2.5. CONCLUSION

documentés et que les évaluations dont ils ont fait l'objet ne sont pas publiées. Toutefois, il est peu probable que les techniques employées diffèrent de beaucoup de celles proposées dans la littérature.

La question de l'évaluation est centrale dans la mise au point des systèmes de détection et reconnaissance des formes. De ce point de vue, on ne peut que regretter l'absence de bases de données publiques de référence permettant une évaluation comparée des performances des différents systèmes. Il n'existe d'ailleurs pas, à notre connaissance, d'étude publiée proposant une comparaison quantitative des performances d'un ensemble représentatif de méthodes existantes. La comparaison de trois méthodes de détection proposée dans le présent ouvrage, sur une base de données réelle qui sera prochainement mise en ligne constitue un premier pas dans cette direction.

Bibliographie

- [1] R. ACH, N. LUTH, et A. TECHMER. Real-time detection of traffic signs on a multi-core processor. Dans *Proceedings of IEEE Intelligent Vehicles Symposium (IV'08)*, pages 307–312, Eindhoven, Pays-bas, juin 2008.
- [2] Y. AOYAGI et T. ASAKURA. A study on traffic sign recognition in scene image using genetic algorithms and neural networks. Dans *Proceedings of IEEE international Conference on Industrial Electronics, Control and Instrumentation*, volume 3, pages 1838–1843, Taipei, Taiwan, août 1996.
- [3] C. BAHLMANN, Y. ZHU, R. VISVANATHAN, M. PELLKOFER, et T. KOEHLER. A system for traffic sign detection, tracking, recognition using color, shape, and motion information. Dans *Proceedings of IEEE Intelligent Vehicles Symposium (IV'05)*, pages 255–260, Las Vegas, USA, juin 2005.
- [4] N. BARNES, A. ZELINSKY, et L.S. FLETCHER. Real-time speed sign detection using the radial symmetry detector. *IEEE Transactions on Intelligent Transportation Systems*, 9(2) :322–332, juin 2008.
- [5] X. BARO, S. ESCALERA, J. VITRIA, O. PUJOL, et P. RADEVA. Traffic sign recognition using evolutionary adaboost detection and forest-ECOC classification. *IEEE Transactions on Intelligent Transportation Systems*, 10 :113–126, 2009.
- [6] R. BELAROUSSI, P. FOUCHER, J.-P. TAREL, B. SOHEILIAN, P. CHARBONNIER, et N. PAPARODITIS. Road sign detection in images : a case study. Dans *Proceeding of the 20th International conference on Pattern Recognition (ICPR 2010)*, Istanbul, Turquie, août 2010. À paraître.
- [7] R. BELAROUSSI et J.-P. TAREL. Angle vertex and bisector geometric model for triangular road sign detection. Dans *Proceedings of IEEE Workshop on Applications of Computer Vision (WACV'09)*, pages 577–583, Snowbird, USA, décembre 2009.
- [8] R. BELAROUSSI et J.-P. TAREL. A real-time road sign detection using bilateral chinese transform. Dans *Proceedings of IEEE International Symposium on Visual Computing (ISVC'09)*, pages 1161–1170, Las Vegas, USA, novembre 2009.
- [9] W. BENESOVA, Y. LYPETSKYY, J.-P. ANDREU, L. PALETTA, A. JEITLER, et E. HÖDL. A mobile system for vision based road sign inventory. Dans *Proceedings of 5th International Symposium on Mobile Mapping Technology (MMT 07)*, Padoue, Italie, mai 2007.
- [10] D. BOULANGER. Système RSR : Système automatisé d'inventaire des panneaux routiers. en ligne, Institut National d'Optique (INO), Ste-Foy, Quebec, Canada, 2006.
- [11] A. BROGGI, P. CERRI, P. MEDICI, P. P. PORTA, et G. GHISIO. Real time road signs recognition. Dans *Proceedings of IEEE Intelligent Vehicles Symposium (IV'07)*, pages 981–986, Istanbul, Turquie, juin 2007.
- [12] C. CARAFFI, E. CARDARELLI, P. MEDICI, P. P. PORTA, G. GHISIO, et G. MONCHIERO. An algorithm for italian de-restriction signs detection. Dans *Proceedings of IEEE Intelligent Vehicles Symposium (IV'08)*, pages 834–839, Eindhoven, Pays-bas, juin 2008.
- [13] S.-Y. CHEN et J.-W. HSIEH. Boosted road sign detection and recognition. Dans *Proceedings of IEEE International conference on Machine Learning and Cybernetics (ICMLC 2008)*, volume 7, pages 3823–3826, Kunming, Chine, juillet 2008.

BIBLIOGRAPHIE

- [14] R. DAHYOT. *Analyse d'images séquentielles de scènes routières par modèles d'apparence pour la gestion du réseau routier*. Thèse de doctorat, Université Louis Pasteur, Strasbourg I, 2001. Collections Etudes et Recherches des Laboratoires des Ponts et Chaussées, CR 31, LCPC.
- [15] A. de la ESCALERA, J. M. ARMINGOL, et M. MATA. Traffic sign recognition and analysis for intelligent vehicles. *Image and Vision Computing*, 21(3) :247–258, mars 2003.
- [16] A. de la ESCALERA, J. M. ARMINGOL, F. RODRIGUEZ, et J. M. PASTOR. Visual sign information extraction and identification by deformable models by intelligent vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 5(2) :57–68, 2004.
- [17] A. de la ESCALERA, L. E. MORENO, J. M. ARMINGOL, et M. MATA. Road traffic sign detection and classification. *IEEE Transactions on Industrial Electronics*, 44(6) :848–859, 1997.
- [18] G. DUTILLEUX et P. CHARBONNIER. Détection de signalisation routière par ajustement de formes prototypes. Dans *20ème colloque GRETSI sur le traitement du signal et des images*, pages 1109–1112, Louvain-la-Neuve (Belgique), septembre 2005.
- [19] G. DUTILLEUX et P. CHARBONNIER. Métaheuristiques biologiques pour la détection de la signalisation routière. Dans P. SIARRY, éditeur, *Optimisation en traitement du signal et de l'image*, Traité IC2, série traitement du signal et de l'image, chapitre 10, pages 271–294. Hermes, février 2007.
- [20] A. A. FARAG et A. E. ABDEL-HAKIM. Detection, categorization and recognition of road signs for autonomous navigation. Dans *Proceeding of Advanced Concepts for Intelligent Vision System*, pages 125–130, Bruxelles, Belgique, août 2004.
- [21] T. FAWCETT. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(9) :861–874, 2006.
- [22] P. FOUCHER, P. CHARBONNIER, et H. KEBBOUS. Evaluation of a road sign pre-detection system by image analysis. Dans *Proceedings of International Conference on Computer Vision theory and Applications (VISAPP 2009)*, pages 362–367, Lisbonne, Portugal, février 2009.
- [23] M. A. GARCIA-GARRIDO, M. A. SOTELO, et E. MARTIN-GOROSTIZA. Fast traffic sign detection and recognition under changing lighting conditions. Dans *Proceedings of IEEE Intelligent Transportation Systems Conference (ITSC'06)*, pages 811–816, Toronto, Canada, septembre 2006.
- [24] D. M. GAVRILA. Traffic sign recognition revisited. Dans *Proceedings of the 21st DAGM Symposium für Mustererkennung*, pages 86–93, Bonn, Allemagne, septembre 1999.
- [25] D. M. GAVRILA et V. PHILOMIN. Real-time object detection for smart vehicles. Dans *Proceedings of IEEE International Conference on Computer Vision (ICCV'99)*, pages 87–93, Kerkyra, Grèce, septembre 1999.
- [26] P. GIL-JIMENEZ, S. LAFUENTE-ARROYO, H. GOMEZ-MORENO, F. LOPEZ-FERRERAS, et S. MALDONADO-BASCON. Traffic sign shape classification evaluation II : FFT applied to the signature of blobs. Dans *Proceedings of IEEE Intelligent Vehicles Symposium (IV'05)*, pages 607–612, Las Vegas, USA, juin 2005.

- [27] A. F. HABIB et M. N. JHA. Hypothesis generation of instances of road signs in color imagery captured by mobile mapping systems. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 36 part 5/C55, pages 159–165, 2007.
- [28] Y. ISHIZUKA et Y. HIRAI. Segmentation of road sign symbols using opponent-color filters. Dans *Proceedings of 11th World Congress on Intelligent Transport Systems and Services (ITSWC)*, Nagoya, Japon, octobre 2004.
- [29] C. G. KELLER, C. SPRUNK, K. BHALMANN, J. GIEBEL, et G. BARATOFF. Real-time recognition of U.S. speed signs. Dans *Proceedings of IEEE Intelligent Vehicles Symposium (IV'08)*, pages 518–523, Eindhoven, Pays-Bas, juin 2008.
- [30] J. F. KHAN, S. M. A. BHUIYAN, et R. R. ADHAMI. Distortion invariant road sign detection. Dans *Proceedings of IEEE International Conference on Image Processing (ICIP 2009)*, pages 841–844, Le Caire, Égypte, novembre 2009.
- [31] T. KINGSTON et C. LAFLAMME. Advances in management : Automated road sign detection and recognition. *International Municipal Signal Association Journal*, 45(1) :46–49, 2007.
- [32] C. G. KIRAN, V. P. LEKHESH, V. Abdu RAHIMAN, et K. RAJEEV. Traffic sign detection and pattern recognition using support vector machine. Dans *Proceedings of seventh International Conference on Advances in Pattern Recognition (ICAPR'09)*, pages 87–90, Kolkata, Inde, février 2009.
- [33] S. LAFUENTE-ARROYO, P. GIL-JIMENEZ, R. MALDONADO-BASCON, F. LOPEZ-FERRERAS, et S. MALDONADO-BASCON. Traffic sign shape classification evaluation I : SVM using distance to borders. Dans *Proceedings of IEEE Intelligent Vehicles Symposium (IV'05)*, pages 557–562, Las Vegas, USA, juin 2005.
- [34] M. LALONDE et Y. LI. Road sign recognition - Survey of the state of art for sub-project 2.4. Technical report, Centre de Recherche en Informatique de Montréal, 1995. Collection scientifique et technique, CRIM-IIT-95/09-35.
- [35] LE MINISTRE DE L'INTÉRIEUR et LE MINISTRE DE L'ÉQUIPEMENT ET DE L'AMÉNAGEMENT DU TERRITOIRE. Instruction interministérielle sur la signalisation routière, parties 1 à 5 et partie 8, Arrêté du 7 juin 1977. Dernière mise à jour 2009.
- [36] G. LOY et N. BARNES. Fast shape-based road sign detection for a driver assistance system. Dans *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'04)*, volume 1, pages 70–75, Sendai, Japon, octobre 2004.
- [37] G. LOY et A. ZELINSKY. Fast radial symmetry for detecting points of interest. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8) :959–973, août 2003.
- [38] S. MAJI et J. MALIK. Object detection using a max-margin hough transform. Dans *Proceedings of Computer Vision and Pattern Recognition (CVPR 2009)*, pages 1038–1045, Miami, USA, juin 2009.
- [39] S. MALDONADO-BASCON, S. LAFUENTE-ARROYO, P. SIEGMANN, H. GOMEZ-MORENO, et F. J. ACEVEDO-RODRIGUEZ. Traffic sign recognition system for inventory purposes. Dans *Proceedings of IEEE Intelligent Vehicles Symposium (IV'08)*, pages 590–595, Eindhoven, Pays-Bas, juin 2008.

BIBLIOGRAPHIE

- [40] R. MALIK, J. KHURSHID, et S. N. AHMAD. Road sign detection and recognition using colour segmentation, shape analysis and template matching. Dans *Proceedings of the 6th International Conference on Machine Learning and Cybernetics, 2007*, volume 6, pages 3556–3560, Hong Kong, août 2007.
- [41] J. MIURA, T. KANDA, et Y. SHIRAI. An active vision system for real-time traffic sign recognition. Dans *Proceedings of IEEE Intelligent Transportation System (ITSC'00)*, pages 52–57, Dearborn, USA, octobre 2000.
- [42] F. MOUTARDE, A. BARGETON, A. HERBIN, et L. CHANUSSOT. Robust on-vehicle real-time visual detection of american and european speed limit signs, with a modular traffic signs recognition system. Dans *Proceedings of IEEE Intelligent Vehicles Symposium 2007 (IV'07)*, Istanbul, Turquie, juin 2007.
- [43] S. MÜLLER-SCHNEIDERS, C. NUNN, et M. MEUTER. Performance evaluation of a real-time traffic sign recognition system. Dans *Proceedings of IEEE Intelligent Vehicles Symposium (IV'08)*, pages 79–84, Eindhoven, Pays-Bas, juin 2008.
- [44] Y. Y. NGUWI et A. Z. KOUZANI. Detection and classification of road signs in natural environments. *Neural Computing and Applications*, 17 :265–289, 2008.
- [45] J. NINOT, L. SMADJA, et K. HEGGARTY. Road sign recognition using a hybrid evolutionary algorithm and primitives fusion. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 39(3), 2010. À paraître.
- [46] N. OTSU. A threshold selection method from gray-level histogram. *IEEE Transactions on System Man and Cybernetics*, 9 :62–66, 1979.
- [47] P. PACLIK, J. NOVOTICOVA, P. SOMOL, et P. PUDIL. Road sign classification using laplace kernel classifier. *Pattern Recognition Letters*, 21(13-14) :1165–1173, 2000.
- [48] C. F. PAULO et P. L. CORREIA. Automatic detection and classification of traffic signs. Dans *Proceedings of 8th International Workshop on Image analysis for Multimedia Interactive Services (WIAMIS'07)*, pages 11–14, Santorini, Grèce, juin 2007.
- [49] G. PICCIOLI, E. De MICHELI, P. PARODI, et M. CAMPANI. A robust method for road sign detection and recognition. *Image and Vision Computing Journal*, 14 :209–223, 1996.
- [50] L. PRIESE, R. LAKMANN, et V. REHRMANN. Ideogramm identification in a realtime traffic sign recognition system. Dans *Proceeding of IEEE intelligent Vehicles Symposium (IV'95)*, pages 310–314, Nagoya, Japon, septembre 1995.
- [51] L. PRIESE, V. REHRMANN, R. SCHIAN, et R. LAKMANN. Traffic sign recognition based on color image evaluation. Dans *Proceedings of IEEE Intelligent Vehicles Symposium (IV'93)*, pages 95–100, Tokyo, Japon, juillet 1993.
- [52] P. L. ROSIN. Measuring shape : ellipticity, rectangularity, and triangularity. *Machine Vision and Applications*, 14(3) :172–184, juillet 2003.
- [53] A. RUTA, Y. LI, et X. LIU. Real-time traffic sign recognition from video by class-specific discriminative features. *Pattern Recognition*, 43 :416–430, 2009.
- [54] D. G. SHAPOSHNIKOV, N. LUBOV, L. N. PODLADCHIKOVA, A. V. GOLOVAN, N.A. SHEVTSOVA, K. HONG, et X.W. GAO. Road sign recognition by single positioning of space-variant sensor window. Dans *Proceedings of 15th International Conference on Vision Interface*, pages 213–217, Calgary, Canada, mai 2002.

- [55] L. SIMON, J.-P. TAREL, et R. BRÉMOND. Towards the estimation of conspicuity with visual priors. Dans *Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP 2008)*, pages 323–328, Funchal, Portugal, 2008.
- [56] G. K. SIOGKAS et E. S. DERMATAS. Detection, tracking and classification of road signs in adverse conditions. Dans *Proceedings of Mediterranean Electrotechnical Conference (MELECON'06)*, pages 537–540, Benalmadena, Espagne, mai 2006.
- [57] B. SOHEILIAN, A. ARLICOT, et N. PAPARODITIS. Extraction de panneaux de signalisation routière dans les images couleurs. Dans *Actes du 17ème congrès francophone en Reconnaissance de Formes et Intelligence Artificielle, RFIA'10*, pages 743–750, Caen, France, janvier 2010.
- [58] J. TORRESEN, J. W. BAKKE, et L. SEKANINA. Efficient recognition of speed limit signs. Dans *Proceedings of 7th IEEE Conference on Intelligent Transportation Systems*, pages 652–656, Washington, USA, octobre 2004.
- [59] R. A. VÁZQUEZ REINA, R. J. LÓPEZ SASTRE, S. LAFUENTE ARROYO, et P. GIL JIMÉNEZ. Adaptive traffic road sign panels text extraction. Dans *Proceedings of the 5th WSEAS International Conference on Signal Processing, Robotics and Automation (IS-PRA'06)*, pages 295–300, février 2006.
- [60] P. VIOLA et M. JONES. Robust real-time face detection. *International Journal of computer Vision*, 57 :137–154, 2004.
- [61] S. VITABILE, G. POLLACIA, G. PILATO, et F. SORBELLO. Road signs recognition using a dynamical pixel aggregation technique in the HSV color space. Dans *Proceedings of 11th International conference On Image Analysis and Processing (ICIAP 2001)*, pages 572–577, Palerme, Italie, septembre 2001.
- [62] Y.J. ZHENG, W. RITTER, et R. JANSSEN. An adaptive system for traffic sign recognition. Dans *Proceedings of IEEE Intelligent Vehicles Symposium (IV'94)*, pages 165–170, Paris, France, octobre 1994.

Chapitre 3

Détection par la couleur et par transformations univariées

*Philippe Foucher,
Pierre Charbonnier*

Ce chapitre présente la méthode de détection de la signalisation verticale développée au LRPC de Strasbourg (ERA 27) en se focalisant uniquement sur les panneaux de danger, d'indication et de prescription. L'algorithme se base sur la transformée à symétrie radiale (pour les cercles) ou la transformée de Hough (pour les polygones). Appliqués directement sur des images de grande taille (1280×1024 ou 1920×1080), les algorithmes de vote univariés impliquent des temps de calcul longs et un risque de taux de fausses alarmes élevé, notamment en milieu urbain, où le nombre de distracteurs est sensiblement plus élevé qu'en milieu rural. Pour diminuer les temps de calcul et le taux de fausses alarmes, un algorithme en trois étapes est proposé :

1. Dans un premier temps, les pixels de l'image (R,V,B) sont filtrés selon leurs couleurs. Pour les panneaux de police, trois couleurs sont principalement utilisées : rouge, bleu et blanc. Nous nous limiterons pour le moment au cas des panneaux à couronne rouge (à l'exception des panneaux « STOP » et « cédez le passage ») ou à fond bleu.
2. Une sélection sur les composantes connexes résultantes du filtre colorimétrique est ensuite effectuée selon des caractéristiques élémentaires de forme et de position dans l'image.
3. La sélection finale est réalisée en utilisant un critère de symétrie (pour les panneaux circulaire) à partir de la transformée en symétrie radiale [6] ou en utilisant des critères d'angles entre les lignes principales obtenues par la transformée de Hough (pour les panneaux triangulaires et rectangulaires).

Chaque étape nécessite d'optimiser les valeurs d'un ou de plusieurs paramètres. Ces valeurs sont déterminées de façon empirique sur des séquences d'images de scènes routières contenant un nombre significatif de panneaux dans des situations parfois complexes. La section 3.1 décrit les séquences d'images utilisées pour d'une part optimiser les valeurs des paramètres et d'autre part valider les résultats. Les différentes étapes de l'algorithme sont détaillées dans la section 3.2. Enfin, les résultats expérimentaux sont présentés et discutés dans la section 3.3.

3.1 Bases de données

3.1.1 Séquences d'images

La base de test est composée de 5 séquences d'images de scènes routières acquises par des caméras frontales montées sur des véhicules d'inspection en circulation. Les caractéristiques des séquences de test sont données dans le tableau 3.1. Les images sont acquises, de jour, tous les 5 ou 10 mètres par différents types de caméras CCD et avec des conditions d'illuminations différentes. Au total, la base contient 26285 images correspondant à 175 kilomètres de routes rurales, urbaines ou 2×2 voies.

| Séquences | Nombre d'images | Pas | type de caméra | Taille en pixels | Taille en Ko | Type de séquence |
|-----------|-----------------|------|----------------|--------------------|--------------|-----------------------------|
| S1 | 3436 | 5 m | Basler A101 | 1280×1024 | 156 | rural et péri-urbain |
| S2 | 3436 | 5 m | Basler A101 | 1280×1024 | 156 | rural et péri-urbain |
| S3 | 10420 | 5 m | Marlin F145-C2 | 1280×1024 | 95,6 | rural et péri-urbain |
| S4 | 890 | 10 m | Pike F210-C | 1920×1080 | 209 | urbain |
| S5 | 8103 | 10 m | Pike F210-C | 1920×1080 | 223 | rural et 2×2 voies |

TAB. 3.1 – Caractéristiques des séquences d'images de scènes routières.

3.1.2 Vérité-terrain

La vérité-terrain a été établie en relevant manuellement les panneaux dans chaque image. Pour effectuer ce relevé, les opérateurs ont utilisé le logiciel SAFOR¹ qui facilite la navigation dans les séquences et le relevé des objets sélectionnés. Plusieurs champs sont enregistrés : numéro de l'image, position dans l'image, type de panneau, rotation, distorsion, occultation éventuelle... Un panneau est relevé si sa boîte englobante a une taille minimale de 25 pixels dans les deux dimensions. Le tableau 3.2 comptabilise le nombre de panneaux relevés par séquence et selon 4 catégories [5] :

- Dans la catégorie « danger », on regroupe tous les panneaux de type A - *cf.* Fig. 3.1(a), les panneaux d'intersection AB1, AB2 et AB25 et les panneaux temporaires de type AK - *cf.* Fig. 3.1(b).
- Dans la catégorie « interdiction », on retrouve tous les panneaux de type B0 à B19 - *cf.* Fig. 3.1(c). On peut remarquer que le panneau « sens interdit » se distingue des autres en termes de géométrie - *cf.* Fig. 3.1(d). Cette particularité sera prise en compte dans l'algorithme.

¹logiciel (SAisie de FORMes) développé par l'ERA 27 du LRPC de Strasbourg en collaboration avec la Division Métrologie et Instrumentation et le LEPSiS, du LCPC

3.1. BASES DE DONNÉES

- Dans la catégorie « obligation », on retrouve tous les panneaux de type B21 à B29 - *cf.* Fig. 3.1(e).
- Dans la catégorie « indication », on retrouve tous les panneaux de type C - *cf.* Fig. 3.1(f). La balise « J5 » - *cf.* Fig. 3.1(g) présente des caractéristiques géométriques très proches et sera ajouté à cette catégorie.

Par ailleurs, un même panneau est apparent dans des images successives, on distingue donc dans la base de référence le nombre de panneaux différents et le nombre total d'apparitions de panneaux dans les images.

| Séq. | Nombre de panneaux différents | | | | Nombre d'apparition des panneaux | | | |
|-------|-------------------------------|----------|------------|------------|----------------------------------|----------|------------|------------|
| | danger | interdit | obligation | indication | danger | interdit | obligation | indication |
| S1 | 26 | 22 | 5 | 18 | 129 | 111 | 19 | 91 |
| S2 | 12 | 6 | 0 | 9 | 81 | 45 | 0 | 64 |
| S3 | 70 | 91 | 12 | 40 | 405 | 497 | 59 | 190 |
| S4 | 9 | 29 | 18 | 27 | 36 | 80 | 48 | 61 |
| S5 | 70 | 115 | 9 | 76 | 385 | 593 | 43 | 380 |
| Total | 187 | 263 | 42 | 170 | 1036 | 1326 | 169 | 786 |

TAB. 3.2 – Vérité-terrain.

Dans le tableau 3.2, on recense un total de 662 panneaux différents qui sont visibles 3317 fois. On peut noter que le nombre d'apparitions des panneaux dans les images est beaucoup plus faible dans une zone urbaine que dans les autres zones. Pour les panneaux d'interdiction par exemple, 29 panneaux différents apparaissent dans 80 images pour la séquence S4 (en moyenne, 1 même panneau est donc visible sur 2,75 images) alors que 22 panneaux d'interdiction différents apparaissent dans 111 images pour la séquence S1 (en moyenne, 1 même panneau est visible sur 5 images).

3.1.3 Métriques d'évaluation

Les performances de l'algorithme sont analysées à partir de courbes COR sur lesquelles le taux de détections correctes (TDC) est représenté en ordonnée et le taux de fausses alarmes (TFA) est représenté en abscisse. Le taux de fausses alarmes est défini de la façon suivante :

$$\text{TFA} = \frac{\text{FA}}{\text{NbImg}}$$

où FA est le nombre d'objets détectés à tort comme panneau et *NbImg* le nombre d'images de la séquence.

Nous distinguons deux types de taux de détections correctes suivant que l'on considère le nombre de panneaux détectés au moins une fois dans les images sur lesquels il apparaît (TDC_p) ou que l'on considère le nombre total de détections de panneaux (TDC_i). Dans ce cas, on comptabilise la détection d'un panneau par rapport à chaque image dans laquelle il apparaît. Ces taux sont définis selon :

$$\text{TDC}_i = \frac{\text{VP}_i}{\text{P}_i} \quad \text{TDC}_p = \frac{\text{VP}_p}{\text{P}_p}$$

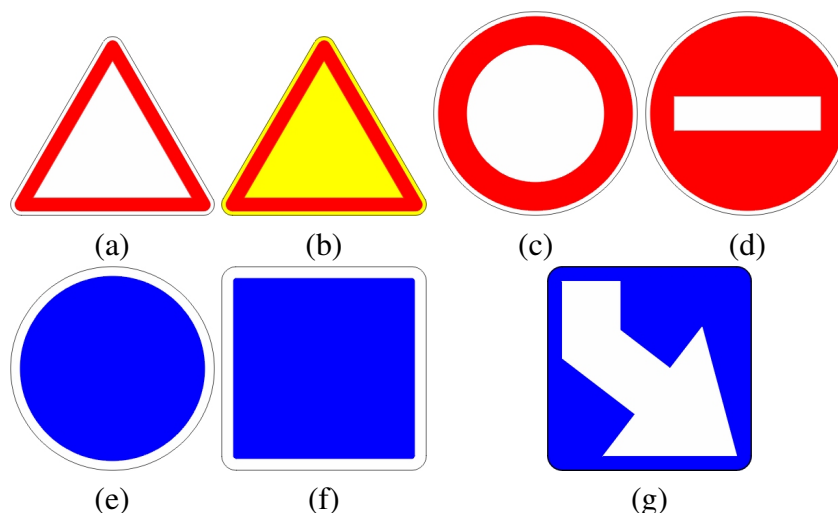


FIG. 3.1 – Ensemble des panneaux répertoriés dans la vérité-terrain : (a) panneau de danger de type A, ainsi que AB1, AB2 et AB25 ; (b) panneau temporaire de type AK ; (c) panneau d'interdiction de type B0 à B19 ; (d) panneau B1 : « sens interdit » ; (e) panneau d'obligation de type B21 à B29 ; (f) panneau d'indication de type C ; (g) balise J5.

où VP_i est le nombre total de détections de panneaux, P_i est le nombre réel d'apparitions des panneaux, VP_p est le nombre de panneaux différents, détectés au moins une fois, P_p est le nombre réel de panneaux différents.

Les règles de détection sont les mêmes que celles données dans la partie 6.2.2.



FIG. 3.2 – (à gauche) Exemple de scène routière ; (au milieu) classification des pixels rouge/non rouge ; (à droite) classification des pixels bleu/non bleu (à droite). La région d'intérêt correspond à la zone non grisée.

3.2 Description de l'algorithme

3.2.1 Filtre colorimétrique

La première étape consiste à construire à partir de l'image en couleurs deux images binaires correspondant à une classification individuelle des pixels selon un critère colorimétrique : rouge /non rouge ou bleu/non bleu -cf. Fig. 3.2.

3.2. DESCRIPTION DE L'ALGORITHME

Pour les panneaux à couronne rouge (danger et interdiction), on sélectionne les pixels de l'image en couleurs (R,V,B) qui répondent aux critères suivants [2, 3] :

$$\begin{aligned} R &> \alpha_r(V + B) \\ R - \max(V, B) &> \beta_r[\max(V, B) - \min(V, B)] \end{aligned} \quad (3.1)$$

La première expression permet de sélectionner les pixels avec une composante rouge normalisée $R/(R + V + B)$ dominante. La normalisation rend le filtre robuste aux variations lumineuses. La seconde expression permet d'éliminer les pixels qui tendraient vers la couleur jaune ou magenta. Pour les panneaux à fond bleu (obligation et indication), le filtre utilisé permet de sélectionner uniquement les pixels dont la composante bleue normalisée est majoritaire selon l'expression :

$$B > \alpha_b(V + R) \quad (3.2)$$

Contrairement aux panneaux rouges, il n'est pas possible d'effectuer une sélection plus étroite car les panneaux bleus observés présentent une gamme de couleur assez étendue, variant du « bleu-cyan » au « bleu-magenta ».

3.2.2 Utilisation des informations géométriques

Les composantes connexes issues de la sélection colorimétrique sont analysées en utilisant les critères suivants :

1. une première sélection est effectuée selon les dimensions de la boîte englobante pour éliminer les objets trop petits ou trop grands,
2. une deuxième sélection prend en compte la position de la composante connexe dans l'image
3. une troisième sélection se base sur la couverture de la composante connexe, définie par le rapport entre l'aire de la composante connexe et l'aire de sa boîte englobante. Sa valeur est comprise entre 0 et 1. La couverture est seuillée.

3.2.3 Sélection des cercles par symétrie radiale

Transformée en symétrie radiale

La Transformée en Symétrie Radiale (TSR) proposée par [6], crée une carte d'accumulation de votes issus de chaque pixel de l'image.

Le calcul de cette transformée s'effectue de la façon suivante. Chaque pixel P d'une image monochrome vote pour un pixel P^+ et P^- correspondant aux points situés à une distance r du pixel P , dans la direction du gradient $g(P)$, dans le sens positif pour P^+ et opposé pour P^- (- cf. Fig. 3.3). Les coordonnées des pixels P^+ et P^- sont données par l'expression :

$$\begin{aligned} P^+(P) &= P + \text{round}\left(\frac{g(P)}{\|g(P)\|} r\right) \\ P^-(P) &= P - \text{round}\left(\frac{g(P)}{\|g(P)\|} r\right) \end{aligned} \quad (3.3)$$

Les gradients sont calculés par l'opérateur Sobel de taille 3×3 .

Le rayon r peut varier dans un intervalle pré-défini $[r_{min}, r_{max}]$.

3. TRANSFORMATIONS UNIVARIÉES

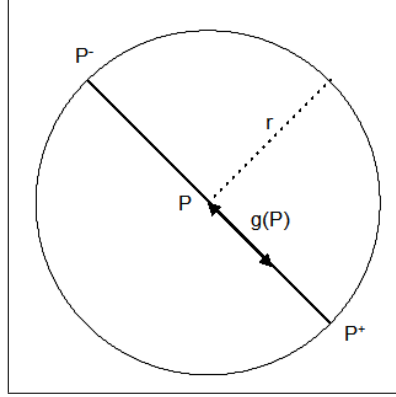


FIG. 3.3 – système de vote [6].

Pour chaque rayon r , une carte d'orientation O_r et d'amplitude M_r est construite de la manière suivante :

$$\begin{aligned} O_r(P^\pm(P)) &= O_r(P^\pm(P)) \pm 1 \\ M_r(P^\pm(P)) &= M_r(P^\pm(P)) \pm \|g(P)\| \end{aligned} \quad (3.4)$$

Pour ne pas favoriser les valeurs des votes calculées sur les grands rayons, il est nécessaire de normaliser toutes les valeurs en les divisant par la valeur maximale attendue k_r pour la carte O_r . La valeur k_r est déterminée de manière empirique et correspond à :

$$k_r = \begin{cases} 8 & \text{si } r = 1 \\ 9,9 & \text{sinon} \end{cases} \quad (3.5)$$

La matrice F_r est calculée selon l'expression :

$$F_r(P) = \frac{M_r(P)}{k_r} \times \left(\frac{\|\tilde{O}_r(P)\|}{k_r} \right)^\alpha \quad (3.6)$$

où $\tilde{O}_r(P)$ est la carte d'orientation saturée :

$$\tilde{O}_r(P) = \begin{cases} O_r(P) & \text{si } O_r(P) < k_r \\ k_r & \text{sinon} \end{cases} \quad (3.7)$$

et α est un facteur égal à 1, 2 ou 3 qui permet d'introduire une certaine souplesse sur la symétrie radiale recherchée. Plus α est élevé, plus on recherche des caractéristiques strictement radiales. La valeur $\alpha = 2$ est un bon compromis et est conservée par la suite.

En réalité, le voisinage de chaque pixel sur lequel on vote est pris en compte en convoluant F_r par une gaussienne bidimensionnelle A_r de taille $r \times r$ et d'écart-type $\sigma = 0,25$:

$$S_r = F_r * A_r \quad (3.8)$$

L'image d'accumulation finale S prend en compte les transformées calculées pour chaque rayon r :

$$S = \frac{1}{N} \times \sum_{r=r_{min}}^{r_{max}} S_r \quad (3.9)$$

N est le nombre de rayons sur lequel est calculée la transformée.

3.2. DESCRIPTION DE L'ALGORITHME

Choix d'une composante colorimétrique

La *TSR* est appliquée ici directement aux images correspondantes aux boîtes englobantes de chaque composante connexe sélectionnée. Les dimensions (h pour la hauteur et w pour la largeur) sont ainsi de taille réduite : h et w sont inférieures à 200.

L'intervalle $[r_{min}, r_{max}]$ est défini directement à partir des dimensions de l'image :

$$\begin{aligned} r_{min} &= \max(h, w)/8 \\ r_{max} &= \max(h, w)/2 + 1 \end{aligned} \quad (3.10)$$

La *TSR* ne s'effectue que sur des images à une seule composante. Dans [4], le choix de la composante s'est porté sur la luminance et les résultats ont montré l'efficacité de la symétrie, sauf dans le cas des panneaux de stationnement, en particulier usagés, pour lesquels le gradient entre la couronne rouge et le cercle intérieur bleu est assez faible. Pour prendre en compte ce cas spécifique sans détériorer *a priori* le taux de détections correctes sur les autres panneaux, une possibilité consiste à utiliser, en entrée du sélecteur de symétrie, les images de la composante bleue ou bleue normalisée.

Pour confirmer cette hypothèse, des tests ont été réalisés sur une séquence composée de 3436 images, dont 93 contiennent un panneau d'interdiction. Sur la figure 3.4, les courbes COR sont tracées pour les trois modalités (luminance, bleue, bleue normalisée) en faisant varier le seuil de symétrie sur un intervalle $[0, 70]$. De façon assez nette, l'utilisation de la composante bleue normalisée permet d'obtenir un meilleur taux de détections correctes. Ceci s'explique par une amélioration du taux de détections correctes des panneaux de stationnement.

Pour la suite, la modalité bleue normalisée sera conservée pour les panneaux à couronne rouge. Par déduction, la modalité rouge normalisée sera utilisée pour les panneaux à fond bleu.

3.2.4 Sélection des polygones par transformée de Hough

La transformée de Hough est classiquement utilisée pour identifier des courbes paramétriques simples (droites, cercles...). De la même manière que la transformée en symétrie radiale, la transformée de Hough est appliquée aux images correspondantes à la boîte englobante de l'objet sélectionné à la fin de l'étape 2. Les images binaires traitées sont en réalité les images des contours de la composante connexe. La transformée vise à identifier les principaux segments de droite et à contrôler si les angles entre ces segments de droite correspondent aux angles théoriques du triangle (60°) et du carré (90°) avec une marge d'erreur de $\pm 5^\circ$ - cf. Fig. 3.5.

La sélection des panneaux de danger (triangle) est réalisée uniquement sur les composantes connexes issues de la carte de rouge et la sélection des panneaux d'indication (carré) est effectuée sur les composantes connexes issues du filtre bleu. Pour ces derniers, le pictogramme à l'intérieur influe sur la détection des droites. Aussi, dans le cas des panneaux carrés, la transformée de Hough est en réalité appliquée sur les images de contour de l'objet dont les trous ont été préalablement remplis.

Pour la détection de panneaux, seuls les huit premiers segments de droite, répondant à une critère d'orientation, ont analysés. Pour les panneaux carrés, les segments de droite sélectionnés doivent être proche de l'horizontale ou de la verticale (à $\pm 10^\circ$ près). Pour les panneaux triangulaires, les segments doivent être orientés selon l'horizontale ou à $\pm 60^\circ$ de celle-ci (à $\pm 20^\circ$ près).

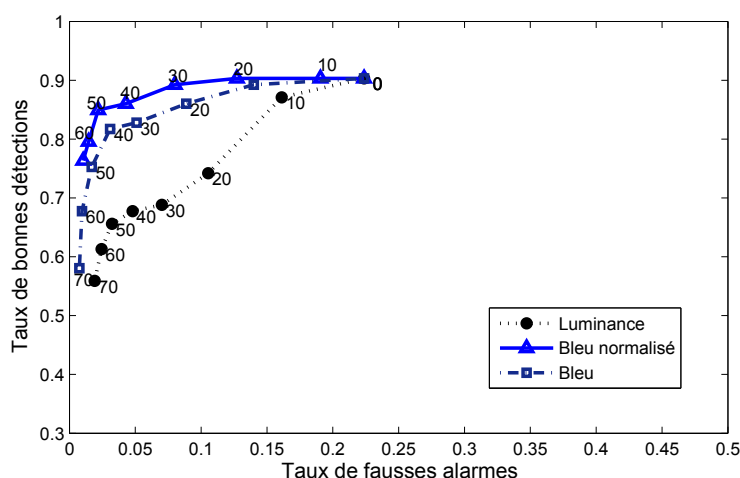


FIG. 3.4 – Influence de la composante colorimétrique pour la détection des panneaux d’interdiction par la TSR. Le seuil de symétrie varie dans l’intervalle $[0, 70]$. Le taux de fausses alarmes correspond aux nombres d’images contenant au moins une fausse alarme sur le nombre total d’images.

Une étude plus approfondie est prévue pour optimiser ce nombre de droites et les paramètres angulaires. Au final, un objet est sélectionné comme panneau si :

- pour un triangle, au moins un angle entre deux droites (parmi les 15 possibilités) est compris dans l’intervalle $[55^\circ, 65^\circ]$.
- pour un carré, au moins deux angles entre deux droites (parmi les 15 possibilités) est compris dans l’intervalle $[85^\circ, 95^\circ]$.

3.3 Résultats expérimentaux

Dans cette partie, les performances des algorithmes sont évaluées. Dans un premier temps, il est nécessaire d’optimiser les valeurs définies préalablement. Les seuils sont déduits des normes officielles sur les caractéristiques des panneaux de la manière suivante :

1. En ce qui concerne la taille de l’objet, les composantes connexes sont conservées si les dimensions de la boîte englobante de la composante connexe sont chacune supérieures à 20 pixels et inférieures à 200 pixels. La valeur minimale correspond au seuil de 25 pixels fixé dans la vérité-terrain, une petite marge est laissée pour prendre en compte le listel blanc autour du panneau, qui est considéré dans le relevé manuel de la vérité-terrain. Le seuil maximum correspond à la dimension maximale observée sur les panneaux, avec également une marge de sûreté.
2. Seules les composantes connexes dont le centre de gravité se trouve dans la zone non grisée de l’image sont conservées - cf. Fig. 3.2.

3.3. RÉSULTATS EXPÉRIMENTAUX

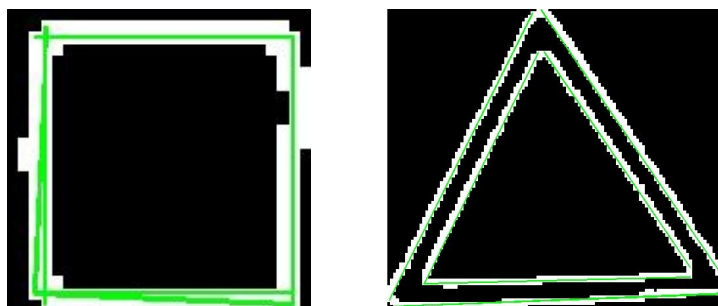


FIG. 3.5 – Détection des lignes principales (en vert) par la transformée de Hough.

3. Pour les panneaux de danger et d'interdiction, la composante connexe correspond à la couronne rouge du panneau, ce qui implique un objet relativement vide, un seuil maximum de couverture est donc défini. La sélection par la couleur bleue permet d'extraire des objets correspondant aux panneaux d'obligation et d'indication, et donne des composantes connexes relativement pleines. Dans ce cas, un seuil minimum de couverture est défini. L'ensemble des valeurs seuils est réuni dans le tableau 3.3. Pour les panneaux d'interdiction, le seuil maximum est volontairement élevé (0,65) car cette catégorie englobe les panneaux « sens interdit » pour lesquels la couverture de la composante dans la boîte englobante est plus importante.

| Type de panneau | Seuil minimum | Seuil maximum |
|---------------------------------------|---------------|---------------|
| Panneau de danger (Triangle rouge) | - | 0,35 |
| Panneau d'interdiction (Cercle rouge) | - | 0,65 |
| Panneau d'obligation (Cercle Bleu) | 0,45 | 1 |
| Panneau d'indication (Carré Bleu) | 0,45 | 1 |

TAB. 3.3 – Choix des seuils pour le critère de couverture.

Pour les autres coefficients, liés à la couleur et à la symétrie, les valeurs optimales sont déterminées empiriquement à partir de deux séquences d'images S1 et S4. L'algorithme est ensuite testé sur les séquences S2, S3 et S5 à partir du meilleur jeu de coefficients.

3.3.1 Réglages des paramètres couleur

Les valeurs des coefficients α_r et α_b ont été déterminées à partir de courbes COR sur les séquences S1 et S4. Pour le filtre rouge, la valeur β_r n'a pas été fixée empiriquement mais le critère $\beta_r = 2\alpha_r$ donne des résultats satisfaisants pour les deux séquences S1 et S4.

Pour fixer les valeurs optimales de chaque filtre, nous avons considéré dans un premier temps le taux de détections correctes maximum par panneau TDC_p et dans un deuxième temps le taux de bonnes détections maximum par nombre d'apparition du panneau TDC_i . Le taux de fausses alarmes TFA comptabilisé correspond au nombre de fausses alarmes dues au filtre colorimétrique. Ce taux n'est pas utilisé ici pour effectuer le choix des valeurs α_r et α_b . On peut cependant noter sur la figure 3.6 que les courbes correspondant à la séquence S4 sont toujours

3. TRANSFORMATIONS UNIVARIÉES

nettement plus à droite que celles de la séquence S1. La nature de la séquence S4 (zone urbaine) est sans doute une des raisons de ce décalage important.

Pour le filtre rouge, on considère les panneaux rouges et les panneaux d'interdiction. La figure 3.6 (a) représente les courbes COR en faisant varier α_r dans l'intervalle $[0,6,0,75]$ par pas de 0,01. Pour la séquence S1, la valeur maximale (91,6%) est atteinte pour toutes les valeurs α_r inférieures à 0,73. Concernant la séquence S4, le taux TDC_p est maximum (92,3%) pour α_r égal à 0,62 ; 0,63 ; 0,65 ou 0,66. Parmi ces 4 valeurs, nous retenons la valeur $\alpha_r = 0,65$. Elle correspond au meilleur taux TDC_i en sommant les résultats des séquences S4 et S1.

La valeur α_b du filtre bleu est fixée de manière analogue et les courbes COR de la figure 3.6 (b) donne les résultats pour α_b variant dans l'intervalle $[0,6,0,65]$ par pas de 0,01. La valeur maximale du TDC_p est égale à 100% pour la séquence S1 et est égale à 93,3% pour la séquence S4 si α_b est supérieur à 0,63. L'analyse des taux de détections correctes par image TDC_i montre un comportement opposé pour les deux séquences. En effet, pour la séquence S1, le TDC_i décroît lorsque α_b augmente alors que pour la séquence S4, le TDC_i augmente avec α_b . La valeur $\alpha_b = 0,63$ offre le meilleur compromis et est retenue pour la suite des travaux.

| Séquence | S1 | S4 |
|-----------------|-------|-------|
| Panneaux rouges | 44/48 | 36/39 |
| Panneaux bleus | 22/23 | 42/45 |

TAB. 3.4 – Nombre maximum de panneaux différents détectés par séquence.

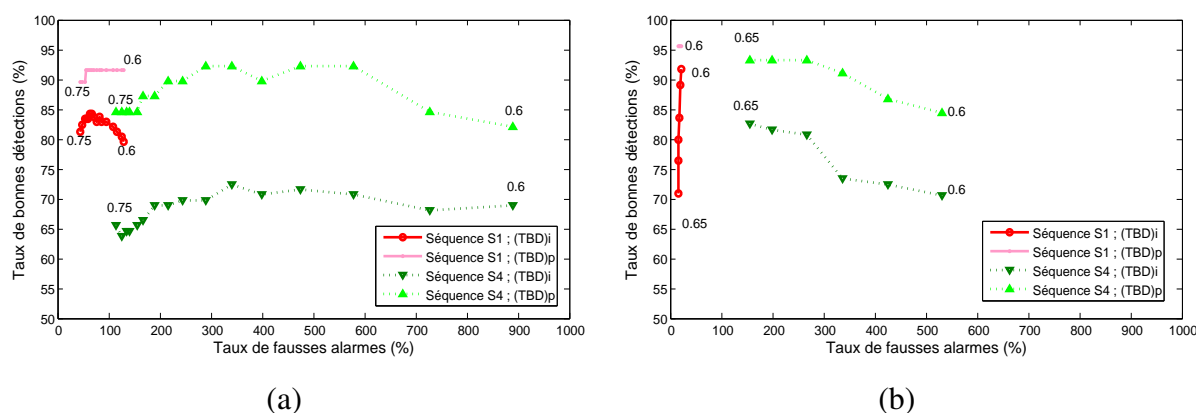


FIG. 3.6 – (a) Détermination du paramètre α_r . Chaque point de chaque courbe correspond à une valeur α_r de l'intervalle $[0,6,0,75]$; (b) Détermination du paramètre α_b . Chaque point de chaque courbe correspond à une valeur α_b de l'intervalle $[0,6,0,65]$

A travers les différents exemples de la figure 3.9, l'algorithme montre son efficacité pour détecter les panneaux dans des situations difficiles : saletés, occultations, orientation non fronto-parallèle. Le nombre de détections correctes s'avère élevé (voir tableau 3.4). Les non-détections sont souvent dues à l'état du panneau en lui-même qui peut présenter une extrême vétusté - cf. Fig. 3.8 (a,c) ou qui ne répond plus aux critères standards actuels - cf. Fig. 3.8 (b). Les panneaux lumineux ne peuvent être détectés puisque les côtés du polygone ne sont pas connexes

3.3. RÉSULTATS EXPÉRIMENTAUX

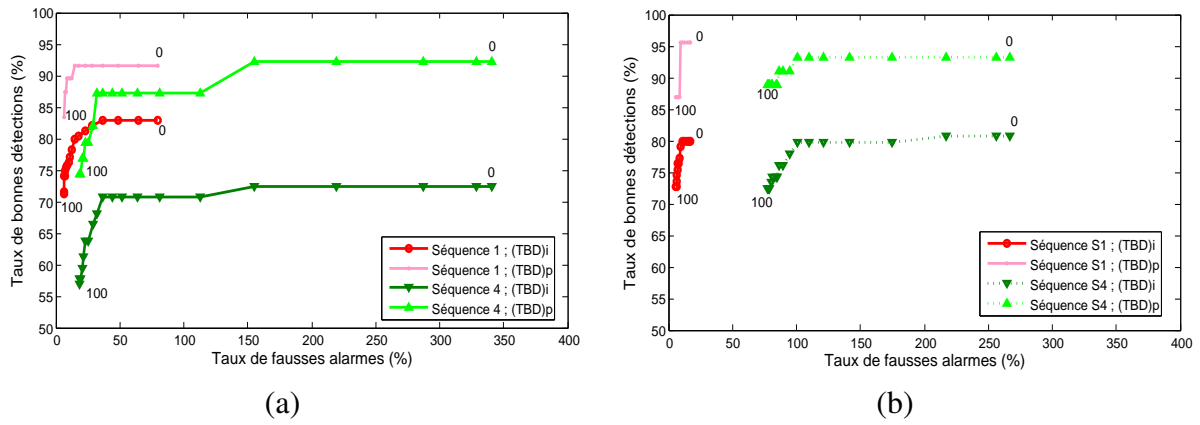


FIG. 3.7 – Influence du seuil de symétrie sur la détection : (a) panneaux d'interdiction ; (b) panneaux d'obligation. Le seuil de symétrie varie dans l'intervalle $[0, 100]$ par pas de 5.

- cf. Fig. 3.8(d). Les autres cas de non-détections sont liés à un problème de segmentation colorimétrique qui ne sélectionne qu'une faible partie de l'objet - cf. Fig. 3.8(e) ou qui fusionne l'objet avec un élément du fond ayant les mêmes propriétés colorimétriques - cf. Fig. 3.8(f). Notons dans ces cas-là, comme pour toutes les non-détections de la séquence S4, que les panneaux non détectés correspondent à des panneaux qui ne sont visibles qu'une seule fois.

3.3.2 Détermination des seuils de symétrie

Les points de l'image d'accumulation, obtenue par la TSR, sont considérés comme centre de symétrie si leur valeur est supérieure à un seuil. La détermination de ces seuils S_{symR} pour les cercles rouges et S_{symB} pour les cercles bleus est effectuée empiriquement par des courbes COR -cf. Fig. 3.7. Sur ces courbes, un seuil égal à 0 correspond à une détection sans utiliser la symétrie.

La sélection par la transformée en symétrie radiale permet de réduire très rapidement le nombre de fausses alarmes. Pour les panneaux rouges de chaque séquence, le facteur de réduction de fausses alarmes est de 2,2 dans l'intervalle $[0, 20]$. Pour la séquence S4 par exemple, le taux de fausses alarmes est en effet de 341% sans symétrie et descend à 155% pour $S_{symR} = 20$. Pour les panneaux bleus, la symétrie engendre également une diminution des fausses alarmes mais avec un facteur moindre. Cette observation est à relativiser car le taux de fausses alarmes est initialement plus faible. Dans l'intervalle $[0, 20]$, le TFA varie d'un facteur 1,88 pour la séquence S4 et 1,55 pour la séquence S1.

Dans l'intervalle $[0, 20]$ considéré, le taux de détections correctes TDC_i diminue très faiblement, et si on considère le TDC_p , il n'y a aucune non-détection supplémentaire aussi bien pour les panneaux d'interdiction (rouges) que pour les panneaux d'obligation (bleus). Pour les panneaux rouges, la diminution de TDC_p commence à partir de $S_{symR} = 25$ (séquence S4) et pour les panneaux bleus, TDC_p baisse à partir de $S_{symR} = 30$ (séquence S1). Les panneaux non-détectés par la symétrie sont dans un premier temps ceux qui ne sont pas fronto-parallèles.

3.3.3 Détection des panneaux polygonaux

Le tableau 3.5 donne les résultats de la détection sur les panneaux polygonaux. Pour les deux types de panneaux, on constate un nombre faible de fausses alarmes. La transformée de Hough s'avère donc très sélective. Le taux de détections correctes TDC_i est relativement élevé pour les deux types de panneaux de la séquence S1 (80% pour les panneaux de danger et 89% pour les panneaux d'indication). Les panneaux manqués correspondent aux cas difficiles dont certains exemples sont représentés sur la figure 3.8. Pour la séquence S4, le TDC_i est très faible (50%) pour les panneaux triangulaires, mais paradoxalement, on constate que tous les panneaux sont au moins une fois détectés. Pour les panneaux d'indication, le TDC_i est égal à 74% avec 3 panneaux jamais détectés.

| Séquence | Type de panneau | TDC_p | TDC_i |
|----------|----------------------|---------|---------|
| S1 | Panneau de danger | 23/26 | 104/129 |
| | Panneau d'indication | 17/18 | 71/91 |
| S4 | Panneau de danger | 9/9 | 18/36 |
| | Panneau d'indication | 24/27 | 45/61 |

TAB. 3.5 – Résultats de détection pour les panneaux de danger et d'indication.

3.3.4 Validation des résultats

L'algorithme est testé sur les séquences S2, S3 et S5 avec le jeu de paramètres $\alpha_r = 0,65$; $\alpha_b = 0,63$; $S_{symR} = 20$ et $S_{symB} = 25$. Les résultats sont rassemblés dans le tableau 3.6. Les résultats sont similaires à ceux obtenus avec des TDC_p souvent supérieurs à 90%, les panneaux manqués correspondant toujours à des situations difficiles. En moyenne les taux de détections correctes sont supérieurs pour les panneaux circulaires (interdiction et obligation) que pour les panneaux polygonaux (danger et indication). Cette situation peut s'expliquer par le fait que les paramètres de la transformée de Hough n'ont pas été empiriquement fixés. Une étude plus approfondie sur cette étape pourrait améliorer les résultats. La séquence S3 présente un nombre plus élevé de non détections mais cette séquence contient un nombre important de panneaux endommagés. Par ailleurs, certaines non-détections pourraient être dues à des artefacts de compression, le taux étant particulièrement élevé sur la séquence S3. Cette hypothèse doit encore être vérifiée. Enfin, la dernière ligne du tableau 3.6 donne les résultats en termes de fausses alarmes. Quatre des cinq séquences (S1, S2, S3 et S5) ont un taux de fausses alarmes du même ordre de grandeur. La spécificité des séquences urbaines, en termes de fausses alarmes, semble ici confirmée, avec un taux nettement supérieur pour la séquence S4.

3.4 Conclusion

L'algorithme de détection se base sur une combinaison d'informations de couleur et de forme. L'analyse des performances montre que l'algorithme semble robuste vis-à-vis des situations complexes (occultations, variations de luminosité, orientation) - cf. Fig. 3.9. Dans le

3.4. CONCLUSION

| Panneaux | Taux de détections correctes | S1 | S4 | S2 | S3 | S5 |
|---------------|------------------------------|------|-------|-----|------|-----|
| Interdiction | TDC_p | 94 | 92 | 100 | 97 | 96 |
| | TDC_i | 86 | 83 | 98 | 9 | 91 |
| Danger | TDC_p | 88 | 100 | 100 | 90 | 97 |
| | TDC_i | 80 | 50 | 96 | 57 | 67 |
| Indication | TDC_p | 95 | 88 | 100 | 82,5 | 100 |
| | TDC_i | 89 | 74 | 75 | 73 | 85 |
| Obligation | TDC_p | 100 | 100 | - | 92 | 89 |
| | TDC_i | 84 | 82 | - | 86 | 90 |
| Tous panneaux | TFA | 46,2 | 276,8 | 20 | 52,5 | 68 |

TAB. 3.6 – Résultats (en %) sur les taux de détections correctes pour les 4 catégories de panneaux et le taux de fausses alarmes total.

chapitre 6, l'évaluation des performances de l'algorithme est étendue aux images de scènes routières fortement urbaines et les résultats sont comparés aux résultats obtenus par les méthodes proposées dans les chapitres 4 et 5.

En moyenne, le temps de calcul est de 1,5 secondes par image (sous *Matlab*® sans prendre en compte la lecture et l'écriture des fichiers). Un développement de l'algorithme en langage C devrait largement accélérer le traitement. D'une façon générale, le problème des non-détections apparaît dès la première étape (comme on peut le voir sur la courbe de la figure 3.6) saturant le taux de détections correctes. Pour de nombreux cas, l'état endommagé du panneau explique ces difficultés. Dans d'autres situations, la saturation est liée soit à une sur-segmentation qui divise les composantes connexes en plusieurs parties, soit à une sous-segmentation qui fusionne différentes composantes connexes. Dans ce dernier cas, La fusion des composantes est souvent due à la complexité de la scène. L'objet est alors détecté mais en incorporant une partie de l'arrière-plan. Dans un avenir proche, des solutions sont envisagées pour tenter de remédier à ce problème, soit en reprenant les travaux de [1] qui étudie les histogrammes de chaque objet pour diviser les composantes soit en effectuant plusieurs filtrages colorimétriques selon différents niveau de sélectivité. Les cartes résultantes seront alors fusionnées. Nos recherches s'orientent désormais vers le développement de la chaîne complète du processus en intégrant la phase de reconnaissance. Les travaux présentés au chapitre 8 serviront de base de travail pour effectuer cette tâche. Nous travaillons également à la détection des panneaux blancs (fin d'interdiction) et des panneaux bleus de fin d'obligation.

3. TRANSFORMATIONS UNIVARIÉES

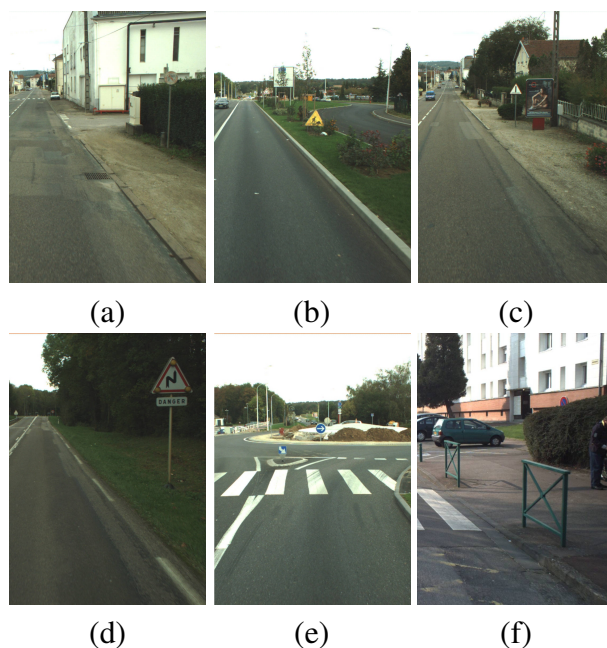


FIG. 3.8 – Exemples de non-détectés : (a) panneau vétuste ; (b) panneau temporaire usagé ; (c) panneau non conforme ; (d) panneau avec flashes aux trois sommets ; (e) balise carrée non sélectionnée par le filtre bleu ; (f) fusion de l'objet panneau avec le mur en arrière plan.

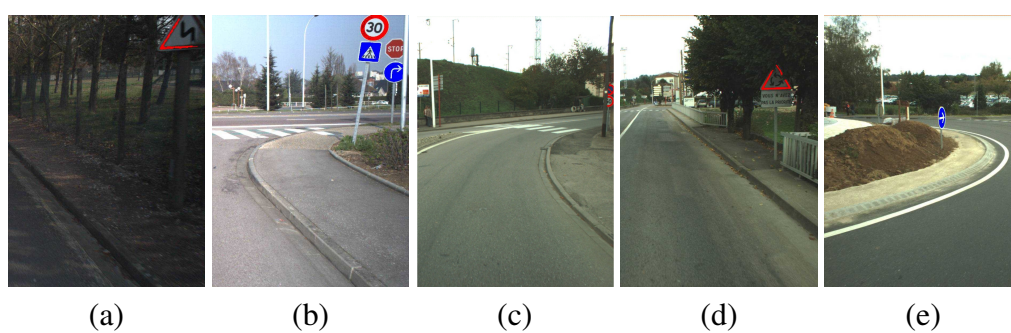


FIG. 3.9 – Exemples de détections : (a) panneau usagé ; (b) cas simples ; (c) deux panneaux proches et occultés ; (d) légère occultation ; (e) panneau non fronto-parallèle. Les détections apparaissent en sur-impression rouge ou bleue sur l'objet considéré.

Bibliographie

- [1] A. BROGGI, P. CERRI, P. MEDICI, P. P. PORTA, et G. GHISIO. Real time road signs recognition. Dans *Proceedings of IEEE Intelligent Vehicles Symposium (IV'07)*, pages 981–986, Istanbul, Turquie, juin 2007.
- [2] G. DUTILLEUX et P. CHARBONNIER. Détection de signalisation routière par ajustement de formes prototypes. Dans *20ème colloque GRETSI sur le traitement du signal et des images*, pages 1109–1112, Louvain-la-Neuve (Belgique), septembre 2005.
- [3] G. DUTILLEUX et P. CHARBONNIER. Métaheuristiques biologiques pour la détection de la signalisation routière. Dans P. SIARRY, éditeur, *Optimisation en traitement du signal et de l'image*, Traité IC2, série traitement du signal et de l'image, chapitre 10, pages 271–294. Hermes, février 2007.
- [4] P. FOUCHER, P. CHARBONNIER, et H. KEBBOUS. Evaluation of a road sign pre-detection system by image analysis. Dans *Proceedings of International Conference on Computer Vision theory and Applications (VISAPP 2009)*, pages 362–367, Lisbonne, Portugal, février 2009.
- [5] LE MINISTRE DE L'INTÉRIEUR et LE MINISTRE DE L'ÉQUIPEMENT ET DE L'AMÉNAGEMENT DU TERRITOIRE. Instruction interministérielle sur la signalisation routière, parties 1 à 5 et partie 8, Arrêté du 7 juin 1977. Dernière mise à jour 2009.
- [6] G. LOY et A. ZELINSKY. Fast radial symmetry for detecting points of interest. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8) :959–973, août 2003.

3. TRANSFORMATIONS UNIVARIÉES

Chapitre 4

Détection et reconnaissance par ajustement de modèles géométriques et radiométriques

*Bahman Soheilian,
Nicolas Paparoditis*

Les panneaux de signalisation routière sont des amers visuels cruciaux pour la navigation dans nos environnements urbains. La détection et la reconnaissance des panneaux dans des images est très utile pour peupler et actualiser des bases de données de routes afin par exemple de calculer des données itinéraires. Ces outils automatiques trouvent également leur intérêt pour créer et enrichir des bases d’amers visuels que l’on peut utiliser dans des systèmes d’aide à la navigation voire de navigation autonome. Une autre application peut également consister à localiser une image acquise par un système nomade par rapport à cette même base dans laquelle les objets sont référencés dans les trois dimensions.

La majorité des articles de l’état de l’art sur ce sujet repose sur des techniques de détection par la couleur et/ou la forme suivie d’une étape de reconnaissance (voir chapitre 2). Le laboratoire MATIS de l’IGN a développé une méthode de détection et de reconnaissance de la signalisation verticale. L’approche est basée sur un appariement avec des modèles géométriques et radiométriques de panneaux. La stratégie générale est décrite dans la section 4.1. Les trois étapes de l’algorithme seront successivement présentés dans la section 4.2 pour la sélection colorimétrique, dans la section 4.3 pour la détection de formes et dans la section 4.4 pour la reconnaissance. Enfin, les résultats expérimentaux seront détaillés dans la section 4.5.

4.1 Notre stratégie

Notre approche est basée sur un appariement avec des modèles grâce à des connaissances *a priori* issues d’un document de référence spécifiant précisément la forme géométrique et les idéogrammes des panneaux de signalisation routière [5]. La figure 4.1(b) montre des exemples de ces panneaux de référence. Notre stratégie générale est montrée sur la Figure 4.1(a). Elle est composée d’un enchaînement en cascade de trois grandes étapes. La première d’entre elles

détecte les pixels d'une couleur donnée. Elle permet d'obtenir des zones de focalisation dans l'image. La seconde étape consiste à détecter des formes géométriques connues de panneaux (cercles, triangles, carrés) par une technique de type RANSAC dans les régions d'intérêts précédemment obtenues. Cette étape propose des candidats potentiels de panneaux. La catégorie de chaque candidat est déterminée en fonction de sa couleur et de sa forme. La dernière étape peut être vue comme une validation permettant de rejeter ou d'accepter ces candidats. Cette étape cherche à apparier par corrélation d'images, chaque candidat avec le bon panneau parmi l'ensemble des panneaux de références de même catégorie. Cependant en milieu urbain les panneaux ne sont pas toujours fronto-parallèles au plan de l'image et les déformations perspectives peuvent donc être très importantes. Grâce aux formes connues et aux formes des panneaux détectées, nous redressons automatiquement l'image localement avant appariement afin de rendre la méthode indépendante de ces déformations perspectives.

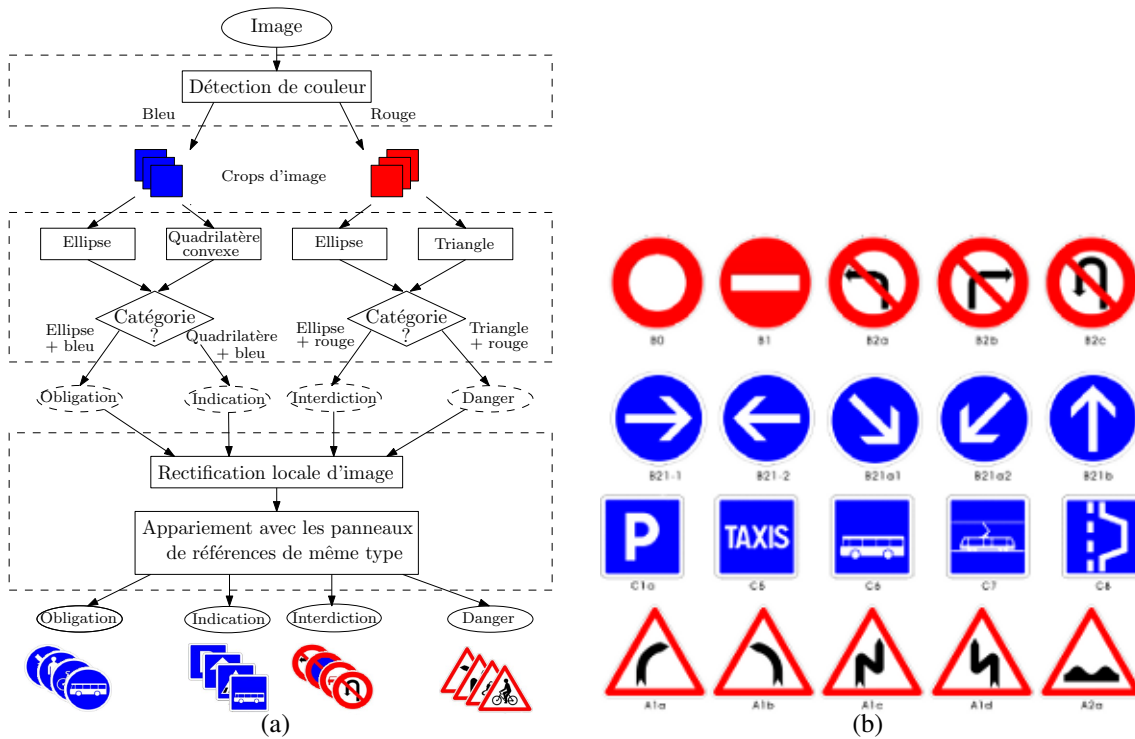


FIG. 4.1 – (a) Notre stratégie générale. (b) Exemples des quatre catégories de panneaux : interdiction, obligation, indication, danger.

4.2 Sélection des couleurs

Nous avons testé et évalué deux méthodes de sélection des couleurs sur un grand nombre d'images obtenues dans des conditions d'illumination variées. Ces deux méthodes sont basées sur les espaces colorimétriques CIE¹ Lab et TSV (Teinte, Saturation, Valeur). Les composantes chromatiques fournies par ces méthodes devraient exprimer la couleur d'objet indépendamment

¹Commission internationale de l'éclairage

4.3. DÉTECTION DES FORMES DES PANNEAUX

de l'éclairement. Mais, en pratique, nous constatons que ces composantes chromatiques ne sont pas complètement indépendantes de l'éclairement. Les résultats fournis ne sont donc pas satisfaisants en présence de variations importantes de l'éclairement. Nous avons donc mis en œuvre une méthode similaire à celle présentée dans [1]. La classification de chaque pixel est faite suivant le critère :

$$I(x,y) \in \begin{cases} C_r & \text{si } \frac{I_r(x,y)}{I_b(x,y)} > T \ \& \ \frac{I_r(x,y)}{I_v(x,y)} > T \\ C_b & \text{si } \frac{I_b(x,y)}{I_r(x,y)} > T \ \& \ \frac{I_b(x,y)}{I_v(x,y)} > T \\ \text{autre} & \text{sinon} \end{cases} \quad (4.1)$$

où :

$I(x,y)$: pixel de coordonnées x et y de l'image I .

I_r, I_v, I_b : valeurs des trois canaux.

C_r, C_b : classes rouge et bleue.

T : seuil à fixer.

Le seuil T est fixé empiriquement sur un grand nombre d'images en faisant un compromis entre sur-détections et sous-détections.

La figure 4.2 montre un exemple d'image et le masque de couleur bleu ainsi calculé. Les pixels connexes sont ensuite labélisés et chaque label est considéré comme une région d'intérêt pour les traitements suivants.

4.3 Détection des formes des panneaux

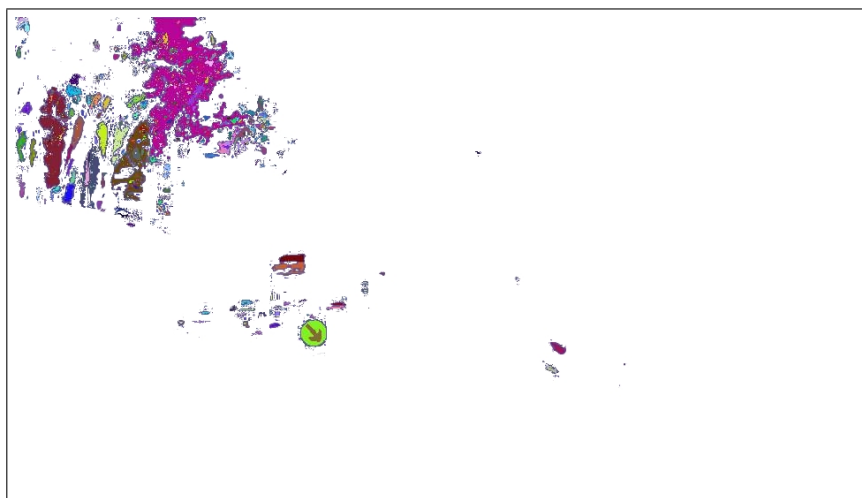
Nous cherchons à détecter les formes géométriques des panneaux (cercles, triangles, carrés) à l'intérieur des régions d'intérêt précédemment obtenues. Cela renforce la fiabilité de la détection et permet de trouver la catégorie du panneau. En général les panneaux ne sont pas fronto-parallèles au plan d'image et sont par conséquent déformés dans les images. C'est pourquoi nous cherchons à détecter des ellipses, des quadrilatères et des triangles (pas forcément équilatéraux). Nous présentons ainsi une méthode de détection pour chacune de ces catégories. Les trois méthodes proposées utilisent les points de contour et l'information de direction du gradient. Nous expliquons tout d'abord le processus d'extraction des points de contours avec leurs tangentes associées. Ensuite nous présentons trois algorithmes dédiés à la détection des ellipses, triangles et quadrilatères. Ces algorithmes sont appliqués à chaque région d'intérêt. La forme finale sera retenue en fonction des degrés de compatibilité fournis par chacun des algorithmes.

4.3.1 Extraction de points de contour orientés

Dans un premier temps, une extraction de contours est effectuée sur chaque région d'intérêt dans l'image en utilisant un opérateur de Canny-Deriche [2]. Les contours sont ensuite chaînés et les positions sous-pixelles des points sont calculées [3]. Pour chaque point de contour on calcule l'orientation de la tangente locale au contour (en faisant une régression linéaire dans un voisinage sur la chaîne de contour). La sortie de cette étape est donc un ensemble de points de contour avec leurs tangentes associées de précision sous-pixelle.



(a)



(b)

FIG. 4.2 – Résultats de détection par le filtre colorimétrique : (a) image initiale ; (b) régions bleues labélisées.

4.3. DÉTECTION DES FORMES DES PANNEAUX

4.3.2 Détection d'ellipse

Une ellipse est définie comme l'ensemble des points vérifiant l'équation :

$$a(x - p)^2 + 2b(x - p)(y - q) + c(y - q)^2 = 1$$

La résolution de ce système non linéaire demande cinq points sur l'ellipse. Notre approche de détection d'ellipse est de type RANSAC et demande le calcul linéaire et rapide des paramètres. Il est possible d'utiliser seulement trois points avec leurs tangentes pour estimer le centre de l'ellipse. Cette méthode est utilisée dans [7] et [6]. Une fois la position du centre de l'ellipse connue, un changement de paramètres permet de ré-écrire l'équation selon :

$$ax^2 + 2bxy + cy^2 = 1$$

Cette équation peut être résolue linéairement avec trois points. On utilise cette technique et on l'intègre dans un algorithme de type RANSAC. Cet algorithme utilise les points de contour (avec tangentes) calculés précédemment et s'écrit de la manière suivante :

1. Sélectionner trois points aléatoirement.
2. Estimer les paramètres d'ellipse.
3. Trouver les points de contour de support.
4. Calculer le degré de compatibilité (C) :

$$C = \frac{\text{Longueur de contour de support}}{\text{Périmètre d'ellipse estimé}}$$

5. Si $(|C - 1| < \epsilon)$ aller à l'étape 1.
6. Sinon enregistrer les paramètres.
7. Répéter les étapes 1-6 n fois.
8. Accepter la plus grande ellipse estimée (s'il y en a une).

Le nombre d'itérations est choisi en fonction du nombre d'observations aberrantes et de la probabilité de réussite [4]. En supposant un maximum de 50% de points aberrants et 95% de probabilité de réussite, le nombre d'itérations théorique est $n = 25$. Pour assurer une détection fiable et précise, nous faisons $n = 100$ itérations. La figure 4.4 montre un exemple de résultat obtenu avec cet algorithme.

4.3.3 Détection de triangle et quadrilatère

Nous proposons un algorithme de type RANSAC dont le schéma est similaire à celui que nous avons présenté précédemment pour la détection d'ellipse. L'estimation du modèle à chaque itération se fait par estimation des sommets d'un triangle ou d'un quadrilatère. A chaque itération trois (resp. quatre) points de contour avec tangentes sont choisis au hasard. Un triangle (resp. un quadrilatère) est ensuite estimé. Grâce à l'information de direction du gradient sur chacun des points, les côtés de la forme sont obtenus et l'intersection des droites correspondantes donne alors les trois (resp. quatre) sommets de l'hypothétique triangle (resp. quadrilatère) - cf. Fig. 4.3(a). Afin de ne garder que les formes plausibles, les deux cas suivants sont éliminés :

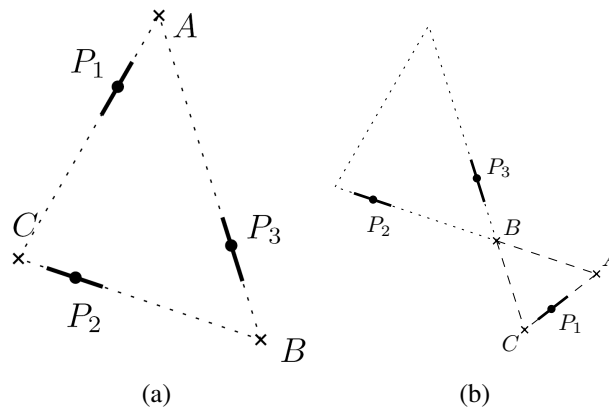


FIG. 4.3 – (a) Estimation d'un triangle avec trois points orientés (l'estimation d'un quadrilatère suit la même méthode) ; (b) Un cas spécial de triangle avec trois points orientés : les sommets estimés sont du même côté par rapport à ces points. Cette configuration est donc éliminée.

- les sommets estimés sont à l'extérieur de la fenêtre d'image.
- les deux sommets estimés avec un point de contour orienté sont du même côté par rapport à ce point, - cf. Fig. 4.3(b).

La seule différence par rapport à l'algorithme de type RANSAC proposé pour l'ellipse est que dans la première étape les trois (resp. quatre) points sont tirés un par un. Après le tirage de chaque point, la droite correspondant à une hypothèse de côté est estimée et les points supports associés sont éliminés. Ces points ne peuvent donc plus être tirés après cette itération. Cela évite le tirage de points qui sont sur la même ligne. Le nombre d'itérations est aussi fixé à $n = 100$. La figure 4.4 montre des exemples de détection de triangle et de rectangle sur les régions rouges et bleues.

4.3.4 Choix de la meilleure forme

Deux algorithmes sont appliqués pour chaque région d'intérêt (détection de quadrilatère et ellipse pour les régions bleues et triangle et ellipse pour les régions rouges). Chaque algorithme propose donc une forme géométrique et fournit aussi son degré de compatibilité (C) avec la région d'intérêt. Nous retenons la forme dont le degré de compatibilité est le plus élevé. A ce stade, grâce à la forme et la couleur, la catégorie (obligation, interdiction, danger, indication) du panneau est estimée - cf. Fig. 4.1(a).

4.4 Validation et identification d'hypothèses de panneaux

4.4.1 Rectification locale d'image

La validation et l'identification des hypothèses panneaux sont effectuées par appariement des textures avec une base de données de panneaux de références. Cependant, les panneaux détectés sont déformés par l'effet de perspective et ne sont donc pas directement comparables avec les panneaux de référence. Cela rend donc l'appariement complexe.

4.4. VALIDATION ET IDENTIFICATION D'HYPOTHÈSES DE PANNEAUX



FIG. 4.4 – (En haut) Les n ellipses, rectangles et triangles estimés par RANSAC. (En bas) Les solutions acceptées.

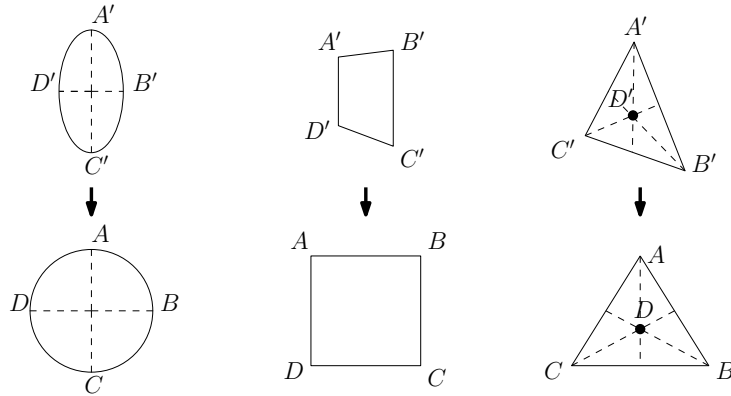


FIG. 4.5 – Les quatre points utilisés en fonction de la forme de panneau pour l'estimation de la transformation projective.

Nous proposons donc de redresser l'image localement afin de la mettre dans la même géométrie que les panneaux de référence. Cela est fait en utilisant une transformation projective à huit paramètres selon l'équation :

$$x = \frac{a_1x' + b_1y' + c_1}{a_3x' + b_3y' + 1} \quad y = \frac{a_2x' + b_2y' + c_2}{a_3x' + b_3y' + 1}$$

Dans le cas des panneaux rectangulaires, ayant quatre points indépendants, ce système peut être résolu de manière exacte quelles que soient la position et l'orientation du panneau dans l'image. Dans le cas des panneaux triangulaires et circulaires on dispose de seulement trois points indépendants (centre et les deux extrémités des axes principaux pour l'ellipse et les trois sommets pour le triangle). Dans le cas général, la solution du système est donc ambiguë. Cette ambiguïté peut être levée de façon approximative en faisant l'hypothèse que les panneaux sont proches de l'axe de visée de la caméra (distance angulaire) et qu'ils sont loin du centre de la caméra

(profondeur). Nous avons utilisé les quatre points présentés dans la figure 4.5 pour calculer la transformation projective. Les images sont ré-échantillonnées en utilisant les paramètres estimés - cf. Fig.4.6. Cela permet de superposer les images d'hypothèses de panneaux avec les panneaux de référence et l'erreur introduite par l'approximation est relativement faible.



FIG. 4.6 – Rectification de panneaux dans le cas de déformations perspectives très fortes.

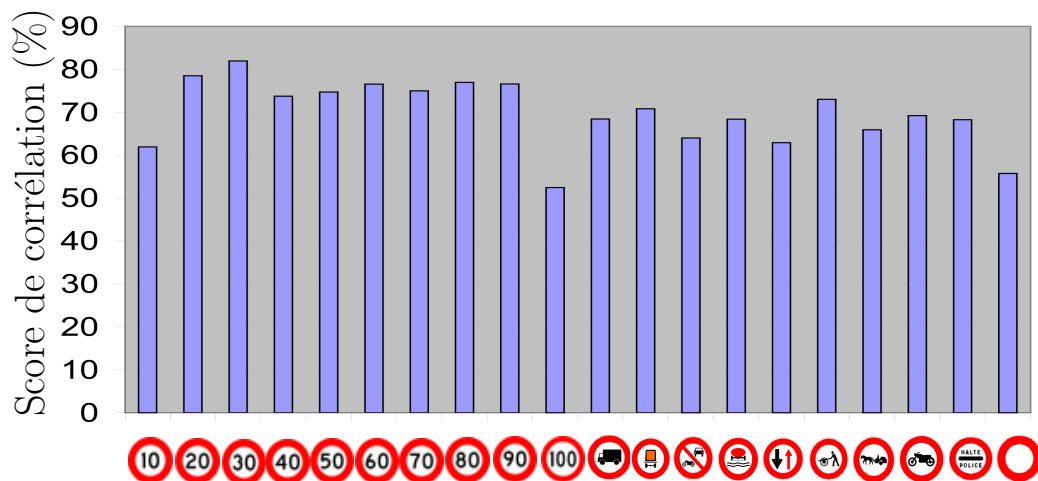


FIG. 4.7 – Résultat d'identification du panneau de limite de vitesse « 30 » détecté dans l'image de la figure 4.6.

4.4.2 Appariement de texture avec les panneaux de référence

Une fois l'image rectifiée, l'hypothèse de panneaux et les panneaux de référence sont dans la même géométrie. L'appariement se fait donc en calculant un score de corrélation entre les pixels de l'image d'hypothèse et les modèles de référence. On a choisi le score de corrélation centrée et normalisée pour être indépendant des possibles variations d'éclairement. Nous rappelons qu'à ce stade de l'algorithme, la catégorie de l'hypothèse de panneau (obligation, interdiction, indication, danger) est déjà déterminée. On apparie donc chaque candidat seulement avec les panneaux de référence de même catégorie. Afin de corréliser uniquement la texture à l'intérieur des panneaux, le score est calculé en appliquant un masque qui dépend de la catégorie

4.5. RÉSULTATS ET ÉVALUATION

de panneau (circulaire, triangulaire ou carré). Dans la majorité des cas le score maximal correspond au bon panneau. Nous acceptons le type de panneaux correspondant au score maximal si le score est supérieur à 75%. Cette tolérance permet de gérer des écarts liés au vieillissement de la texture du panneau mais aussi des différences liées aux effets de ré-échantillonnage des deux images qui ne sont donc pas parfaitement superposables. La figure 4.7 montre le résultat de corrélation de panneau de limite de vitesse « 30 » précédemment détecté sur la figure 4.6. Ce panneau étant fortement déformé par la perspective, la texture est relativement floue après le re-échantillonnage. Cependant nous trouvons correctement le score maximal (85%) pour le bon panneau.

La figure 4.8 montre le résultat de l'ensemble de chaîne de traitement sur une image.

4.5 Résultats et évaluation

4.5.1 Résultats qualitatifs

Nous avons appliqué l'ensemble de la chaîne de traitement (détection et reconnaissance) sur 1000 panoramiques constituées de 1000×8 images de taille 1920×1080 pixels. Ces images ont été obtenues dans le douzième arrondissement de Paris par Stereopolis (le véhicule de cartographie mobile développé au laboratoire MATIS de l'IGN). Les résultats sont intégrés dans le logiciel de navigation dans des séquences d'images panoramiques développé dans le cadre du projet ANR iTOWNS². La figure 4.9 représente des copies d'écrans de ce logiciel. Sur ces images, les bords des panneaux sont représentés par les polygones. Des informations supplémentaires comme le type de panneau et le degré de confiance sont aussi attachées à chaque panneau. La navigation visuelle sur ces panoramiques montre des résultats très satisfaisants du point de vue de la complétude, de la cohérence géométrique et de l'exactitude de reconnaissance du type de panneau.

Les fausses alarmes correspondent souvent aux feux arrières des voitures généralement confondus avec le panneau « sens interdit » -cf. Fig. 4.10. D'autres objets ressemblants aux panneaux sont aussi détectés. Les lettres « O » sur les devantures de commerces sont souvent confondues avec le panneau « circulation interdite » -cf. Fig. 4.11.

4.5.2 Résultats quantitatifs

Pour comparer notre méthode avec d'autres algorithmes une évaluation quantitative est nécessaire. Pour effectuer cette évaluation nous avons utilisé la base de référence créée conjointement par le LCPC/LEPSiS et l'ERA 27 du CETE de l'Est. Cette base contient 847 images. Sur ces images les panneaux circulaires, triangulaires et rectangulaires sont marqués par leurs boîtes englobantes. L'étape de détection de notre algorithme est appliquée aux images de référence (détection de couleurs et de formes géométriques). Les taux de détection et de fausses alarmes sont ensuite calculés. L'influence du paramètre le plus sensible est étudiée (le seuil de l'étape de détection de couleur T de l'équation 4.1). Les résultats sont présentés en détail dans le chapitre 6 sous forme de courbes COR. Les meilleurs taux de bonnes détections sont obtenus pour les valeurs faibles de T ($T = 1, 20$). L'influence de ce seuil est plus importante pour les

²<http://itowns.ign.fr>

4. AJUSTEMENT DE MODÈLE

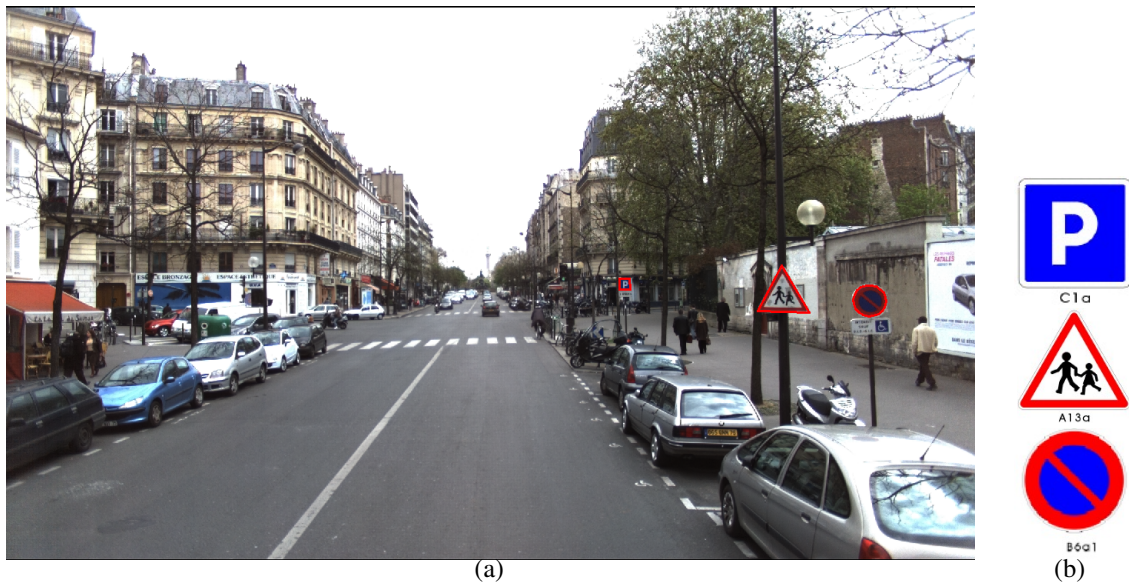


FIG. 4.8 – (a) Résultat de détection et d'identification du panneau dans l'image ; (b) panneaux identifiés.

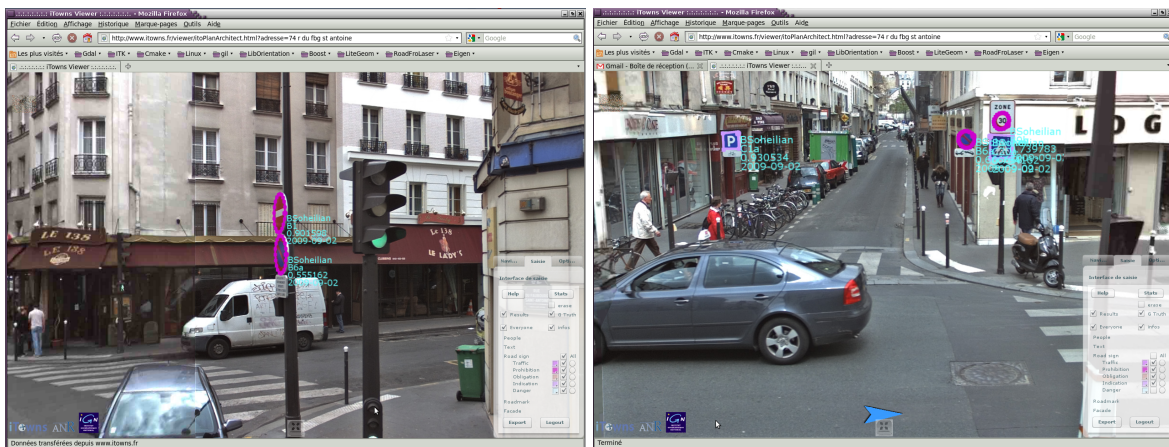


FIG. 4.9 – Intégration des panneaux détectés dans le logiciel de navigation d'images panoramiques développé dans le cadre du projet ANR iTOWNS. Résultats disponibles sur 8000 images.

4.5. RÉSULTATS ET ÉVALUATION

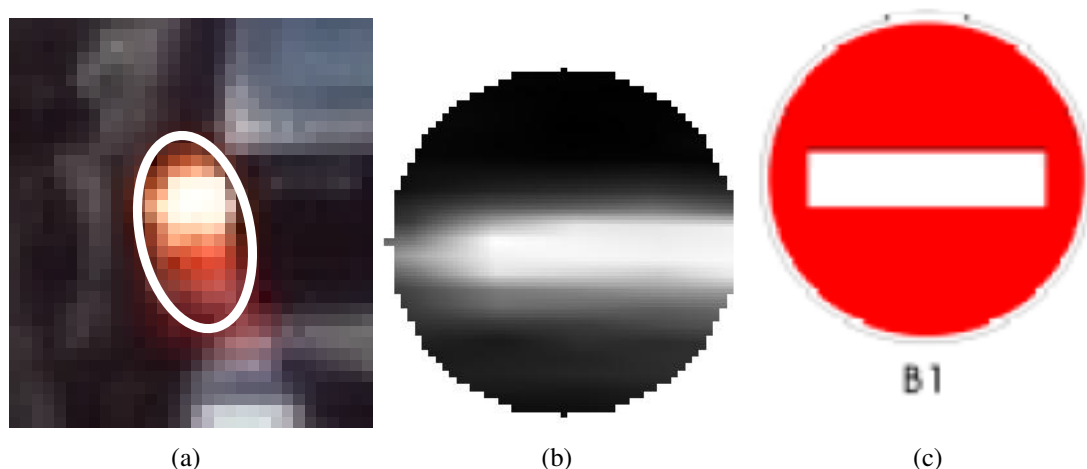


FIG. 4.10 – Exemple de fausse alarme : (a) feu arrière de voiture détecté comme une ellipse, (b,c) ressemblance de l'image rectifiée avec le panneau de sens interdit.



FIG. 4.11 – Exemple de fausse alarme : ressemblance de la lettre « O » et le panneau « circulation interdite ».

panneaux de petite taille ($16 < \text{taille} < 32$). La figure 4.12 montre un exemple de région bleue brisée en morceaux à cause d'une valeur forte de ce seuil ($T > 1,4$). Dans le cas des régions brisées, la détection de formes géométriques est lancée sur chaque région brisée et en conséquence le processus échoue car le bon modèle géométrique ne peut être trouvé.

En choisissant le paramètre qui maximise le taux de bonnes détections pour l'ensemble des panneaux de référence, nous obtenons des taux de bonnes détections et de faux positifs présentés sur les diagrammes de la figure 4.13. Comme on peut le constater sur ces deux diagrammes la qualité de la sélection augmente avec la taille des panneaux (le taux de détection augmente et le taux de fausses alarmes diminue). Quand la taille du panneau diminue, d'une part la détection de couleur a plus de risque d'échouer et d'autre part la qualité du contour extrait diminue de manière importante. Par exemple un taux de fausses alarmes de 200% correspond à 2 fausses alarmes par image. Nous rappelons que la majorité des fausses alarmes peut être éliminée par la

phase de reconnaissance qui n'a pas été évaluée dans cette étude. Dans cette étude, c'est le taux de bonnes détections qui est prépondérant. Le taux de détection est relativement faible pour les panneaux de tailles petites (72% pour les panneaux de taille supérieure à 16 pixels) . Ce taux monte au dessus de 93% pour les panneaux de taille supérieure à 64 pixels. Dans un contexte d'inventaire du mobilier de route avec des images acquises par des systèmes de cartographie mobile cela ne pose aucun problème car l'objet sera observé sur de multiples images. Cela augmentera la probabilité de détection.

4.6 Conclusion

Nous avons présenté un algorithme très prometteur de détection et d'identification des panneaux de signalisation routière à partir d'une unique image en couleurs (non issue d'une séquence) qui enchaîne en cascade trois types de critères discriminants : couleur, forme générale et texture. Notre algorithme est particulièrement performant sur des panneaux non fronto-parallèles avec des déformations importantes, en raison du redressement opéré antérieurement à l'appariement avec une base de données de panneaux de référence.

L'évaluation des performances de l'étape de détection montre l'efficacité de notre algorithme pour la détection des panneaux dans un contexte particulièrement difficile. Un grand nombre de fausses alarmes de l'étape de détection peut être éliminé dans la phase de reconnaissance. Une évaluation quantitative plus poussée doit être faite pour valider les performances de l'étape de reconnaissance. Il faut noter qu'il reste cependant quelques fausses alarmes (voir section 4.5.1) car en milieu urbain beaucoup d'objets peuvent ressembler à des panneaux. Ces ambiguïtés ne peuvent bien entendu être levées qu'avec l'introduction d'imagerie stéréoscopique (par le mouvement ou par des bases stéréoscopiques rigides) ou d'imagerie de profondeur de type LASER. Ces dernières introduisent toutes deux une métrique qui permet par conséquent de filtrer les objets qui n'ont pas en 3D la taille réelle d'un panneau. Ces types d'imagerie sont disponibles sur de nombreux véhicules de numérisation mobile et permettent par conséquent d'envisager l'enrichissement automatique de bases de données de routes pour des applications d'aide à la navigation ou de navigation autonome.

En ce qui concerne le temps de calcul, 2 à 2,3 secondes sont nécessaires pour traiter une image de taille 960×1024 pixels, sans compter le temps de lecture, d'écriture et de l'étape de reconnaissance. Ce temps peut être encore amélioré en utilisant des algorithmes de détection de contours plus rapides de type Prewitt ou Sobel. Pour information, le temps de calcul pour l'étape de reconnaissance est inférieur à 0,5 seconde.

La reconstruction 3D des objets détectés par des techniques de stéréo vision est notre perspective principale, à la fois pour rendre plus robuste le problème de détection mais également pour géoréférencer sa localisation.

4.6. CONCLUSION

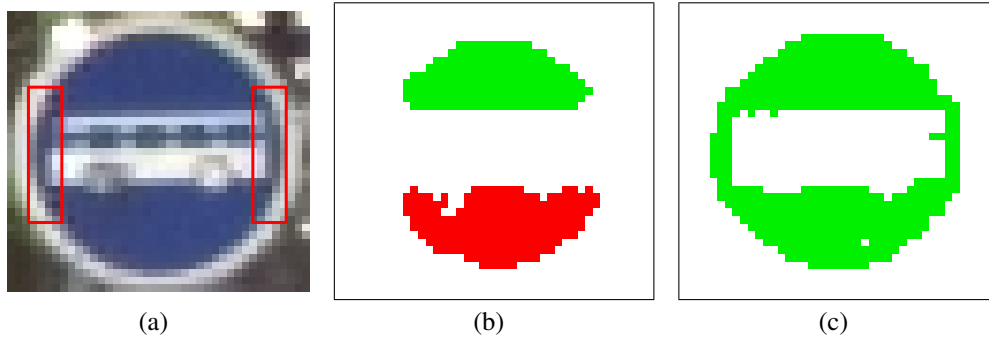


FIG. 4.12 – Exemple d'un panneau bleu (B27a) de petite taille. (a) Les rapports de la valeur bleue sur les valeurs rouge et verte sont plus faibles à l'intérieur des rectangles rouges ; (b) Le panneau est brisé en deux régions à cause de la forte valeur du paramètre T . (c) Une seule région est obtenue en fixant une valeur faible pour le paramètre T .

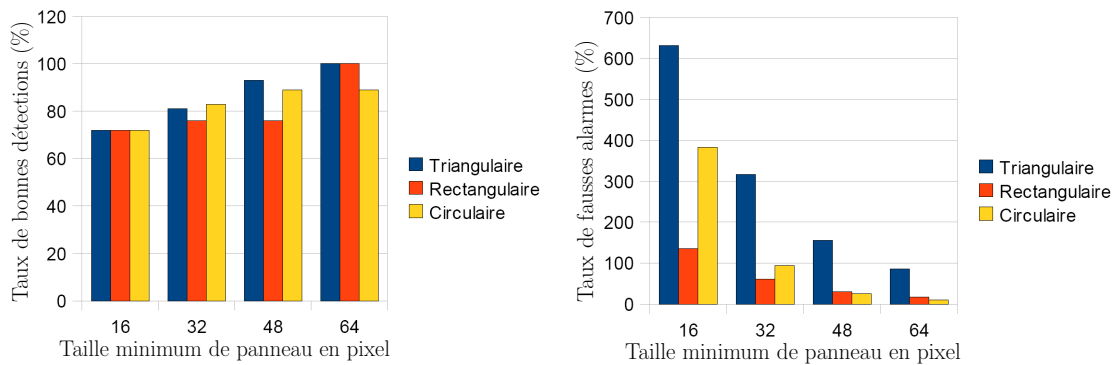


FIG. 4.13 – Taux de bonnes détections et de fausses détections de panneaux triangulaires, rectangulaires et circulaires de différentes tailles. Les statistiques sont obtenues sur 847 images contenant 251 panneaux.

Bibliographie

- [1] A. de la ESCALERA, L.E. MORENO, M.A. SALICHS, et J.M. ARMINGOL. Road traffic sign detection and classification. *IEEE Transactions on Industrial Electronics*, 44(6) :848–859, 1997.
- [2] R. DERICHE. Using canny’s criteria to derive a recursively implemented optimal edge detector. *The International Journal of Computer Vision*, 1(2) :167–187, 1987.
- [3] F. DEVERNAY. A non-maxima suppression method for edge detection with sub-pixel accuracy. Rapport technique RR-2724, INRIA, 1995.
- [4] M. A. FISCHLER et R. C. BOLLES. Random sample consensus : A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6) :381–395, 1981.
- [5] LE MINISTRE DE L’INTÉRIEUR et LE MINISTRE DE L’ÉQUIPEMENT ET DE L’AMÉNAGEMENT DU TERRITOIRE. *Instruction interministérielle sur la signalisation routière - Version consolidée - Deuxième - Cinquième parties*. Paris, France, Arrêté du 7 juin 1977. Dernière mise à jour 2009.
- [6] G. SONG et H. WANG. A fast and robust ellipse detection algorithm based on pseudo-random sample consensus. *Lecture Notes in Computer Science*, Volume 4673/2007 :669–676, 2007.
- [7] S.-C. ZHANG et Z.-Q. LIU. A robust, real-time ellipse detector. *Pattern Recognition*, 38(2) :273–287, 2005.

Chapitre 5

Détection des Panneaux de Signalisation Routière par Accumulation Bivariée

*Rachid Belaroussi,
Jean-Philippe Tarel*

Nous présentons dans ce chapitre une méthode géométrique utilisant l'orientation du gradient le long des contours pour la détection de la signalisation verticale dans des images couleurs, indépendamment de leur position et de leur orientation. La détection est réalisée par une transformation de type accumulateur de Hough bivariée, fondée sur l'utilisation de paires de points avec des contraintes sur les orientations des gradients. Les panneaux circulaires et polygonaux (non triangulaires) sont détectés par la Transformation Chinoise Bilatérale **TCB**. Cette transformation est rapide et ne fait pas de distinction entre les cercles et les polygones de plus de 4 côtés. Le cas des panneaux triangulaires est traité par la Transformation en Sommet et Bissectrice **TSB**, capable de détecter précisément les bissectrices et les sommets d'un triangle.

La section 5.1 fait un rappel de la Transformation en Symétrie Généralisée TSG et introduit les Transformations Chinoise standard TC et Bilatérale TCB. On y propose une analyse comparée de ces opérateurs pour la tâche de détection de panneaux de signalisation circulaires et rectangulaires.

La section 5.2 introduit la Transformation en Symétrie de Coins TSC, et la Transformation en Sommet et Bissectrice TSB. La différence entre ces opérateurs est discutée et l'originalité de l'approche est expliquée.

Les résultats expérimentaux, sur des images synthétiques et des images de route en milieu urbain, sont présentés à la section 5.3. On montre notamment que la TCB et la TSB sont indépendantes de l'orientation globale du panneau, indépendantes du contraste clair/sombre ou sombre/clair, et sont rapides et efficaces en environnement extérieur.

5.1 Détection de panneaux circulaires et rectangulaires

5.1.1 Transformation en Symétrie Généralisée TSG

La Transformation en Symétrie Généralisée (**TSG**) [12] est un opérateur sans contrainte d'orientation absolue, qui dans son principe cherche à détecter des portions de contours symé-

5. TRANSFORMATIONS BIVARIÉES

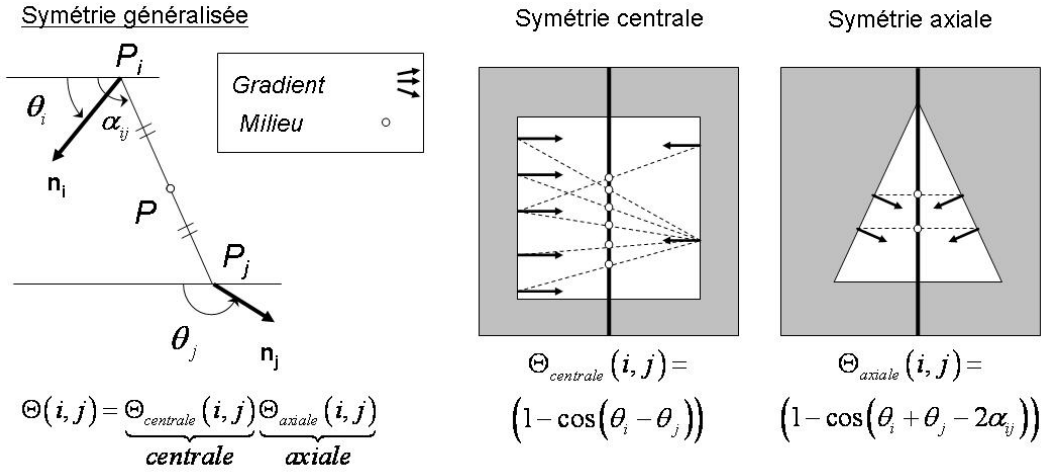


FIG. 5.1 – La symétrie centrale et symétrie axiale entre deux vecteurs gradient \mathbf{n}_i et \mathbf{n}_j peut être modélisée par des contraintes sur les paramètres $(\theta_i, \theta_j, \alpha_{ij})$.

triques. La TSG a inspiré différents algorithmes, dédiés à un type de symétrie ou à un type de forme à détecter.

La TSG calcule un indice de symétrie à partir des orientations des gradients de deux pixels (P_i, P_j) en votant au point milieu P du segment $[P_i P_j]$. Pour chaque point P d'une image, un ensemble de votants est défini par :

$$\Gamma(P) = \{(P_i, P_j) \mid \frac{P_i + P_j}{2} = P\} \quad (5.1)$$

Chaque pixel P_i a un vecteur gradient \mathbf{n}_i faisant un angle θ_i par rapport à l'horizontale. Chaque paire de vecteurs gradients $(\mathbf{n}_i, \mathbf{n}_j)$ est plus ou moins proche d'une certaine symétrie axiale et centrale selon les orientations (θ_i, θ_j) . Cette proximité est estimée par une fonction à deux termes explicités dans la figure 5.1 et nommée fonction de phase. Cette fonction est ensuite pondérée par une fonction de la distance $\|\mathbf{P}_i \mathbf{P}_j\|$, et par la norme des gradients $r_i = \|\mathbf{n}_i\|$ et $r_j = \|\mathbf{n}_j\|$. La fonction $D(i, j)$ utilisée pour la distance permet d'atténuer les votes de paires éloignées, donc de rendre la détection locale à une certaine échelle. L'indice de symétrie en un point P est la somme sur l'ensemble Γ_P de toutes ces contributions :

$$S(P) = \sum_{(i,j) \in \Gamma_P} \underbrace{D(i, j)}_{\text{distance}} \underbrace{\Theta_{axiale}(i, j)}_{\text{Symétrie axiale}} \underbrace{\Theta_{centrale}(i, j)}_{\text{Symétrie centrale}} \underbrace{r_i r_j}_{\text{Module gradients}} \quad (5.2)$$

En pratique, r_i peut être une fonction logarithme du module du gradient : $r_i = \log(\|\mathbf{n}_i\|)$ ou une autre fonction croissante. Au contraire, $D(i, j)$ est une fonction décroissante avec la distance $\|\mathbf{P}_i \mathbf{P}_j\|$, typiquement :

$$D(i, j) = \frac{1}{\sqrt{2\pi}\sigma} \exp^{-\frac{\|\mathbf{P}_i - \mathbf{P}_j\|}{2\sigma}} \quad (5.3)$$

où le facteur d'échelle σ dépend de la taille des objets recherchés. La formule de $\Theta(i, j)$ est indiquée sur la figure 5.1. Elle combine un indice de symétrie centrale et un indice de symétrie axiale : $\Theta(i, j) = \Theta_{refl}(i, j) \times \Theta_{rad}(i, j)$.

5.1.2 Transformation Chinoise TC

Pour modéliser la symétrie axiale, une autre approche a récemment été proposée, la TC [10], et appliquée à la détection des yeux sur des images de visages.

Comme pour la TSG, l'espace d'accumulation de la TC est celui de l'image, et une paire de points de contour vote pour leur milieu. La différence entre ces deux transformations est la fonction de phase $\Theta(i, j)$, c'est-à-dire les conditions requises pour former une paire. La Transformation Chinoise ne fait appel qu'à un terme de symétrie centrale, mais avec une condition sur l'alignement qui permet d'englober aussi le cas de la symétrie axiale :

$$\Theta(i, j) = \underbrace{W_\beta(\theta_i - \alpha_{ij})}_{\text{Alignement restreint}} \underbrace{W_\delta(|\theta_i - \theta_j| - \pi)}_{\text{Symétrie centrale}} \quad (5.4)$$

où W_R est un noyau uniforme de rayon R : $W_R(x) = 1$ si $|x| < R$, 0 sinon. Le premier terme de la fonction de phase est une contrainte sur l'alignement entre $\mathbf{P}_i \mathbf{P}_j$ et \mathbf{n}_i . La valeur de δ est petite pour assurer que la deuxième contrainte $|\theta_i - \theta_j| = \pi$ est vérifiée numériquement, de sorte que seuls les gradients ayant une direction opposée sont inclus dans l'ensemble des votants. Comme pour la TSG, un facteur d'échelle σ est utilisé pour définir la région d'influence de chaque pixel par le biais de $D(i, j)$.

La TC peut sembler proche de la TSG mais elle est assez différente. En effet, bien qu'elle ne modélise pas explicitement la symétrie axiale, elle est aussi capable de détecter des axes de symétrie (cas de la détection des doigts de la main [2]). Autre exemple, les accumulateurs obtenus avec ces transformées sont comparés dans la figure 5.2, dans le cas d'une image avec un rectangle blanc sur fond noir. Avec la TC, la majeure partie des votes est accumulée près des axes, tandis que pour la TSG, la contribution des paires de gradients se diffuse loin des axes. Les deux profils obtenus sur les deux accumulateurs au même endroit (indiqué par des flèches rouges) montrent la distribution horizontale de l'indice de symétrie pour chacune des transformations. Le pic central est le point d'intersection des axes. On observe que la TC résulte en une distribution quasi-uniforme des votes tandis que les votes de la TSG forment un dôme : la TC est donc plus sélective que la TSG. La figure 5.3 illustre le cas d'un disque blanc sur fond noir : la TC est aussi plus précise que la TSG. La figure 5.5 montre l'accumulateur obtenu par la TC sur des panneaux circulaire et polygonal.

5.1.3 Transformation Chinoise Bilatérale TCB

La TC fait toutefois l'hypothèse que l'objet est clair sur un fond sombre dans l'image, de sorte que les paires de points à considérer est réduit aux couples (P_1, P_2) ayant des gradients convergents - cf. Fig. 5.4(b). Cette hypothèse est généralement vérifiée dans le cas d'un iris ou d'un œil, mais pas dans le cas de la détection de panneaux. Elle peut être assurée par une segmentation couleur, notamment si on cherche à détecter des panneaux d'une certaine couleur [7]. Pour prendre en compte les cas de gradients d'orientations divergents, nous proposons plutôt de modifier le terme de phase de la TC. Nous avons testé deux alternatives : dans la première, les gradients divergents ont une contribution négative (c'est la TCB signée, illustrée par les figures 5.4 et 5.6), et dans la seconde, ils ont la même contribution positive (c'est la TCB non signée) :

5. TRANSFORMATIONS BIVARIÉES

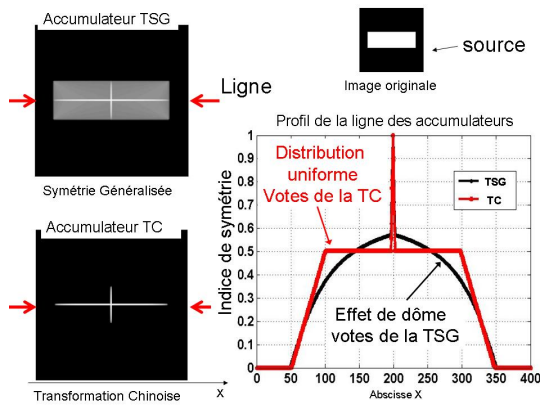


FIG. 5.2 – Accumulateurs de la TSG et de la TC pour une forme rectangulaire : les deux axes de symétrie (horizontal et vertical) sont mis en valeur.

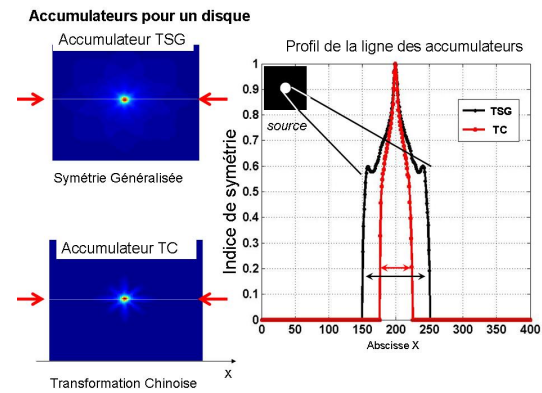


FIG. 5.3 – Accumulateurs de la TC et de la TSG dans le cas d'une forme circulaire. La localisation du centre par la TC est plus précise et le pic de la TSG est deux fois plus large que celui de la TC.

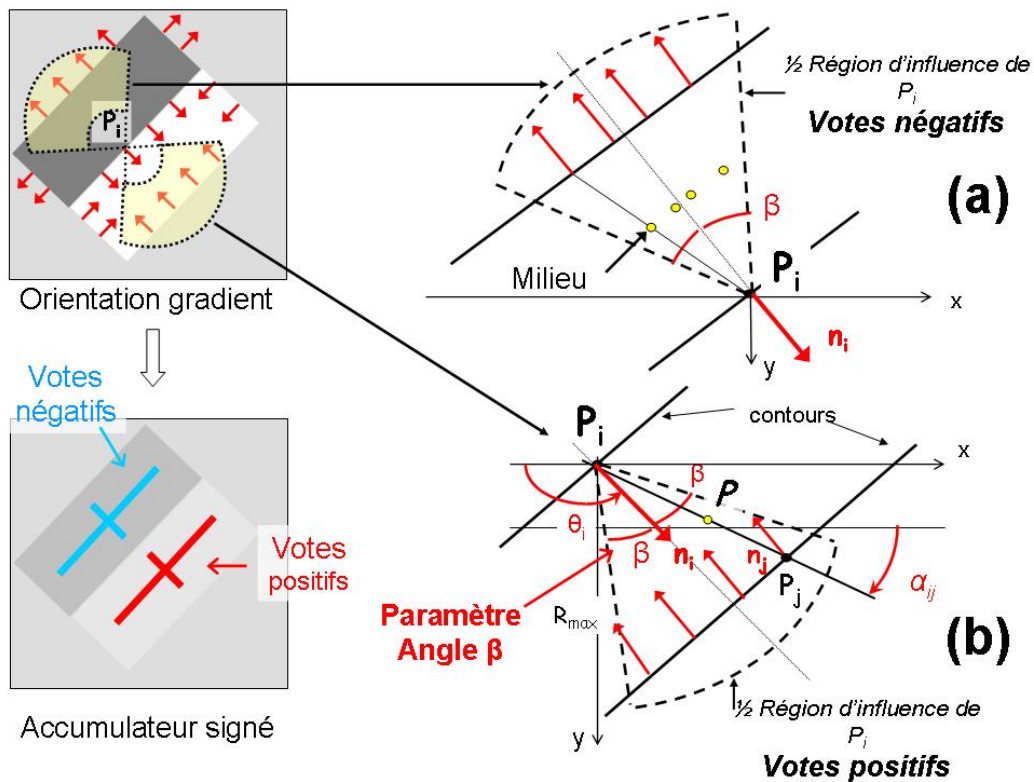


FIG. 5.4 – Transformation Chinoise Bilatérale signée : (a) Dans le cas d'un contraste sombre/clair, une paire de points de contour avec des gradients d'orientations opposées (et divergents) ($\mathbf{n}_i, \mathbf{n}_j$) résulte en une contribution négative dans l'accumulateur. (b) Une paire de points de contour avec des gradients d'orientations opposées ($\mathbf{n}_i, \mathbf{n}_j$) mais pointant l'un vers l'autre contribue positivement à l'accumulateur (rectangle clair/sombre).

5.1. DÉTECTION DE PANNEAUX CIRCULAIRES ET RECTANGULAIRES

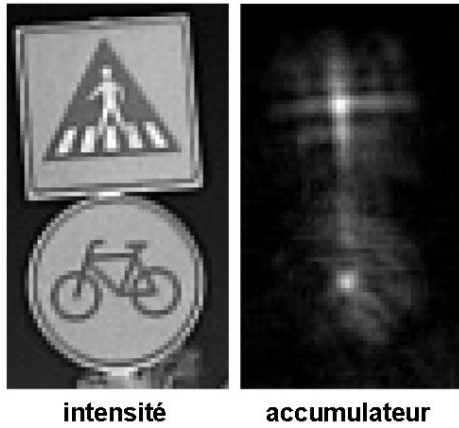


FIG. 5.5 – Accumulateur de la TC standard pour une forme rectangulaire et une autre circulaire : le centre de symétrie est bien détecté dans les deux cas.

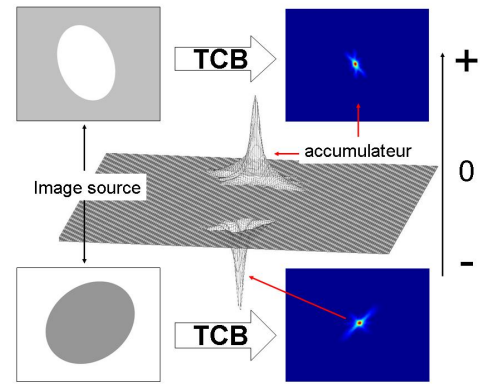


FIG. 5.6 – Accumulateur de TCB signée dans le cas d'ellipses clair/sombre et sombre/clair : la TCB est invariante aux rotations dans le plan image.

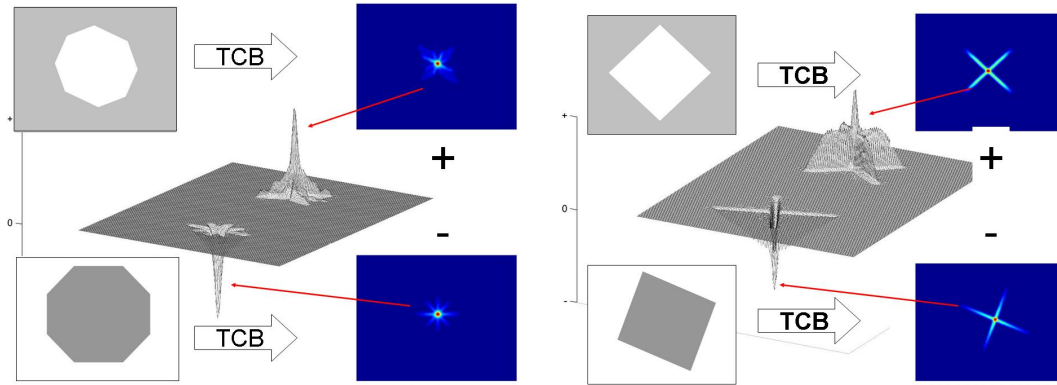


FIG. 5.7 – TCB signée dans le cas de formes polygonales : image en niveaux de gris et accumulateur correspondant. Les contours des objets de contraste clair/sombre (en haut) ont une contribution positive dans l'accumulateur ; ceux de contraste sombre/clair (en bas) contribuent avec un vote négatif.

$$\Theta(i, j) = \left(\underbrace{W_{\beta}(\theta_i - \alpha_{ij})}_{\text{clair/sombre}} \underbrace{\pm}_{\text{signe}} \underbrace{W_{\beta}(\theta_i - \alpha_{ij} - \pi)}_{\text{sombre/clair}} \right) \times \underbrace{W_{\delta}(|\theta_i - \theta_j| - \pi)}_{\text{Symétrie centrale}} \quad (5.5)$$

où δ est la précision avec laquelle l'assertion $|\theta_i - \theta_j| = \pi$ est vérifiée numériquement ($\delta = \frac{2\pi}{N}$ si on quantifie l'orientation du gradient sur N valeurs). Comme précédemment, α_{ij} est l'angle que fait le vecteur $\mathbf{P}_i\mathbf{P}_j$ avec l'axe horizontal -cf. la convention en figure 5.1). β est la tolérance angulaire entre les vecteurs \mathbf{n}_i et $\mathbf{P}_i\mathbf{P}_j$ et sur l'alignement entre les gradients \mathbf{n}_i et \mathbf{n}_j . Ce paramètre délimite la zone d'influence du point P_i comme l'indique la figure 5.4. Une valeur petite de β est adaptée à une forme strictement circulaire tandis que la valeur $\beta = \pi/2$ peut être utilisée pour des formes plus variées. Néanmoins $\beta = \pi/2$ augmente substantiellement le nombre de vo-

tants et par voie de conséquence le temps de calcul, et en en général le nombre de faux positifs. Concernant la pondération $D(i, j)$ par la distance, il est plus rapide d'utiliser un seuillage sur la distance entre deux points, plutôt qu'une gaussienne :

$$D(i, j) = W_{R_{max}}(\|P_i - P_j\|) - W_{R_{min}}(\|P_i - P_j\|) \quad (5.6)$$

où R_{min} et R_{max} sont respectivement la taille minimale et maximale d'un panneau. Les panneaux sont donc recherchés dans la gamme de taille $[R_{min}, R_{max}]$. L'accumulateur est incrémenté d'une valeur proportionnelle au produit de $\Theta(i, j)$ avec les fonctions du module des gradients $r_i = \log(1 + \|\mathbf{n}_i\|)$, soit :

$$Accu_{TCB}(P) = \sum_{(i,j) \in \Gamma(P)} D(i, j) \Theta(i, j) r_i r_j \quad (5.7)$$

Dans le cas où l'on cherche à détecter des formes polygonales ayant quatre côtés ou plus, il faut utiliser une valeur de $\beta = \frac{\pi}{8}$ au moins. Cela peut se faire en quantifiant l'orientation du gradient sur 8 directions comme dans le code de Freeman.

La TCB a pour sortie deux accumulateurs, un pour les centres de symétrie des objets, et un autre dont est déduit leur rayon. Le centre de symétrie d'un objet peut être extrait de l'accumulateur (5.7). Par contre, pour déterminer l'étendue spatiale de l'objet détecté, un autre accumulateur est nécessaire pour enregistrer la somme des distances entre votants $\|P_i - P_j\|$, ce qui permet de calculer leur moyenne :

$$Rayon(P) = \frac{1}{2N} \sum_{(i,j) \in \Gamma(P)} \|P_i - P_j\| \quad (5.8)$$

où $N = Card(\Gamma(P))$. L'estimation du rayon de l'objet détecté est avantageuse, en termes de taille mémoire, vitesse et précision, comparée aux méthodes multi-échelles telles que celles présentées dans [1, 9] qui utilisent un accumulateur par rayon.

Dans la TCB signée, les objets sombre/clair ont une contribution négative tandis que ceux de contraste clair/sombre ont un vote positif. La figure 5.6 montre l'accumulateur dans le cas de formes elliptiques avec des contrastes et des orientations différents : on remarque que la TCB est invariante aux rotations dans le plan de l'image. La robustesse à de petites rotations hors du plan de l'image est illustrée avec des polygones sur la figure 5.7. Comme l'expliquent Park *et al.* [11], l'emploi d'une méthode de votes signés renforce la sélectivité de la transformation. Cela dit l'application de la TCB signée est problématique dans le cas de panneaux non plein (indication et interdiction). Les deux types de contraste sont alors présents et les votes négatifs diminuent la contribution des votes positifs pour le centre du panneau. Les résultats expérimentaux montrent ainsi que la TCB non signée améliore notablement les performances en détection par rapport à la TCB signée.

5.2 Détection de panneaux triangulaires

Le cas des panneaux triangulaires ne peut pas être traité par les transformées précédentes. Nous proposons donc une transformation spécifique pour les triangles qui combine détection des sommets et du centre du triangle.

5.2.1 De la TSG à la détection de coins

La Transformation en Symétrie de Coins (TSC) introduite dans [6], adapte la TSG aux caractéristiques de symétrie d'un angle γ de valeur angulaire donnée. Ainsi, la contribution d'une paire de points (P_i, P_j) à la fonction de phase $\Theta(i, j)$ est constituée de deux termes : le premier modélise la symétrie axiale et le second estime le degré d'appartenance à la classe des coins d'angle γ . Le vote du couple (P_i, P_j) est aussi accumulé en leur milieu P :

$$\Theta_\gamma(i, j) = \underbrace{\Theta_{axiale}(i, j)}_{\text{Symétrie axiale}} \exp \left(- \frac{(\beta_{ij} (|\theta_i - \theta_j| - \pi - \gamma))^2}{2\sigma_\gamma} \right) \quad (5.9)$$

Le terme de symétrie axiale est maximal lorsque les gradients sont symétriques par rapport à la médiatrice du segment $[P_i P_j]$. Le second terme a une valeur maximale lorsque le coin d'angle $||\theta_i - \theta_j| - \pi|$ est égal à γ . σ_γ est un paramètre contrôlant l'adéquation à γ . $\beta_{ij} = \left[\pi \sqrt{1 - \cos^2(\theta_i - \theta_j)} \right]^{-1}$ est une fonction de rejet assurant que P_i et P_j ne sont pas alignés sur le même contour. Enfin, l'indice de symétrie de la TSC en P , pour un angle γ donné, est la somme des contributions des paires de points de Γ_P :

$$S_\gamma(P) = \sum_{(i,j) \in \Gamma_P} D(i, j) \Theta_\gamma(i, j) r_i r_j \quad (5.10)$$

A notre avis, la pertinence du terme de symétrie axiale est discutable : deux points appartenant à des côtés adjacents d'un angle ne vérifient pas nécessairement cette condition de symétrie comme le montre la figure 5.9. Il semble préférable d'exclure ce critère de symétrie par rapport à la médiatrice. De plus, dans la TSC, le vote d'une paire de points est accumulé en leur milieu, ce qui ne met pas en valeur chaque coin mais un nuage de points situés près du coin : les résultats de la TSC sont donc biaisés.

5.2.2 Transformation en Sommet et Bissectrice TSB

Dans l'approche que nous proposons, une paire de points (P_i, P_j) vote pour le sommet du coin A auquel ils appartiennent :

$$\Gamma(A) = \{(P_i, P_j) \mid \mathbf{AP}_i \cdot \mathbf{n}_i = 0, \quad \mathbf{AP}_j \cdot \mathbf{n}_j = 0\} \quad (5.11)$$

$\mathbf{n}_i = \left[\frac{\partial I}{\partial x} \quad \frac{\partial I}{\partial y} \right]^T$ étant le gradient de l'image I au point $P_i = (x_i, y_i)$, l'équation de la perpendiculaire au gradient en P_i est :

$$y = a_i (x - x_i) + y_i \quad \text{où} \quad a_i = - \frac{\frac{\partial I}{\partial x}}{\frac{\partial I}{\partial y}} \quad (5.12)$$

En conséquence, le sommet A de l'angle formé par les droites perpendiculaires aux gradients en P_i et P_j a pour coordonnées :

$$\begin{cases} x_A = \frac{1}{a_i - a_j} (y_j - a_j x_j - y_i + a_i x_i) \\ y_A = a_i (x_A - x_i) + y_i \end{cases} \quad (5.13)$$

5. TRANSFORMATIONS BIVARIÉES

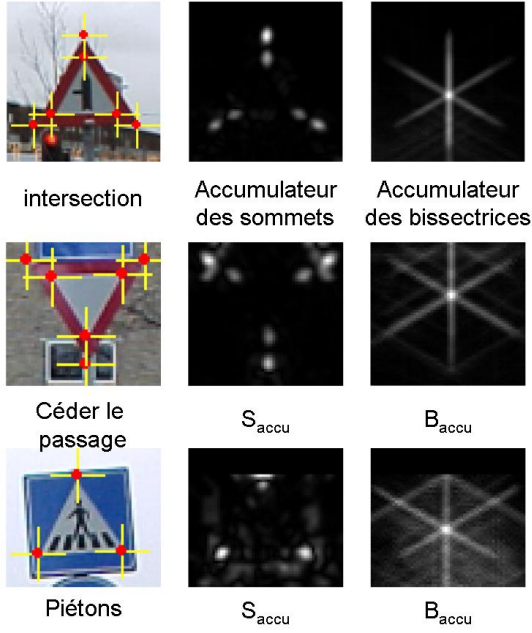


FIG. 5.8 – Les trois types de formes triangulaires présents dans la base de test, avec les sommets détectés (croix jaunes) en seuillant l'accumulateur S_{accu} de la TSB.

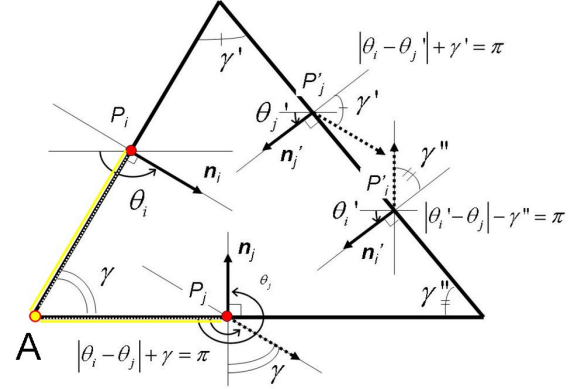


FIG. 5.9 – Orientation du gradient des côtés d'un triangle. La relation entre les orientations θ_i et θ_j de deux côtés adjacents dépend de l'angle au sommet γ , car chaque vecteur gradient \mathbf{n}_i et \mathbf{n}_j est orthogonal à son côté.

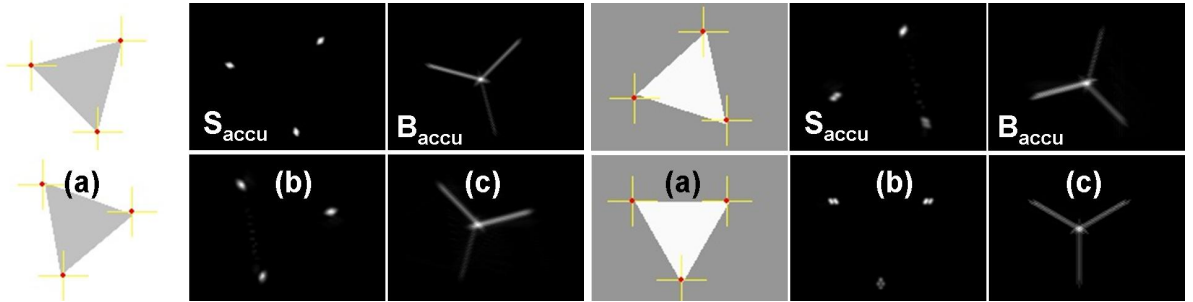


FIG. 5.10 – Cas de triangles clair/sombre ou sombre/clair avec différentes orientations. Colonne (a) : image source (niveau de gris) avec les sommets détectés en jaune. Colonne (b) : Accumulateur des sommets S_{accu} . Colonne (c) : Accumulateur pour les bissectrices B_{accu} . La TSB est relativement invariante aux rotations dans le plan.

On peut alors construire un tableau accumulant la présence d'un sommet en A comme dans le cas de la TSG :

$$S_{accu}(A) = \sum_{(i,j) \in \Gamma(A)} D(i,j) \Theta_{\gamma}(i,j) r_i r_j \quad (5.14)$$

avec

$$D(i,j) = W_{L_{max}}(\|\mathbf{P}_i \mathbf{P}_j\|) \quad (5.15)$$

où $W_{L_{max}}$ est un noyau uniforme de rayon L_{max} . De cette manière, les pixels trop éloignés ne sont

5.2. DÉTECTION DE PANNEAUX TRIANGULAIRES

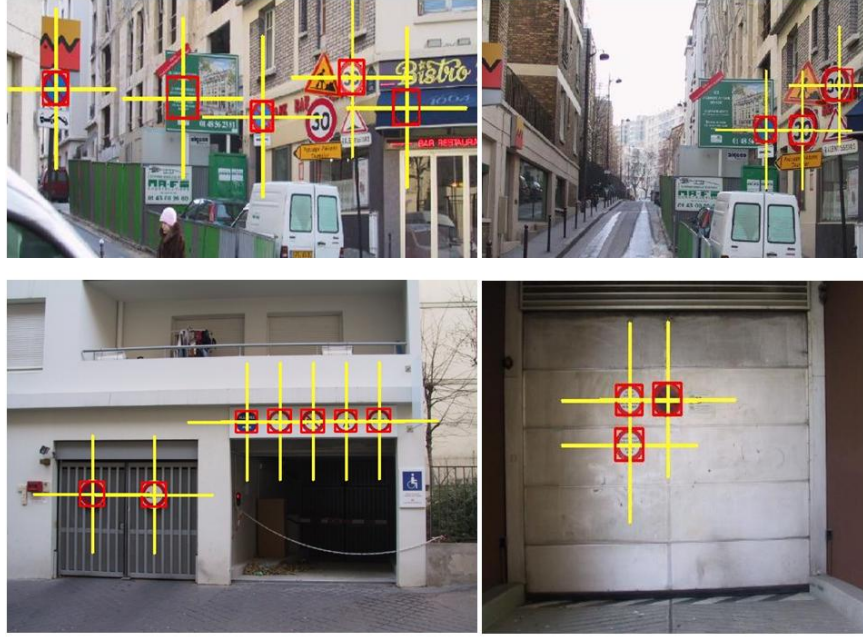


FIG. 5.11 – Exemples de segmentation de panneaux par TCB sur une scène urbaine (base de test Paris 15°).



FIG. 5.12 – Exemples de détection de panneaux triangulaires de la base de test avec leurs deux accumulateurs S_{accu} et B_{accu} .

5. TRANSFORMATIONS BIVARIÉES

pas appariés. On choisit pour la pondération par rapport aux modules du gradient une fonction logarithme r_i et r_j la même fonction que pour la TCB :

$$r_i = \log(1 + \|\mathbf{n}_i\|) \quad (5.16)$$

On peut implanter une transformation générale si on ne spécifie pas d'angle γ , la fonction de phase étant alors définie par :

$$\Theta(i, j) = \Theta_{//}(i, j) = 1 - \delta(\cos(\theta_i - \alpha_{ij})) \quad (5.17)$$

où $\alpha_{ij} = \angle(O\mathbf{x}, \mathbf{P}_i\mathbf{P}_j)$ est l'angle de $\mathbf{P}_i\mathbf{P}_j$ par rapport à l'horizontale - cf. Fig. 5.1, et δ est le symbole de Kronecker. La fonction $\Theta_{//}(i, j)$ rejette les cas où $|\theta_i - \alpha_{ij}| = \pi/2$ modulo π , P_i et P_j étant alors sur des contours parallèles. Les maxima locaux du tableau S_{accu} , après seuillage permettent de détecter les sommets d'angles.

Pour détecter le centre du panneau, la paire de points (P_i, P_j) vote aussi pour une portion $[AB]$ de la bissectrice de l'angle au sommet A, donc d'angle $\angle(P_iAB) = \angle(P_iAP_j)/2$:

$$\Gamma(AB) = \{(P_i, P_j) \mid \mathbf{AP}_i \cdot \mathbf{n}_i = 0, \mathbf{AP}_j \cdot \mathbf{n}_j = 0, \angle(P_iAB) = \angle(P_iAP_j)/2\} \quad (5.18)$$

Le segment $[AB]$ est construit de façon à ce que sa longueur soit la taille maximale L_{max} du triangle à détecter dans l'image. Un indicateur de la localisation des bissectrices est donc simultanément construit en accumulant les votes selon les segments $[AB]$:

$$B_{accu}([AB]) = \sum_{(i,j) \in \Gamma(AB)} D(i, j) \Theta_\gamma(i, j) r_i r_j \quad (5.19)$$

Par construction, chaque bissectrice d'un angle apparaît sous la forme d'un segment dans l'accumulateur B_{accu} . Si une forme triangulaire est présente dans l'image, son centre est donc un maximum local de B_{accu} du fait que les trois bissectrices se coupent en un même point.

Nous illustrons les résultats obtenus par la transformation TSB sur des images synthétiques en niveaux de gris, en particulier les deux accumulateurs. Sur la figure 5.10, des triangles blancs sur fond gris ou gris sur fond blanc sont traités. Dans les deux cas, on observe que la TSB est invariante aux rotations dans le plan de l'image, aux erreurs d'échantillonnage près.

5.2.3 Détection de Signalisation Triangulaire

Dans le cas des panneaux triangulaires, des contraintes supplémentaires nous amènent à définir une fonction de phase plus spécifique $\Theta(i, j)$. Lorsque le triangle est approximativement équilatéral : l'angle de base est $\gamma_0 = \pi/3$ à une précision $\Delta\gamma$ donnée. Deux points de contour de gradient \mathbf{n}_i et \mathbf{n}_j , ayant pour orientation respectivement θ_i et θ_j , appartiennent à deux côtés adjacents du triangle équilatéral si l'une des deux conditions suivantes est vérifiée :

$$||\theta_i - \theta_j| - \pi - \gamma_0| \leq \Delta\gamma \quad \text{ou} \quad ||\theta_i - \theta_j| - \pi + \gamma_0| \leq \Delta\gamma \quad (5.20)$$

Le choix de $\Delta\gamma$ détermine la précision avec laquelle l'égalité $|\theta_i - \theta_j| - \pi = \gamma_0$ ou $-\gamma_0$ doit être vérifiée numériquement : ces deux cas sont illustrés sur la Fig 5.9.

De plus, il ne faut pas être trop sévère sur le choix de $\Delta\gamma$. En effet, la projection d'un panneau de signalisation dans le plan image dépend de son orientation dans le repère caméra : son image

5.3. ÉVALUATIONS EXPÉRIMENTALES

est en toute généralité un triangle quelconque, mais l'intervalle $\gamma \in]\frac{\pi}{6}, \frac{\pi}{2}[$, correspondant à $\Delta\gamma = \pi/6$, regroupe un large spectre de points de vue habituels de la caméra. On peut envisager une précision plus fine si on a une connaissance *a priori* des angles à détecter. Nous avons ainsi testé les cas $\Delta\gamma = \pi/6$, $\Delta\gamma = \pi/12$ et $\Delta\gamma = \pi/18$. Les meilleurs résultats sont obtenus pour une précision intermédiaire $\Delta\gamma = \pi/12$ comme illustré par la figure 5.15.

Afin de modéliser plus spécifiquement l'angle $\gamma = \pi/3$, la fonction de phase est conçue pour n'apparier que les points appartenant aux côtés adjacents d'un angle de $\gamma_0 = \pi/3$:

$$\Theta(i, j) = \Theta_{//}(i, j) \left(\underbrace{W_{\frac{2\pi}{N}}(|\theta_i - \theta_j| - \pi - \gamma)}_{\text{angle aigu}} + \underbrace{W_{\frac{2\pi}{N}}(|\theta_i - \theta_j| - \pi + \gamma)}_{\text{angle obtus}} \right) \quad (5.21)$$

La largeur du noyau $W_{\Delta\gamma}$ est la précision $\Delta\gamma$. On la calcule à partir du choix d'une constante N qui est le nombre de bissectrices considérées : $\Delta\gamma = \frac{2\pi}{N}$.

Comme expliqué précédemment, on calcule deux accumulateurs, celui des sommets S_{accu} et celui des centres B_{accu} . Notre stratégie pour détecter un triangle est de commencer par détecter les centres de triangles en utilisant B_{accu} , puis de regrouper les sommets correspondant par triplet en exploitant S_{accu} . Un centre est détecté en recherchant un maximum local de l'accumulateur B_{accu} supérieur à un certain seuil τ_B . Pour chaque détection, on cherche dans S_{accu} les trois maxima locaux les plus proches supérieurs à un seuil τ_S . Si un point de l'accumulateur B_{accu} est associé à trois sommets, ils sont définis respectivement comme centre et sommets d'un triangle.

5.3 Évaluations expérimentales

5.3.1 Les bases de test

Les performances des algorithmes et les variantes sont évaluées sur 3 bases d'images différentes, dans lesquelles les panneaux pris en compte sont ceux dont la taille est comprise entre 32×32 et 70×70 :

- Base *traffic sign image database* [8]. Elle est constituée de 48 images de taille 360×270 pixels contenant 37 panneaux bleus (passage piétons, vélos) et 18 panneaux rouges (intersection, céder le passage). Parmi ces 55 panneaux routiers, 15 sont circulaires et 40 contiennent un triangle. Ces derniers se divisent en trois catégories illustrées par la figure 5.8 :
 - passage piétons **C20** : un triangle blanc dans un panneau rectangulaire bleu,
 - avertissement d'intersection : un triangle blanc, contenant un pictogramme, bordé par un triangle rouge, dont un sommet est situé au dessus de la base horizontale,
 - avertissement céder le passage : un triangle blanc bordé par un triangle rouge avec un sommet situé en dessous de la base horizontale.
- Base *Paris 15* : 89 images, de taille 640×480 , prises dans les rues de Paris 15° - cf. Fig. 5.11, contenant 81 panneaux circulaires rouges et un total de 92 panneaux de signalisation (11 bleus ou jaune, rectangulaires ou circulaires).
- Base *Stéréopolis* : 847 images de dimensions 960×540 contenant 9 panneaux triangulaires et 21 panneaux circulaires ou rectangulaires.

| Méthode | Nombre de cibles | Vrais Positifs | Faux Positifs | Temps de calcul |
|----------------|------------------|----------------|---------------|-----------------|
| TCB Bilatérale | 92 | 79 (86%) | 25 | 30 ms/img |
| TC Standard | 92 | 69 (75%) | 24 | 30 ms/img |

TAB. 5.1 – Performances en détection de panneaux sur la base *Paris 15* de 89 images 640×480 en milieu urbain, contenant 92 panneaux dont 81 rouges et 11 bleus ou jaunes.

Les performances de la TCB (panneaux circulaires et rectangulaires) ont été mesurées sur la base *Paris 15* et la base *Stéréopolis*. Les performances de la TSB ont été estimées sur la base *traffic sign image database* et la base *Stéréopolis*.

5.3.2 Performances en détection de panneaux circulaires et rectangulaires

Les algorithmes de la TC et de la TCB sont testés avec les mêmes étapes de pré-traitement pour l'estimation du champ de gradients d'une image. La comparaison est réalisée sur la base *Paris 15* : la figure 5.11 illustre quelques exemples de détections correctes et de faux positifs pour la TCB. Comparé à la TCB, l'algorithme de la TC est moins efficace comme le montrent les courbes de la figure 5.13. Les courbes COR de la figure 5.13(a) montrent le Taux de Détections Correctes TDC en fonction du Taux de Fausses Alarmes (TFA) pour chaque algorithme. La figure 5.13(b) montre le coefficient nommé Dice pour la TC et la TCB en fonction du seuil sur l'accumulateur avec :

$$\text{TDC} = \frac{\text{VP}}{\text{P}} \quad \text{TFA} = \frac{\text{FP}}{\text{NbImg}} \quad \text{Dice} = \frac{2\text{VP}}{\text{VP} + \text{FP} + \text{P}}$$

où VP est le nombre de vrais positifs, FP le nombre de faux positifs, P est le nombre total de panneaux dans la base de test, et NbImg le nombre total d'images. Le Tab. 5.1 indique les performances de ces deux algorithmes pour un point donné des courbes COR : la TCB est capable de détecter 79 panneaux sur 92 avec 25 faux positifs, tandis que la TC standard n'en détecte que 69 pour 24 faux positifs. En effet, comme on traite le canal rouge normalisé, les panneaux rouges ont de grandes chances d'avoir un contraste clair/sombre et auquel cas ils peuvent être détectés par la TC et la TCB. Les panneaux ayant une autre couleur, en particulier les signalisations bleues, ont plus de chances d'apparaître avec un contraste sombre/clair : la plupart d'entre eux sont détectés par la TCB mais pas par la TC standard. Le coefficient Dice atteint une valeur maximale de 82% pour la TCB contre 77% seulement pour la TC. Avec une différence de 5%, la Transformation Chinoise Bilatérale améliore significativement la Transformation Chinoise standard.

Enfin, la figure 5.13 compare aussi les performances des TCB signée et non signée. Le coefficient Dice atteint une valeur maximale de 85% pour la TCB non signée contre 82% pour la TCB signée. La supériorité des performances de la TCB non signée sur la TCB signée est aussi vérifiée sur la base *Stéréopolis*.

Remarquons que les images de la base *Stéréopolis* sont de dimensions 1920×1080 et des panneaux y apparaissent avec une grande taille. Afin d'accélérer les calculs, nous réduisons

5.3. ÉVALUATIONS EXPÉRIMENTALES

| Images | Panneaux (triangles) | Vrais Positifs | Faux Positifs | Cadence |
|--------|-------------------------|-------------------|------------------|---------------|
| 48 | 40 | 33 (82,5%) | 2 | <50 ms/img |

TAB. 5.2 – Performances en détection de panneaux de signalisation routière triangulaires sur la base *traffic sign database* de 48 images 360×270 . Les seuils τ_B et τ_S sur B_{accu} et S_{accu} sont déterminés expérimentalement sur une autre base d’images : $\tau_B = 1000$, $\tau_S = 20$.

| Images | Panneaux (triangles) | Vrais Positifs | Faux Positifs | Seuil τ_B |
|--------|-------------------------|-------------------|------------------|----------------|
| 847 | 9 | 7 (77,7%) | 0 | 1000 |
| 847 | 9 | 8 (88,8%) | 29 | 500 |

TAB. 5.3 – Performances en détection de panneaux de signalisation routière triangulaires sur la base *Stéréopolis* de 847 images 960×540 contenant 9 panneaux triangulaires. Deux seuils τ_B sur B_{accu} sont proposés, mais $\tau_S = 20$ dans les deux cas.

chaque dimension de l’image d’un facteur 2 : les images traitées ont alors pour dimensions 960×540 . De plus les panneaux ne sont cherchés que sur le côté droit de l’image. Enfin, on se limite aux panneaux dans l’intervalle $[R_{min}, R_{max}] = [32, 70]$ dans l’image réduite. En particulier, ils ne sont pas détectés lorsqu’ils sont trop lointain (trop petit) ou trop proche (trop grand et généralement tronqués) du capteur. Pour pallier à cela, les images en pleine résolution à différentes échelles sont traitées lors de la comparaison entre les différentes approches de détection du chapitre 6.

La figure 5.14 montre que les résultats sont améliorés si le module du gradient privilégie ceux rouges et ceux bleus. Ainsi, quand on traite une combinaison des gradients des canaux rouge et bleu normalisés, on augmente le contraste des panneaux bleus par rapport à ceux dans le canal du rouge normalisé seul. Les meilleures performances sont obtenues en combinant le rouge et le bleu, et en détectant les symétries par Transformée Bilatérale non signée.

5.3.3 Détection de panneaux de signalisation routière triangulaires

Notre algorithme a été évalué en premier sur la base *traffic sign image database* [8]. L’algorithme est capable de détecter correctement 33 panneaux sur 40 avec seulement 2 faux positifs en un temps inférieur à 50 ms/image, comme le résume Tab. 5.2.

Quelques exemples sont donnés dans la figure 5.12 : ils illustrent le cas d’une détection des sommets par seuillage de S_{accu} (sans utilisation de B_{accu}). Il est intéressant de voir que les sommets intérieurs et extérieurs des triangles, en bordure de panneaux rouges, sont correctement détectés. En effet, un panneau rouge contient un total de six sommets de triangles, tandis que les panneaux bleus contiennent seulement trois sommets.

Les tests sur la base *Stéréopolis* montrent une sensibilité du résultat de la TSB pas trop importante par rapport à la valeur du seuil τ_B . Le tableau 5.3 donne les résultats obtenus pour deux valeurs de τ_B . Les meilleures performances rapportées dans les Tab. 5.2 et 5.3 correspondent à une incertitude $\Delta\gamma = 2\pi/N = \pi/12$. La figure 5.15 compare les performances obtenues par la TSB pour différentes incertitudes. Les courbes COR de la figure 5.15(a) donnent pour ces méthodes le graphique du Taux de Détections Correctes TDC en fonction du Taux de Fausses Alarmes (TFA). La figure 5.15(b) représente le coefficient Dice pour la TC et la TCB en fonction du seuil sur l'accumulateur B_{accu} avec $\tau_S = 20$.

Pour $N = 12$, la gamme d'angles détectables est $]\pi/6, \pi/2[$. Pour $N = 24$, cette gamme est $]\pi/4, 5\pi/12[$, et pour $N = 36$, elle est de $]\pi/18, 7\pi/18[$, avec $\gamma_0 = \pi/3$. La TSB avec $N=24$ est plus sélective que dans le cas $N=12$, car l'incertitude est plus faible. Mais lorsque N augmente, les performances baissent parce que l'incertitude $\Delta\gamma$ devient inférieure aux incertitudes dues aux bruits et aux diverses perturbations (discrétisation,...). Les performances de la TSB avec $N = 12$ dépassent largement celles des deux autres cas avec un coefficient Dice maximal à 87,5% contre 61,5% pour les deux autres.

5.4 Conclusion

Nous avons présenté une méthode de détection des panneaux circulaires et rectangulaires utilisant un détecteur de symétrie efficace, la Transformation Chinoise Bilatérale. Elle peut traiter une image 640×480 en 30 ms avec un taux de détection élevé : 86% pour 25 faux positifs sur un ensemble de 89 images contenant 92 panneaux de différentes couleurs et formes (à l'exception des panneaux triangulaires). Elle améliore les résultats de la Transformation Chinoise standard qui ne détecte que 75% des panneaux pour 24 faux positifs sur cette base de test. De plus, la TCB est plus facile à généraliser que la TSR car elle peut détecter de la même manière cercles, carrés, rectangles, losanges et polygones. Elle est aussi plus précise que la TSG car elle est spécialisée sur les paires de points ayant des gradients symétriques par rapport à leur milieu. Cette méthode a fait l'objet d'une publication dans les actes de la conférence ISVC 2009 [4].

Pour traiter le cas des panneaux triangulaires, nous avons introduit une nouvelle transformation pour la détection du sommet d'un angle et de sa bissectrice simultanément, même en cas d'occultations partielles. Elle peut traiter efficacement une image 360×270 en moins de 50 ms pour détecter des panneaux de signalisation triangulaires. Elle atteint un taux de détection élevé avec 33 panneaux sur 40 correctement détectés pour 2 faux positifs, sur une base de 48 images 360×270 , et 7 panneaux sur 9 sur une base de 847 images 960×540 pour 0 faux positifs. Cette approche a été présentée aux conférences WACV 2009 [3] et RFIA 2010 [5].

5.4. CONCLUSION

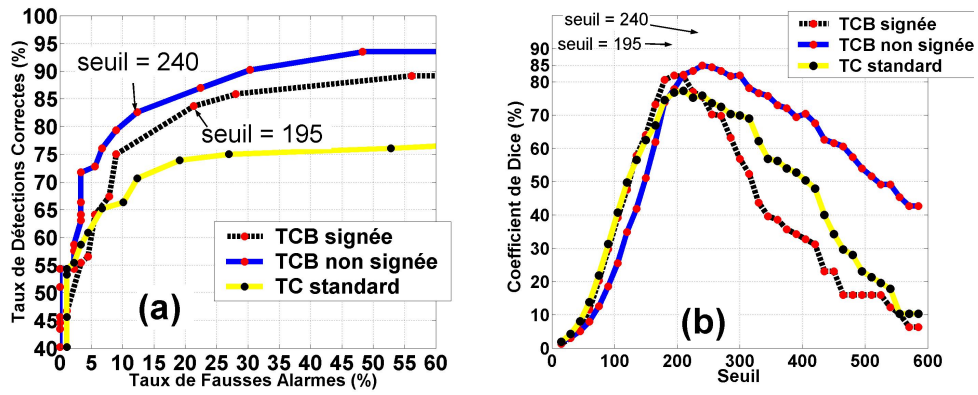


FIG. 5.13 – Comparaison des performances de la TC Bilatérale (noir) et de la TC standard (jaune), sur une base de test de 89 images contenant 92 panneaux : (a) Courbes COR et (b) Coefficient Dice (%) en fonction du seuil sur l'accumulateur.

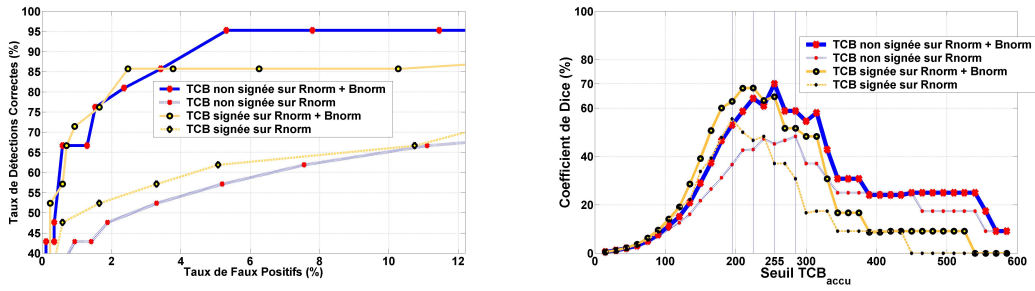


FIG. 5.14 – Courbes COR et Coefficient Dice de la TC Bilatérale sur la base *Stéréopolis*. Plusieurs variantes sont représentées : TCB signée et non signée, module du gradient estimé sur le canal rouge normalisé (Rnorm) ou calculé par la somme de celui du canal rouge normalisé et de celui du canal bleu normalisé (Rnorm+Bnorm).

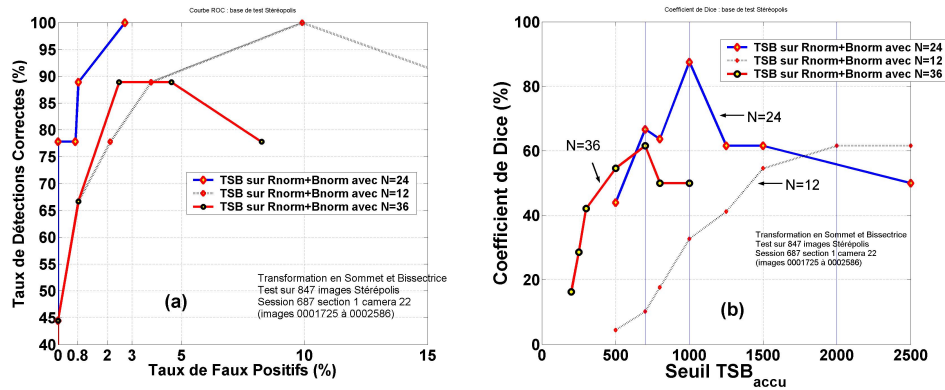


FIG. 5.15 – Comparaison des performances de la TSB pour différentes valeurs de la précision $\Delta\gamma$: $\pi/6$ ($N = 12$), $\pi/12$ ($N = 24$) et $\pi/18$ ($N = 36$) : (a) Courbes COR et (b) Coefficient Dice (%) en fonction du seuil sur l'accumulateur B_{accu} . La base de test est la base *Stéréopolis* ayant 847 images 960×540 contenant 9 panneaux triangulaires. Le seuil sur l'accumulateur S_{accu} est toujours $\tau_S = 20$ et le seuil τ_B est utilisé pour tracer les courbes COR.

Bibliographie

- [1] N. BARNES, A. ZELINSKY, et L. FLETCHER. Traffic sign recognition and analysis for intelligent vehicles. *Image and Vision Computing*, 21(3) :247–258, 2003.
- [2] R. BELAROUSSI et M. MILGRAM. A real time fingers detection by symmetry transform using a two cameras system. Dans *Proceedings of the 4th International Symposium on Visual Computing ISVC'08*, pages 703–712, Las Vegas, USA, décembre 2008.
- [3] R. BELAROUSSI et J.-P. TAREL. Angle vertex and bisector geometric model for triangular road sign detection. Dans *Proceedings of IEEE Workshop on Applications of Computer Vision WACV'09*, pages 577–583, Snowbird, USA, décembre 2009.
- [4] R. BELAROUSSI et J.-P. TAREL. A real-time road sign detection using bilateral chinese transform. Dans *5th International Symposium on Visual Computing ISVC'09*, pages 1161–1170, Las Vegas, USA, décembre 2009.
- [5] R. BELAROUSSI et J.-P. TAREL. Modèle géométrique du sommet et de la bissectrice d'un angle pour la détection de panneaux triangulaires. Dans *Actes du 17ème congrès francophone en Reconnaissance des Formes et Intelligence Artificielle, RFIA'10*, pages 757–764, Caen, France, décembre 2010.
- [6] I. CHOI et S.-I. CHIEN. A generalized symmetry transform with selective attention capability for specific corner angles. *Signal Processing Letters*, 11(2) :255–257, 2004.
- [7] P. FOUCHER, P. CHARBONNIER, et H. KEBBOUS. Evaluation of a road sign pre-detection system by image analysis. Dans *International Conference on Computer Vision Theory and Applications VISAPP'09*, pages 362–367, Lisbonne, Portugal, février 2009.
- [8] C. GRIGORESCU et N. PETKOV. Distance sets for shape filters and shape recognition. *IEEE Transactions on Image Processing*, 12(10) :1274–1286, 2003.
- [9] G. LOY et N. BARNES. Fast shape-based road sign detection for a driver assistance system. Dans *Proceedings of International Conference on Intelligent Robots and Systems IROS'04*, pages 70–75, Sendai, Japon, septembre 2004. IEEE/RSJ.
- [10] M. MILGRAM, R. BELAROUSSI, et L. PREVOST. Multi-stage combination of geometric and colorimetric detectors for eyes localization. Dans *Proceedings of 13th International Conference Image Analysis and Processing ICIAP'05*, pages 1010–1017, 2005.
- [11] C.-J. PARK, K.-S. SEOB, et H.-M. CHOIB. Symmetric polarity in generalized symmetry transformation. *Pattern Recognition Letters*, 27(7) :854–857, 2006.
- [12] D. REISFELD, H. WOLFSON, et Y. YESHURUN. Context free attentional operators : the generalized symmetry transform. *International Journal of Computer Vision*, 14(2) :119–130, 1995.

Chapitre 6

Etude comparative de trois méthodes

*Rachid Belaroussi,
Jean-Philippe Tarel,
Philippe Foucher,
Pierre Charbonnier,
Bahman Soheilian,
Nicolas Paparoditis*

Ce chapitre propose une comparaison des performances des trois algorithmes présentés dans les chapitres 3, 4 et 5 à partir d'une même base d'images de scènes routières en milieu urbain. Cette base d'images a été collectée dans le cadre du projet iTOWNS et peut-être téléchargée sur l'url [http ://www.itowns.fr/benchmarking.html](http://www.itowns.fr/benchmarking.html). La comparaison des trois algorithmes porte sur un échantillon de cette base. Cette sous-base sera décrite dans la section 6.1. Les résultats de la comparaison seront présentés et analysés dans la partie 6.2 avant la conclusion 6.3.

6.1 Base de données

Les séquences d'images ont été capturées par le véhicule Stéréopolis de l'IGN, en circulation, équipé de douze caméras très haute résolution permettant d'avoir une vision panoramique (360°) de la scène urbaine. La distance entre deux acquisitions est de 4 mètres. Chaque acquisition correspond à l'enregistrement de douze images de taille 1920×1080 dans le spectre « visible ». Dans le projet iTOWNS, un des objectifs est de rechercher des amers visuels dans la scène urbaine. La signalisation verticale est un élément caractéristique d'une voie de circulation, elle peut être utilisée comme vecteur d'entrée pour une recherche, ou pour aider à la navigation.

La base de données d'images issues du projet iTOWNS contient plus de 11000 panoramiques du 12^{ème} arrondissement de Paris. L'échantillon utilisé pour la comparaison correspond à une séquence de 847 images provenant de la caméra frontale (Caméra 22 du véhicule Stéréopolis). Une vérité terrain a été établie sur ce corpus d'images dans laquelle on distingue trois catégories de panneaux : Panneaux circulaires ou rectangulaires, panneaux de type C20 et panneaux triangulaires.

6.2 Comparaison des détecteurs de panneaux

6.2.1 Protocole d'évaluation

| Nombre d'images | Nombre de panneaux circulaires ou rectangulaires | Panneaux piétons C20 | Nombre de panneaux triangulaires |
|-----------------|--|-------------------------|--|
| 847 | 239 | 9 | 28 |

TAB. 6.1 – Dénombrement de la vérité terrain (Stéréopolis : caméra 22) : les panneaux recensés ont une taille minimale de 25×25 .

Les trois approches proposées peuvent être résumées de la manière suivante :

- Ajustement de contour **CF** (pour Contour Fitting) : les régions d'intérêts (ROI) obtenues par segmentation des pixels rouges sont traitées pour y détecter des triangles ou des cercles. Les formes rectangulaires et circulaires sont recherchées dans les ROIs obtenues par segmentation des pixels bleus (voir chapitre 4).
- Algorithme de vote univarié **SPV** (pour Single Pixel Voting) : deux variantes de la segmentation des pixels rouges et bleus sont utilisées. Une série de sélections par les caractéristiques de forme est opérée sur les ROIs obtenues. En particulier, une TSR analyse une composante couleur pour y détecter les cercles, une HT traite les ROIs rouges pour y trouver des triangles, et une autre HT détecte les rectangles sur les ROIs bleues (voir chapitre 3).
- Algorithme de vote bivarié **PWPV** (pour Pair-Wise Pixels Voting) : le gradient est calculé en sommant les normes des gradients du rouge et du bleu normalisés, la TCB détecte sur ces contours les cercles et les rectangles tandis que la TSB détecte les triangles (voir chapitre 5).

Nous avons comparé ces trois approches sur une même section de la base Stéréopolis, en utilisant les mêmes critères et la base vérité terrain précédemment décrite. La détection des panneaux de signalisation est réalisée seulement sur la moitié droite de chaque image. Une marge de 10 pixels sur le bord droit de l'image a été utilisée pour filtrer les panneaux sortant de la scène : lorsque le côté droit du rectangle englobant un panneau se situe dans cette marge, il est exclu de la vérité terrain. Un total de 251 panneaux de signalisation ont ainsi été relevés. On a divisé l'ensemble de ces panneaux suivant 4 catégories en fonction de leur taille : ≥ 16 pixels, ≥ 32 pixels, ≥ 48 pixels et ≥ 64 pixels. La figure 6.1 indique la population de chaque classe de panneaux : la taille d'un panneau est définie par le minimum entre la largeur et la hauteur de sa boîte englobante. On obtient 30 panneaux de taille supérieure à 64 pixels, 74 panneaux de taille supérieure 48 pixels, 173 panneaux de taille supérieure à 32 pixels et 251 panneaux de taille supérieure à 16 pixels.

6.2.2 Métriques d'évaluation

Notons (x_c^*, y_c^*) le centre du rectangle englobant un panneau de la vérité terrain, et $w^* \times h^*$ sa taille. Soit (x_c, y_c) le centre du rectangle englobant une détection, et $w \times h$ sa taille. On considère

6.2. COMPARAISON DES DÉTECTEURS DE PANNEAUX

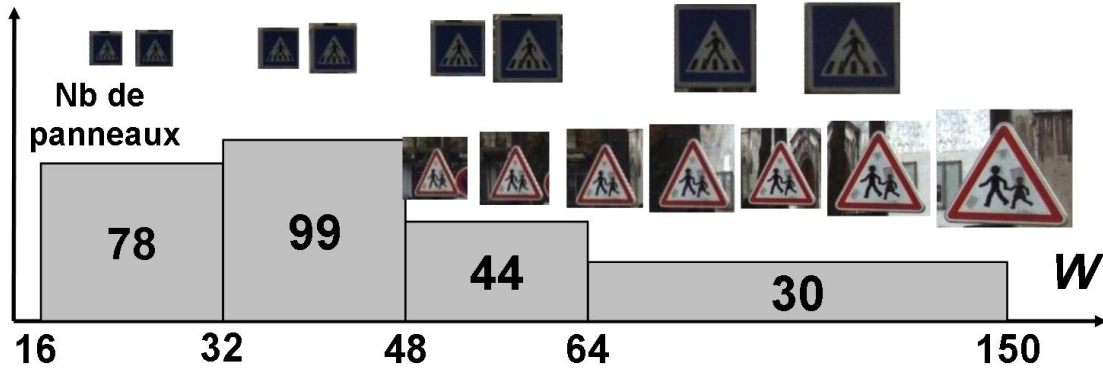


FIG. 6.1 – Distribution de la taille ($\min(\text{width}, \text{height})$) des 251 panneaux recensés.

que ce dernier est un vrai positif si son centre est situé à une distance inférieure ou égale à 20% de $\max(w^*, h^*)$ pixels du vrai centre et si ses dimensions correspondent aux vraies dimensions du panneau à $\pm 45\%$ près :

- précision sur le centre de la détection :

$$\sqrt{(x_c - x_c^*)^2 + (y_c - y_c^*)^2} \leq 0,2 \max(w^*, h^*) \quad (6.1)$$

- précision sur les dimensions du panneau :

$$|w - w^*| \leq 0,45 w^* \quad \text{et} \quad |h - h^*| \leq 0,45 h^* \quad (6.2)$$

Si une détection ne vérifie pas ces trois conditions pour aucun des panneaux de la base vérité terrain, elle est considérée comme une fausse détection, ou faux positif. Le taux des fausses alarmes est rapporté au nombre d'images. Il peut donc être supérieur à 1, possibilité également donnée dans [3]. Un panneau manqué est un panneau de la vérité terrain qui n'a pas été détecté.

Rappelons qu'une courbe COR (Caractéristiques Opérationnelles du Récepteur) affiche le taux de détections correctes (TDC) en ordonnées et le taux de fausses alarmes (TFA) en abscisses, en fonction d'un paramètre lié au détecteur. Ces deux taux sont définis par :

$$\text{TDC}(\%) = \frac{\text{VP}}{N} \quad \text{TFA}(\%) = \frac{\text{FP}}{\text{NbImg}}$$

où VP est le nombre de vrais positifs, FP le nombre de faux positifs, N est le nombre total d'apparition de panneaux dans la base de test (par exemple, $P = 251$ pour les panneaux de taille supérieure à 16 pixels), et NbImg le nombre total d'images ($\text{NbImg} = 847$).

6.2.3 Performances de détection

En termes de rapidité, les performances des différents algorithmes sont relativement proches. Le temps moyen de traitement est de 2,3 secondes par image pour la méthode CF, de 1,5 secondes par image pour la méthode SPV, et de 0,8 secondes pour la méthode PWPV. Ces résultats ont été obtenus pour la détection de tous les types de panneaux de taille supérieure à 16 pixels. La détection réduite à des tailles plus grandes ne change quasiment pas les temps de calculs pour la SPV. Notons que cette méthode est programmée en *Matlab*®. Contrairement

| Taille Min. | N de panneaux | couleur+CF | | couleur+SPV | | PWPV | |
|-------------|---------------|------------|------|-------------|------|------|------|
| | | TCD | TFA | TCD | TFA | TDC | TFA |
| 64 | 30 | 97% | 0,32 | 97% | 0,43 | 100% | 0,18 |
| 48 | 74 | 87% | 0,83 | 86% | 0,76 | 92% | 22 |
| 32 | 173 | 82% | 4,7 | 83% | 2,8 | 88% | 11 |
| 16 | 251 | 72% | 11,5 | 71% | 2,7 | 73% | 11 |

TAB. 6.2 – TDC maximal pour les trois méthodes de détection des panneaux.

aux autres, elle peut donc facilement être accélérée. Lorsque la détection est réduite aux panneaux de taille supérieure à 64 pixels, la méthode CF est accélérée d'un facteur 2 et celle PWPV d'un facteur 4.

Les courbes COR de la figure 6.2 présentent les performances des trois algorithmes sur l'ensemble des panneaux de la base de test, triangles inclus. Les courbes COR des méthodes SPV et PWPV ont été construites en faisant varier le seuil portant sur la valeur contenue dans les accumulateurs. La courbe COR de la méthode CF a été construite en faisant varier le seuil sur la sélection des couleurs ce qui explique des courbes plus courtes. Les courbes COR obtenues sont relativement proches pour les trois méthodes, en particulier pour les panneaux de la catégorie ≥ 32 pixels entre SPV et PWPV - cf. Fig. 6.2(c) et pour les panneaux de taille supérieure à 48 pixels entre SPV et CF - cf. Fig. 6.2(b). Toutefois, dans le cas des panneaux de grande taille (≥ 64 pixels) - cf. Fig. 6.2(a), la méthode PWPV est légèrement meilleure que celle CF et SPV, en atteignant un taux de réussite de 100% pour une moyenne de 0,2 faux positif par image. Cela peut s'expliquer par des cas de sur-segmentation colorimétrique, divisant un panneau en plusieurs régions, en particulier dans le cas des panneaux barrés (fin d'interdiction, fin de voies de bus, stationnement). A l'opposé, pour les panneaux de petite taille ≥ 16 pixels (Fig. 6.2(d)), la méthode SPV fournit clairement la meilleure courbe COR.

En comparant les courbes COR, on évalue les performances intrinsèques de chaque détecteur. Sachant que l'étape de détection peut aussi être suivie d'une étape de reconnaissance ou de filtrage temporel, capables de filtrer encore plus de faux positifs, il est intéressant de comparer les taux maximaux de détections correctes que peut atteindre chaque approche, à des taux raisonnables de faux positifs. Les TDC maximaux sont indiqués dans le Tab. 6.2 avec le taux correspondant de FP/image, pour chacune des quatre catégories de taille et pour chaque méthode. Il apparaît que le TDC maximal décroît à mesure que la taille des panneaux considérés diminue, pour les trois approches. C'est une conséquence de l'utilisation d'un modèle géométrique. Pour la méthode CF, cela est dû au fait que les panneaux petits ont une forme moins précisément définie. Pour la méthode PWPV, quand les panneaux sont trop petits, le nombre de pixels votant est réduit et devient trop faible pour être correctement discriminé du fond de l'image. Par contre, pour l'algorithme SPV, cela semble plutôt causé par la segmentation couleur sans que nous sachions à présent l'expliquer. Le TDC maximal obtenu par l'algorithme PWPV est, quelle que soit la taille des panneaux, supérieur à celui obtenu par l'approche CF et SPV. Ces deux approches sont, de nouveau, limitées par les performances de la segmentation en composantes connexes qui est sujette à la sur-segmentation et à la sous-segmentation. Toutefois, la PWPV atteint des taux de TDC supérieurs pour les tailles supérieures à 16, 32 et 48 pixels à des taux de fausses alarmes de l'ordre de 10 FP/image, largement au-dessus de ceux de la SPV.

6.3. COMPARAISON DES DÉTECTEURS DE PANNEAUX

| Taille | Triangles | | CF | | TH | | TSB | |
|--------|-----------|------|-----------|--------------|-----------|----------|------------|----------|
| | Rouge | Bleu | VP | FP | VP | FP | VP | FP |
| 64 | 8 | 2 | 8 80% | 249 29% | 8 80% | 19 2% | 10 100% | 7 1% |
| 48 | 15 | 4 | 14 74% | 434 51% | 13 68% | 21 2% | 14 74% | 13 2% |
| 32 | 22 | 7 | 15 52% | 802 94% | 16 55% | 21 2% | 15 52% | 32 4% |
| 16 | 25 | 9 | 15 44% | 1546 183% | 16 47% | 21 2% | 15 44% | 37 4% |

TAB. 6.3 – Performances de détection pour les panneaux triangulaires rouge ou contenant un triangle blanc sur fond bleu.

Ainsi, sur l'ensemble des panneaux de taille ≥ 16 , la SPV obtient pratiquement le même taux de réussite que la PWPV et CF pour un nombre de fausses alarmes cinq fois moindre. C'est une indication de l'intérêt de l'utilisation de la couleur pour les petits panneaux. Il semble donc intéressant de renforcer l'utilisation de la couleur dans la méthode PWPV. Pour les panneaux de taille supérieure à 16 pixels, le TDC maximal est de 73% donc 27% des panneaux ne sont pas détectés. Ce taux de panneaux manqués demeure élevé, et c'est un des principaux inconvénients observé sur les approches testées. Cela peut être amélioré en tenant compte de la cohérence temporelle de la séquence : les panneaux détectés à une grande taille peuvent être recherchés plus efficacement dans les images précédentes où ils apparaissent avec une taille plus petite.

6.2.4 Cas des panneaux triangulaires

Les trois algorithmes évalués proposent une détection spécifique pour les triangles. Il est donc intéressant de comparer les performances obtenues sur ce type de panneaux, d'autant que ces derniers sont assez peu traités dans la littérature. Les résultats en détection de panneaux triangulaires sont indiqués dans le Tab. 6.3. Les panneaux triangulaires sont pour la plupart rouge et blanc (Danger : école, ralentisseur, piétons) mais il existe aussi des panneaux rectangulaires bleus contenant un triangle blanc avec un pictogramme tels que ceux des passages pour piétons. La figure 6.1 illustre ces deux types de panneaux. Contrairement à la TSB, les algorithmes CF et SPV ne cherchent pas à détecter ce type de triangle blanc sur fond bleu. Ils sont détectés comme des panneaux rectangulaires bleus mais n'apparaissent donc pas dans le Tab. 6.3. En détectant ce type de panneaux de deux façons (rectangle bleu et triangle blanc), on augmente la redondance. Si cela nécessite l'utilisation d'une étape de fusion, cela peut permettre de rater moins de panneaux en particulier lorsque les contours rectangulaires du panneau ne sont pas très contrastés ou perturbés. Cette idée pourrait être reprise par les algorithmes CF et SPV.

En termes de TDC, les performances indiquées dans le Tab. 6.3 sont à peu près équivalentes, la principale différence étant le taux de fausses alarmes. On peut voir que la SPV donne le moins de faux positifs pour les différentes échelles. Cela s'explique par la nécessité pour la TSB d'avoir plus de votants que la SPV.

6.3 Conclusion

Trois algorithmes fondés sur des approches assez différentes ont été présentés et évalués indépendamment dans les chapitres précédents. Dans ce chapitre, nous les avons comparés de manière objective, dans un cadre unifié. La base de références et les critères utilisés sont assez stricts. En particulier, nous avons choisi une évaluation en termes d'image ou chaque apparition d'un panneau dans une nouvelle image doit être détectée. Une autre possibilité aurait été de faire une évaluation en termes de panneau en considérant que le panneau est détecté s'il est détecté dans au moins une image comme dans [2]. Cela augmente le taux de détections correctes, mais ne permet plus l'analyse par taille de panneaux.

La comparaison a été réalisée en termes de performances intrinsèques de détection, dans un compromis entre le taux de détection et le nombre de fausses alarmes. Elle a aussi été réalisée en considérant que la détection est suivie d'une étape de reconnaissance capable d'éliminer de nombreuses fausses détections. Dans ce contexte, c'est le taux de détection maximum avec un nombre raisonnable de fausses détections qui devient important.

L'algorithme avec ajustement d'un modèle géométrique (CF) et celui avec une transformation par votes univariés (SPV) utilisent en pré-traitement une segmentation des pixels rouges et une segmentation des pixels bleus. Si le fait de considérer la couleur s'est avéré être important, il apparaît que la segmentation par agglomération de pixels connexes est le point faible de ce pré-traitement. La segmentation limite le taux de détections correctes à cause du problème de sur-segmentation (région brisée en morceaux) ou de sous-segmentation (région trop agglomérée), même si cette étape permet une accélération drastique du traitement. Il serait bien de l'améliorer pour tenir compte de ces problèmes ou de trouver une alternative. En comparant les résultats obtenus par vote univarié à ceux de la méthode bivariée, il apparaît aussi que la couleur est un critère de sélection particulièrement efficace pour la détection des petits panneaux. Il serait donc profitable pour l'approche bivariée, d'avoir une étape de recherche des contours plus sélective que la somme des gradients du rouge et bleu normalisés.

L'algorithme avec ajustement d'un modèle géométrique, tout comme celui avec les transformations univariées, est capable de séparer les panneaux en trois classes : triangulaires, circulaires et rectangulaires. En ajoutant un classifieur entre formes circulaires et rectangulaires, l'algorithme fondé sur les transformations bivariées devrait être capable de faire de même.

Pour les panneaux de grande taille, l'algorithme PWPV fondé sur des transformations bivariées a obtenu la meilleure performance avec un taux de bonnes détections de 100% (0 panneau raté sur 30). L'algorithme SPV fondé sur les transformations univariées a obtenu une très bonne performance avec un taux de bonnes détections de 97% (1 panneau raté sur 30). L'algorithme avec ajustement d'un modèle géométrique (CF) obtient des performances aussi bonnes avec aussi un taux de bonnes détections de 97% (1 panneau raté sur 30). Notons que ces performances sont obtenues sur un ensemble réduit de panneaux et elles se dégradent pour des tailles de panneaux plus petites. L'algorithme fondé sur l'utilisation de transformations univariées semble moins sujet à cette dégradation lorsque l'on regarde les courbes COR obtenues. La Figure 6.3 montre les panneaux qui sont ratés pour les panneaux de grande taille. On remarque que le panneau raté par la méthode SPV est aussi raté par la méthode CF, à cause du même problème de segmentation dû à la présence d'un fond rouge autour du panneau. Si les performances en termes de bonnes détections maximum sont assez semblables, il y a des différences en nombre de fausses alarmes. Il paraît donc intéressant de faire collaborer ces trois détecteurs,

6.3. CONCLUSION

ce qui permettrait d'obtenir un détecteur qui atteindrait un très bon taux de bonnes détections avec probablement moins de fausses alarmes.

Le problème de baisse des performances de détection avec la taille peut être traité en utilisant la cohérence temporelle des séquences d'images. En effet, après la détection d'un panneau de grande taille, il est possible d'aller chercher dans les images précédentes, dans une zone réduite, la présence de ce même panneau avec un seuil de décision plus faible. Une autre possibilité est d'utiliser les données laser pour tester la cohérence 3D de la zone détectée comme un panneau (zone plane de taille et position 3D compatible avec un panneau). En ajoutant ce filtrage, il doit être possible de diminuer les seuils de détection sans ajouter trop de faux alarmes et donc d'améliorer les performances des détecteurs. Cette étude comparative des trois méthodes a fait l'objet d'une présentation à la conférence ICPR 2010 [1].

6. ÉTUDE COMPARATIVE

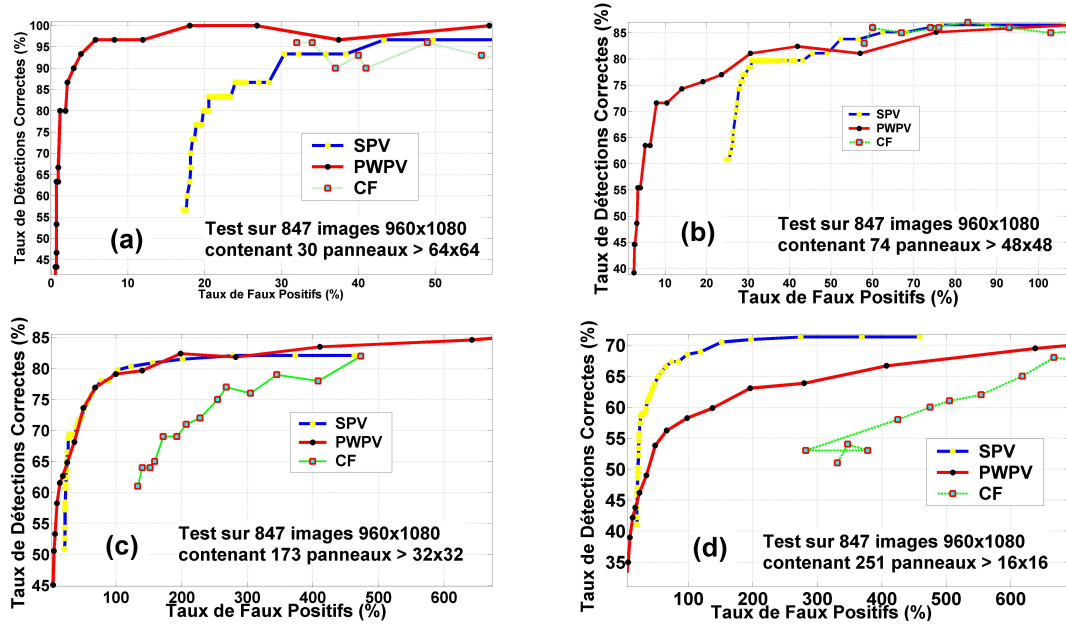


FIG. 6.2 – Courbes COR pour 4 tailles de panneaux : (a) taille ≥ 64 pixels ; (b) taille ≥ 48 pixels ; (c) taille ≥ 32 pixels ; (d) taille ≥ 16 pixels.



FIG. 6.3 – Pour la taille ≥ 64 , le panneau B6a1 est raté par la méthode CF et SPV à cause d'un problème de segmentation couleur lié au fond rouge en arrière plan.

Bibliographie

- [1] R. BELAROUSSI, P. FOUCHER, J.-P. TAREL, B. SOHEILIAN, P. CHARBONNIER, et N. PAPARODITIS. Road sign detection in images : a case study. Dans *Proceeding of the 20th International conference on Pattern Recognition (ICPR 2010)*, Istanbul, Turquie, août 2010. A paraître.
- [2] P. FOUCHER, P. CHARBONNIER, et H. KEBBOUS. Evaluation of a road sign pre-detection system by image analysis. Dans *Proceedings of International Conference on Computer Vision theory and Applications (VISAPP 2009)*, pages 362–367, Lisbonne, Portugal, février 2009.
- [3] S. MAJI et J. MALIK. Object detection using a max-margin hough transform. Dans *Proceedings of Computer Vision and Pattern Recognition (CVPR 2009)*, pages 1038–1045, Miami, USA, juin 2009.

6. ÉTUDE COMPARATIVE

Chapitre 7

Ajustement de forme prototype par méta-heuristiques biologiques

*Pierre Charbonnier,
Guillaume Dutilleux*

Nous présentons dans ce chapitre l'adaptation de techniques de reconnaissance de formes, les modèles déformables, au problème de l'analyse de la signalisation verticale. Les modèles déformables sont des outils désormais classiques en analyse d'images et dont l'essor date du début des années 1990. Il s'agit d'« objets » mathématiques que l'on vient ajuster sur les images en les déplaçant et en les déformant. La plupart du temps, l'évolution du modèle est pilotée par l'optimisation d'une fonctionnelle, d'inspiration mécanique ou dérivant d'une modélisation probabiliste. Cette fonctionnelle quantifie l'adéquation du modèle à l'image ainsi que son acceptabilité. Notons que la méthode présentée au chapitre 4 réalise également un ajustement de forme, sur des primitives extraites des images, mais celui-ci obéit à des critères purement géométriques, la validation de la forme étant effectuée *a posteriori* à partir du contenu image.

De nombreuses variantes de modèles déformables ont été proposées durant ces 20 dernières années. Certaines d'entre elles sont très génériques, permettant ainsi d'aborder des problématiques très variées. C'est le cas des contours actifs, techniques dans lesquelles les frontières des objets sont représentées par des courbes (en 2D) ou des surfaces (en imagerie 3D). Le lecteur intéressé pourra, par exemple, se référer à [2] pour un état de l'art sur le sujet. D'autres, au contraire, sont très spécifiques, limitées à la représentation de classes plus restreintes d'objets. Le niveau de contrainte plus important qu'elles imposent aux objets manipulés peuvent cependant les rendre plus robuste aux perturbations.

Nous nous intéressons ici à ce dernier type de modèle, et, plus particulièrement, aux formes prototypes [17]. Ces approches décrivent les formes à l'aide d'assemblages de primitives géométriques. Ces modèles sont donc bien adaptés à la représentation d'objets manufacturés. La modélisation que nous employons est inspirée des travaux de Dubuisson et Jain [5], dédiés à la reconnaissance de véhicules automobiles dans des images fixes prises depuis le bord de la chaussée. Elle débouche sur un problème d'optimisation difficile, caractérisé par la présence de multiples minima locaux. Dans ce contexte, l'utilisation d'algorithmes stochastiques, aptes à atteindre l'optimum global de la fonctionnelle, est particulièrement bien indiquée. Ainsi, un algorithme de type recuit simulé, utilisé seul dans [5], est associé avec succès dans [13] à un

7. FORMES PROTOTYPES ET MÉTA-HEURISTIQUES BIOLOGIQUES

algorithme local d'optimisation. Le recuit simulé est une métaphore d'un processus physique, utilisé en métallurgie pour figer le métal dans des configurations atomiques très stables, de faible énergie. Les algorithmes que nous avons mis en œuvre sont également des métaphores. Cependant, elles s'inspirent plutôt de mécanismes biologiques, intervenant dans l'évolution des espèces, les mécanismes immunitaires, ou le comportement grégaire de certains animaux : on parle de méta-heuristiques biologiques. En ce qui concerne l'ajustement de formes prototypes, ces techniques ont pour la première fois été employées dans les travaux de Mignotte et Collet [13], consacrés à la détection de mines sous-marines dans des images SONAR. Notons que les développements rapportés dans ce chapitre ont également été décrits plus en détails dans [8], où le lecteur pourra également trouver des informations sur de nombreuses méthodes alternatives d'optimisation.

Nous décrirons, dans un premier temps, la position du problème de détection de panneaux de signalisation de danger sous forme d'optimisation de fonctionnelle pour l'ajustement de forme prototype. Nous proposerons ensuite une présentation des algorithmes employés, en détaillant celui des essaims de particules, qui a montré les meilleures performances dans nos expérimentations.

7.1 Définition du problème d'estimation

7.1.1 Représentation de la forme : le prototype

Nous nous intéressons ici à la détection de la signalisation verticale de danger. En France, les panneaux de danger ont une forme de triangle équilatéral, arrondi aux trois extrémités. C'est également le cas de 3 des panneaux de signalisation d'intersection [11]. Ils se composent d'une couronne triangulaire de couleur rouge, bordée d'un listel blanc et encadrant un triangle blanc où sont tracés des pictogrammes. Ces derniers sont, à deux exceptions près, noirs (il existe 29 pictogrammes différents pour les panneaux de danger, et 3 pictogrammes pour les panneaux triangulaires de signalisation des intersections). La couronne triangulaire est, à la fois, le point commun de l'ensemble des panneaux de danger et leur principal élément distinctif dans les scènes routières. Le prototype que nous définissons est donc formé de deux triangles équilatéraux emboîtés concentriques, représentant les bords de la couronne rouge - cf. Fig. 7.2(a). Leur longueur de base est dans un rapport fixe de 0,78. Ce rapport est défini de manière réglementaire [11] et ne varie pas en fonction de la gamme de taille du panneau.

Comme nous le faisons remarquer en introduction, ce modèle est très spécialisé : pour prendre en compte les autres types de signalisation, il conviendrait de considérer des modèles elliptiques, ou d'autres formes de polygones. Notons que les panneaux triangulaires de « Cédez-le-passage » ont une couronne plus épaisse. Le modèle triangulaire doit donc également être adapté si l'on souhaite détecter cette signalisation.

7.1.2 Représentation des déformations

Le prototype est défini dans un repère qui lui est propre : il est inscrit dans un cercle de rayon arbitraire - cf. Fig. 7.2(a). Afin de l'ajuster aux objets d'intérêt dans les images, il est nécessaire de le déplacer et de le déformer. De nouvelles instances du prototype peuvent être

7.1. DÉFINITION DU PROBLÈME D'ESTIMATION

générées en appliquant une transformation affine aux coordonnées des points qui le composent, que l'on peut écrire sous la forme [13] :

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} 1 & h \\ 0 & 1 \end{pmatrix} \begin{pmatrix} s_1 & 0 \\ 0 & s_2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad (7.1)$$

Les paramètres de la transformation affine sont rassemblés dans un vecteur $\Theta = (\alpha, h, s_1, s_2, t_x, t_y)^T$. Chaque valeur de ce vecteur correspond donc à une instance déformée et déplacée du modèle - cf. Fig. 7.2(b).

Notons que cette façon de générer les déformations est indépendante de la forme du prototype : elle peut donc être employée quel que soit le type de signalisation considéré.

7.1.3 Fonctionnelle d'énergie

Nous nous plaçons, comme dans [5], dans le cadre de l'estimation bayésienne : la détection de panneau dans une image I est vue comme un problème d'estimation de paramètres au sens du *Maximum A Posteriori* (MAP). Le vecteur de paramètres optimal minimise une énergie :

$$\hat{\Theta} = \arg \min_{\Theta} \{ \mathcal{U}_d(I|\Theta) + \lambda \mathcal{U}_p(\Theta) \}. \quad (7.2)$$

Dans cette expression, \mathcal{U}_p est un *a priori*, qui reflète l'acceptabilité de la configuration Θ dans l'absolu (en dehors de toute observation) et \mathcal{U}_d mesure l'adéquation du modèle aux données. Le paramètre λ règle le compromis entre les deux termes du critère.

A priori

Le positionnement des panneaux de signalisation le long des voies de circulation obéit à des règles bien précises [11]. Leur taille appartient à une gamme également bien définie. Enfin, la géométrie de prise de vues est fixée et les panneaux sont observés de manière quasiment frontale. Il est donc possible de déterminer *a priori* un intervalle de valeurs admissibles des paramètres de position, taille et orientation des panneaux dans les images.

Pour mesurer l'adéquation d'une instance du prototype vis-à-vis d'un paramètre donné, on introduit une fonction de pénalisation en forme de vallée - cf. Fig. 7.2(c). Sa valeur est nulle dans l'intervalle de validité du paramètre, $[\theta_{\min}, \theta_{\max}]$, et augmente linéairement au-delà des limites de celui-ci. Les pentes des droites de pénalisation sont fixées empiriquement.

Finalement, l'énergie *a priori* \mathcal{U}_p est une somme pondérée des différents termes de pénalité définis chacun pour un paramètre d'intérêt.

Adéquation à l'image

Le prototype que nous avons choisi modélise la couronne rouge du panneau de signalisation. Une instance parfaitement adaptée à un panneau dans une image sera telle que la zone comprise entre les deux triangles contiendra exclusivement des pixels rouges, tandis que le triangle intérieur n'en contiendra pas (ou peu, dans le cas particulier des 2 panneaux à pictogramme en couleurs). Par ailleurs, les bords du prototype seront alignés sur les contours du panneau dans l'image. De façon à améliorer la robustesse de la segmentation, nous exploitons simultanément ces deux informations.

7. FORMES PROTOTYPES ET MÉTA-HEURISTIQUES BIOLOGIQUES

Utilisation de la couleur Nous avons proposé [7, 8] un critère simple permettant de classer les pixels des images en « rouge » ou « non-rouge ». Celui-ci examine la proportion de rouge dans le pixel. Ce critère, décrit plus en détails au chapitre 3, a l'avantage d'être insensible aux variations d'intensité lumineuse. Il nécessite *in fine* le réglage d'un seul paramètre. Une carte de couleur rouge est ainsi calculée pour chaque image avant la segmentation - cf. Fig. 7.3(a-b), où le blanc correspond à la classe « non rouge » et le noir, à la classe « rouge ».

Le terme région de la fonctionnelle \mathcal{U}_d est calculé par intégration numérique [7, 8]. Il dépend de la proportion \mathcal{U}_c de rouge à l'intérieur de la couronne du prototype, déformé et déplacé selon les valeurs courantes du vecteur de paramètres Θ . Il dépend également de la proportion \mathcal{U}_t de rouge dans le triangle intérieur du prototype. Plus précisément, ce terme région est défini par :

$$\mathcal{U}_{reg}(I|\Theta) = \min(1, 1 + \mathcal{U}_t(I|\Theta) - \mathcal{U}_c(I|\Theta)) \quad (7.3)$$

Ainsi, lorsque le prototype est parfaitement ajusté sur un panneau, $\mathcal{U}_{reg}(I|\Theta)$ vaut 0. Si le prototype se trouve sur une région rouge uniforme (un toit, par exemple), ou blanche (ciel, mur), l'énergie région vaut 1.

Utilisation des contours Le terme contour de la fonctionnelle \mathcal{U}_d est de la même forme que celui défini dans [5], qui intègre sur chaque arête du triangle le produit de la norme du gradient de l'image et d'une fonction du produit scalaire entre la direction du gradient et le vecteur directeur du segment. Ce produit est normalisé de façon à tendre vers 0 lorsque le prototype déformé est aligné sur les contours du panneau dans l'image, et vers 1 sinon. Toutefois, afin de rendre ce terme plus attractif, nous ne l'appliquons pas sur la norme du gradient, mais sur une carte de distance au point de contour le plus proche, calculée à l'aide de l'algorithme *Fast Marching* [16].

7.2 Optimisation par méta-heuristiques biologiques

La fonctionnelle d'énergie, où fonction coût, obtenue en sommant les termes d'*a priori* et d'adéquation aux données est non convexe et possède de nombreux minima locaux. De ce fait, un algorithme d'optimisation globale est nécessaire. Nous avons proposé dans [8] une comparaison expérimentale de trois algorithmes inspirés de mécanismes biologiques : la stratégie d'évolution (ES) [15], la sélection clonale (CS) [4] et l'essaim de particules (PSO) [3]. Cette étude a montré la supériorité de ce dernier algorithme dans le problème qui nous intéresse. Aussi, nous le décrivons ici de manière plus détaillée. Le lecteur intéressé pourra se référer à [8] pour une description plus complète de l'ES et de la CS, et à [6] pour un descriptif des algorithmes accompagné d'une implantation en langage Scilab.

7.2.1 La stratégie d'évolution : une métaphore darwinienne

Si l'on considère une seule espèce vivante dans son environnement et un problème d'optimisation, une analogie peut être établie entre l'ensemble des individus de l'espèce et l'ensemble des solutions, et entre la performance dans l'environnement et la valeur de la fonction coût. Les opérateurs de base pour la production de nouvelles solutions candidates sont inspirés de ceux

7.2. OPTIMISATION PAR MÉTA-HEURISTIQUES BIOLOGIQUES

qu'implique l'évolution d'une population : *croisement*, *mutation*. Dans le même esprit, un opérateur de *sélection* fait office de prédateur, en éliminant les solutions candidates de mauvaise qualité à la fin de chaque itération. Cette analogie a mené à la formulation d'algorithmes habituellement rangés dans la classe des « algorithmes évolutionnaires », dont les algorithmes génétiques (GA) constituent sans doute le représentant le plus connu [12]. Les stratégies d'évolution (ES) [1], font également partie de cette classe, mais sont directement adaptées à l'optimisation de variables continues.

7.2.2 La sélection clonale : une métaphore immunitaire

La sélection clonale est un mécanisme qui permet à un organisme vivant de faire face à de nouveaux pathogènes, et d'améliorer sa résistance face à un pathogène récurrent. L'affinité élevée d'un antigène avec un anticorps stimule la multiplication du lymphocyte B qui produit celui-ci. Les clones obtenus vont devenir soit des plasmocytes qui libèrent massivement des anticorps dans la circulation, soit des lymphocytes mémoire, qui tombent en dormance en vue d'une prochaine intrusion. Le processus de clonage n'est pas exempt d'erreurs, ce qui permet de diversifier les anticorps produits. On retrouve donc les ingrédients d'un algorithme d'optimisation : génération de solutions candidates, mesure d'affinité (évaluation de la fonction coût), sélection. Les anticorps de la population manipulée par l'algorithme introduit dans [4] représentent chacun un jeu de valeurs pour les paramètres à optimiser. A chaque itération, la population, composée en partie d'anticorps sélectionnés au fil des itérations précédentes et de nouveaux anticorps, produits par tirage aléatoire, fait l'objet d'une sélection selon la valeur de la fonction coût. Les meilleurs individus sont retenus et subissent alors une phase de reproduction par clonage, puis de maturation, qui introduit des variations dans les clones obtenus. La boucle est reprise à son début, jusqu'à ce que le critère de sortie soit atteint. Cet algorithme s'avère d'une efficacité remarquable sur une large gamme de problèmes, moyennant un réglage préalable de ses paramètres de contrôle [6].

7.2.3 L'essaim de particules : une métaphore sociale

L'harmonie qui se dégage des mouvements d'un banc de poissons ou d'un essaim d'oiseaux ne doit pas faire oublier que ce mode de déplacement constitue un avantage évolutif majeur pour les espèces qui le pratiquent, en matière de risque de prédation comme de recherche de nourriture. L'harmonie résulte d'une interaction permanente entre les membres du groupe. L'algorithme de l'essaim de particules [10, 3], proposé par un chercheur en psychologie sociale, s'en inspire. L'exploration de l'environnement par un essaim à la recherche de nourriture peut être vue comme un processus d'optimisation, faisant entrer en jeu des comportements individuels et des interactions sociales. La métaphore consiste à passer de l'espace physique à l'espace de paramètres de dimension arbitraire et de la fonction « nourriture disponible » à une fonction mathématique quelconque.

Dans l'algorithme PSO, les particules sont des vecteurs, au nombre de N , qui se déplacent sur l'espace de recherche, de dimension D , associé à la fonction à optimiser, \mathcal{U} . Dans notre cas, la position de chaque particule correspond à une instance du modèle déformable : nous la noterons donc Θ_i , pour $i = 1, \dots, N$. Chaque particule est dotée d'un vecteur de changement de

7. FORMES PROTOTYPES ET MÉTA-HEURISTIQUES BIOLOGIQUES

position, ou *vélocité*, noté v_i : le déplacement d'une particule d'une itération à l'autre est donné par : $\Theta_i \leftarrow \Theta_i + v_i$.

```

Algorithme Optimisation de la fonction  $\mathcal{U}(\times)$  par essaim de particules
Début
Initialiser la population
  Répéter
    Pour  $i$  variant de 1 à N Faire
      Si  $\mathcal{U}(\Theta_i) < \mathcal{U}(p_i)$  Alors
         $p_i \leftarrow \Theta_i$ 
      FinSi
       $g \leftarrow i$ 
      Pour  $j \in \text{Voisins}(i)$  Faire
        Si  $\mathcal{U}(\Theta_j) < \mathcal{U}(p_g)$  Alors
           $g \leftarrow j$ 
        FinSi
      FinPour
      Pour  $d$  variant de 1 à D Faire
         $v_{id} \leftarrow \chi(v_{id} + \varphi_1(p_{id} - \Theta_{id}) + \varphi_2(p_{gd} - \Theta_{id}))$ 
         $\Theta_{id} \leftarrow \Theta_{id} + v_{id}$ 
      FinPour
    FinPour
  Jusqu'à ce que le critère de fin soit satisfait
Fin

```

FIG. 7.1 – Pseudo-code de l'algorithme PSO

Par ailleurs, chaque individu garde la mémoire de la meilleure position qu'il ait jamais occupée, p_i . Il appartient à un voisinage de particules, dont elle connaît à chaque instant la mieux positionnée, p_g , c'est-à-dire celle qui minimise \mathcal{U} . L'idée centrale de la méthode est que le déplacement de chaque particule résulte d'une pondération de son expérience individuelle et des informations obtenues par communication avec ses voisines. En termes plus mathématiques, la variation de vélocité entre deux itérations est une combinaison linéaire (coordonnée par coordonnée) des vecteurs $\Theta_i - p_i$ et $\Theta_i - p_g$. Les coefficients de cette combinaison sont choisis par tirage aléatoire sur un intervalle $[0, \varphi_{\max}]$, où φ_{\max} est un paramètre de l'algorithme. Les particules ont donc tendance à osciller autour des meilleures positions p_i et p_g . Notons qu'il est possible de limiter l'amplitude des oscillations en bornant l'amplitude de v_i . Dans le même ordre d'idées, un coefficient global de *constriction* [3], χ , est appliqué à la vélocité de façon à contrôler la convergence de l'algorithme. Le pseudo-code de ce dernier est proposé Fig. 7.1.

7.3. CONCLUSION

Dans notre application, la taille de la population est fixée à $N = 20$ particules. Le voisinage est étendu à l'ensemble des particules de l'essaim. La valeur de ϕ_{\max} est fixée à 4,5 et celle du coefficient de constriction est obtenue en prenant la valeur usuelle $\kappa = 0,8$ dans le modèle proposé dans [3]. La sortie de l'algorithme intervient lorsqu'un nombre pré-défini d'itérations est atteint ou lorsque l'énergie du meilleur individu devient inférieure à un certain seuil.

On peut noter que l'algorithme PSO offre plusieurs avantages. Il est défini dans un formalisme continu. Il est de forme très simple, ne nécessite que peu de mémoire et de calculs. Enfin, il s'avère d'une grande efficacité dans de nombreuses situations [3]. Si la convergence de l'algorithme utilisant un coefficient de constriction a été étudiée dans le cas d'une particule unique [3], l'analyse théorique de l'interaction sociale entre particules reste à faire.

7.2.4 Mise en œuvre et validation

Bien que les algorithmes choisis soient conçus pour l'optimisation globale, nous les initialisons à l'aide du résultat d'une phase de détection. Celle-ci consiste à analyser les composantes connexes extraites des cartes de pré-détection. Les composantes répondant à certains critères de forme et de position fournissent une partie de la population initiale, le complément étant apporté par tirage aléatoire.

Par ailleurs, nous avons mis en œuvre, dans l'esprit de [13], une technique d'hybridation entre méthode globale et méthode locale. Plus précisément, une descente de gradient est effectuée avant le lancement de l'algorithme méta-heuristique. Cela peut suffire à obtenir une solution de faible énergie. De même, à la sortie de l'optimisation globale (lorsque celle-ci intervient sur le critère de valeur de l'énergie), les paramètres du modèle déformable sont affinés par descente de gradient.

Les algorithmes ont été comparés de manière objective sur une séquence de 48 images ayant fait l'objet d'un dépouillement manuel par un opérateur. Un exemple de résultat est montré Fig. 7.3(c), superposé sur la carte de rouge extraite d'une image en contre-jour. L'expérimentation était fondée sur l'utilisation de courbes COR et d'histogrammes de distance à la vérité-terrain. Elle a montré [8] la supériorité de l'algorithme PSO sur l'ES et la CS : on atteint un taux de bonnes détections de 100 % sur cette séquence de test, pour un nombre de fausses alarmes négligeable. Par ailleurs, elle a permis de vérifier que le détecteur était relativement bien formulé, au sens où une faible valeur d'énergie correspond effectivement à une bonne détection. Cela permet de définir un critère de décision pouvant éliminer d'éventuelles fausses alarmes sur la base de la valeur de l'énergie, ce qui ouvre la voie à une automatisation de l'algorithme.

7.3 Conclusion

Dans ce chapitre, nous avons abordé à l'aide du formalisme des modèles déformables le problème difficile de la détection de la signalisation verticale dans des images acquises en lumière naturelle visible. Nous avons décrit une approche combinant modélisation par forme prototype et optimisation par méta-heuristiques biologiques. La fonction de coût optimisée repose sur l'utilisation d'informations de couleur et de contour, par le biais de cartes de distance. L'algorithme d'optimisation retenu est celui de l'essaim de particules. Par ailleurs, notre étude a inclus le développement d'un logiciel démonstrateur de détection de la signalisation de danger, réalisé

7. FORMES PROTOTYPES ET MÉTA-HEURISTIQUES BIOLOGIQUES

par la Division Informatique du CETE de l'Est - cf. Fig. 7.4. Le logiciel a permis l'évaluation des différents schémas de détection et réglages des paramètres internes sur des séquences plus importantes.

Si les résultats en termes de détection se sont montrés excellents sur des séquences de quelques dizaines d'images, le coût calculatoire associé au traitement de séquences très volumineuses d'images demeure important. Or, les séquences de scènes routières ne contiennent qu'une très faible proportion d'images comportant des panneaux de signalisation. Il peut donc être intéressant d'adopter une démarche en deux étapes : pré-sélection des images susceptibles de contenir effectivement un panneau et détection proprement dite. En proposant un algorithme simple de détection des pixels rouges et des critères géométriques de sélection des composantes connexes, nous avons jeté les bases de la méthode décrite au chapitre 3. Cette étape de pré-détection a été incorporée dans le logiciel démonstrateur. Au final, nous avons atteint, sur des séquences de plusieurs milliers d'images, des taux de bonnes détections supérieurs à 80 % et des taux de fausses alarmes inférieurs à 10 % (il est possible d'atteindre 90 % de bonnes détections si l'on tolère 25 % de fausses alarmes). Les panneaux manqués correspondent le plus souvent à des panneaux très usés, ou à des modèles ne répondant pas aux hypothèses de l'algorithme de pré-sélection (voir chapitre 3).

De plus, l'automatisation complète du détecteur nécessite de gérer aussi bien l'absence de signalisation dans les images que la présence d'une ou plusieurs occurrences de panneaux. La valeur résiduelle de l'énergie fournit un critère de détection susceptible d'aider à filtrer les fausses alarmes issues de la phase de pré-sélection. L'ajustement de plusieurs panneaux peut s'envisager soit par mise à jour de l'observation, soit par modification du critère énergétique, soit enfin par utilisation d'algorithmes conçus pour la localisation de plusieurs minima de bonne qualité, comme cela peut être le cas de la sélection clonale.

Enfin, notons que nos travaux sont à la base des développements récents de l'équipe de l'entreprise *VIAMETRIS*, rapportés dans [14]. J. Ninot et son équipe proposent notamment de remplacer l'algorithme de descente de gradient utilisé pour l'ajustement local de la forme par un algorithme de type ICP (*Iterative Closest Point*) inspiré de [9], ce qui permet d'améliorer la vitesse de convergence de l'algorithme. Par ailleurs, la détection des panneaux d'interdiction est implantée et la méthode inclut une étape de reconnaissance par corrélation.

7.3. CONCLUSION

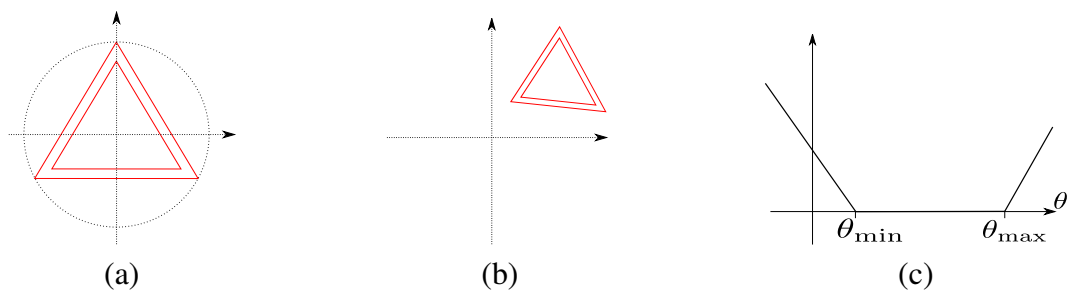


FIG. 7.2 – (a) Prototype ; (b) instance déformée selon (7.1) ; (c) fonction « vallée » .

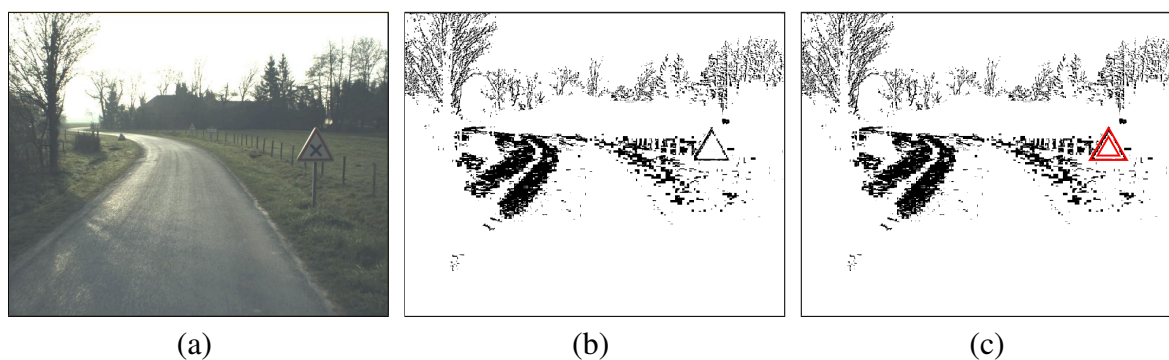


FIG. 7.3 – (a) Image originale ; (b) carte de la couleur rouge ; (c) résultat de l'optimisation.

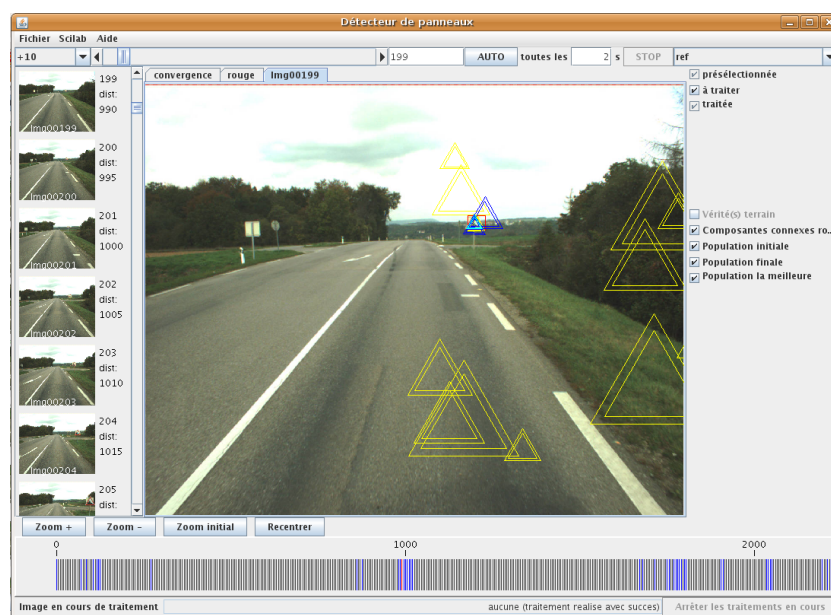


FIG. 7.4 – Interface graphique du logiciel démonstrateur.

Bibliographie

- [1] H.-G. BEYER. *The Theory of Evolution Strategies*. Springer Berlin Heidelberg, 2010.
- [2] P. CHARBONNIER. *Modèles de forme et d'apparence en traitement d'images*. Habilitation à diriger les recherches, Université de Strasbourg, septembre 2009.
- [3] M. CLERC et J. KENNEDY. The particle swarm - explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation*, 6(1) :58–73, 2002.
- [4] L. de CASTRO et F. von ZUBEN. Learning and optimization using the clonal selection principle. *IEEE Transactions on Evolutionary Computation*, 6(3) :239–251, 2002.
- [5] M.P. DUBUISSON JOLLY, S. LAKSHMANAN, et A. JAIN. Vehicle segmentation and classification using deformable templates. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 18(3) :293–308, mars 1996.
- [6] G. DUTILLEUX. Algorithmes pour l'optimisation globale. Dans Pierre ARGOUL, Nelly POINT, et Guillaume DUTILLEUX, éditeurs, *Problèmes inverses en génie civil*, numéro 15 dans Collection Études et Recherches des Laboratoires des Ponts et Chaussées (ERLPC) - Série sciences pour le génie civil, chapitre 9, pages 187–215. Laboratoire Central des Ponts et Chaussées, Paris, janvier 2009.
- [7] G. DUTILLEUX et P. CHARBONNIER. Détection de signalisation routière par ajustement de formes prototypes. Dans *20ème colloque GRETSI sur le traitement du signal et des images*, pages 1109–1112, Louvain-la-Neuve, Belgique, septembre 2005.
- [8] G. DUTILLEUX et P. CHARBONNIER. Métaheuristiques biologiques pour la détection de la signalisation routière. Dans P. SIARRY, éditeur, *Optimisation en traitement du signal et de l'image*, Traité IC2, série traitement du signal et de l'image, chapitre 10, pages 271–294. Hermes, avril 2007. ISBN 978-2-7462-1463-7.
- [9] A. FITZGIBBON. Robust registration of 2D and 3D point sets. *Image and Vision Computing*, 21(12-13) :1145–1153, décembre 2003.
- [10] J. KENNEDY et R. EBERHART. Particle swarm optimization. Dans *IEEE International Conference on Neural Networks*, volume IV, pages 1942–1948, 1995.
- [11] LE MINISTRE DE L'INTÉRIEUR et LE MINISTRE DE L'ÉQUIPEMENT ET DE L'AMÉNAGEMENT DU TERRITOIRE. Instruction interministérielle sur la signalisation routière, parties 2 et 3, Arrêté du 7 juin 1977. Dernière mise à jour 2009.
- [12] Z. MICHALEWICZ. *Genetic algorithms + data structures = evolution programs (2nd, extended ed.)*. Springer-Verlag New York, Inc., New York, NY, USA, 1994.
- [13] M. MIGNOTTE, C. COLLET, P. PEREZ, et P. BOUTHEMY. Hybrid genetic optimization and statistical model-based approach for the classification of shadow shapes in sonar imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(2) :129–141, février 2000.
- [14] J. NINOT, L. SMADJA, et K. HEGGARTY. Road sign recognition using a hybrid evolutionary algorithm and primitives fusion. *ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 39(3), 2010. À paraître.

BIBLIOGRAPHIE

- [15] I. RECHENBERG. *Evolutionssstrategie'94*. Frommann-Holzboog, 1994. en allemand.
- [16] J.A. SETHIAN. A fast marching level set method for monotonically advancing fronts. *Proc. Nat. Acad. Sci.*, 93(4) :1591–1595, 1996.
- [17] A. YUILLE, P. HALLINAN, et D. COHEN. Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 8(2) :99–111, août 1992.

7. FORMES PROTOTYPES ET MÉTA-HEURISTIQUES BIOLOGIQUES

Chapitre 8

Modèles probabilistes d'apparence et estimation robuste

*Pierre Charbonnier,
Rozenn Dahyot,
Torbjørn Vik,
Fabrice Heitz*

La détection d'objets d'intérêt dans des images est un sujet difficile, qui a justifié de nombreux travaux de recherche depuis plusieurs décennies. Nous considérons dans ce chapitre les approches par *modèles d'apparence*, qui rencontrent un succès certain depuis leur application, dès le début des années 90, en reconnaissance de visages [27, 31]. Elles permettent la représentation de classes d'objets et l'apprentissage de variations de forme, d'orientation ou d'illumination. Dans cette approche de la reconnaissance de formes, toute image est représentée par un vecteur, dont chaque composante est la valeur de niveau de gris ou de couleur d'un pixel. Une telle représentation, globale, de l'apparence est évidemment peu parcimonieuse. Or, l'ensemble des apparences d'un objet n'occupe en général qu'un petit sous-espace de l'espace des apparences possibles. Il peut donc être décrit par une variable de faible dimensionnalité.

Classiquement, des techniques linéaires telles que l'analyse en composantes principales (ACP) sont utilisées pour « capturer » la relation entre cette variable latente, dite variable dans l'espace propre, et l'apparence visuelle de l'objet [27, 31]. D'autres techniques linéaires, telles que l'analyse discriminante de Fisher [1], l'analyse factorielle ou l'analyse en composantes indépendantes peuvent aussi être utilisées. Afin de capturer plus finement des caractéristiques locales du sous-espace de représentation, des méthodes non linéaires ont été proposées. Dans une représentation par ACP, Murase et Nayar [21] utilisent une modélisation par *spline* de la variété décrite par la variable dans l'espace propre, ou variable latente. Dans le même registre d'idées, une extension non linéaire de l'ACP connue sous le nom de courbes et surfaces principales a été proposée dans [14]. Une autre façon, classique en classification et reconnaissance des formes, d'introduire de la non-linéarité est l'utilisation de noyaux, qui conduit ici à la *kernel PCA* [25]. Des approches neuronales sont également envisageables. Enfin, des approches locales de l'ACP, ou plus généralement, l'utilisation de modèles linéaires locaux [24] représentent une alternative intéressante. Pour un état de l'art plus complet sur les méthodes de représentation de l'apparence, le lecteur pourra se référer à [6, 32, 16].

Nous nous en tiendrons ici à l'approche probabiliste de l'ACP [29, 23], qui considère pour la formation des images un modèle linéaire bruité de la variable latente (*cf.* Fig. 8.1). On associe à ce modèle linéaire deux modèles de distributions statistiques : l'un pour la variable dans l'espace propre, l'autre pour le bruit. Ces approches sont, en général, limitées à des hypothèses gaussiennes (ou, au mieux, multi-gaussiennes [20, 30]). De telles hypothèses sont réductrices : les distributions rencontrées dans des applications réelles sont rarement gaussiennes. Utiliser une représentation non gaussienne de la variable latente permet également de contrebalancer le fait que le modèle génératif linéaire est, forcément, limité lorsqu'il s'agit de représenter certaines variétés non linéaires d'apparence. Par ailleurs, les perturbations importantes parfois rencontrées, telles que les occultations, ne sont pas gérées par les modèles gaussiens de bruit. S'affranchir des hypothèses gaussiennes et proposer des techniques efficaces pour la mise en œuvre pratique des modèles proposés a constitué le fil conducteur de nos travaux dans ce domaine.

Nous présentons au paragraphe 8.1 un cadre bayésien original pour la détection et la reconnaissance, développé dans [6, 8]. Celui-ci associe un apprentissage statistique par ACP à des méthodes d'estimation robuste et tient compte de la distribution de la variable latente. En utilisant la théorie semi-quadratique, on bâtit une nouvelle famille d'algorithmes de détection et de reconnaissance, capables de gérer des variabilités relativement complexes d'apparence, dans les situations défavorables souvent rencontrées dans les scènes routières. Nous avons, dans un premier temps, recours à une approximation pour gérer le cas où la distribution de la variable latente est modélisée de façon non paramétrique. Nous décrivons au paragraphe 8.2 la version modifiée de l'algorithme *Mean Shift* [11], proposée dans [34, 32], pour s'affranchir de cette approximation, ce qui ouvre la voie à la modélisation de variétés arbitrairement complexes.

8.1 Détection d'objets par modèles probabilistes d'apparence

8.1.1 Décomposition sur espace propre

Les approches par décomposition sur espace propre impliquent une phase préliminaire d'apprentissage. On constitue pour cela une base \mathcal{B} rassemblant $N_{\mathcal{B}}$ images \mathbf{x}_n , de taille N pixels, en niveaux de gris ou en couleurs (pour plus de détails sur le traitement des images couleurs, voir [6]). La base constitue un échantillon représentatif des objets à modéliser sous leurs différentes apparences. Les échantillons sont centrés, en leur soustrayant la moyenne empirique de la base : $\mu_{\mathcal{B}} = \frac{1}{N_{\mathcal{B}}} \sum_{n=1}^{N_{\mathcal{B}}} \mathbf{x}_n$. L'ACP consiste à diagonaliser la matrice de covariance de l'échantillon, définie par :

$$\Sigma_{\mathcal{B}} = \frac{1}{N_{\mathcal{B}}} \sum_{n=1}^{N_{\mathcal{B}}} (\mathbf{x}_n - \mu_{\mathcal{B}}) (\mathbf{x}_n - \mu_{\mathcal{B}})^T, \quad (8.1)$$

ce qui s'écrit : $\Sigma_{\mathcal{B}} = \mathbf{R}_N \Lambda_N \mathbf{R}_N^T$. La matrice Λ_N est diagonale et contient les N valeurs propres de $\Sigma_{\mathcal{B}}$, rangées par ordre décroissant, qui s'interprètent comme des variances. Les N vecteurs propres orthogonaux associés, rangés en colonne dans \mathbf{R}_N , représentent donc les axes des principales variations observables sur l'échantillon d'apprentissage. En général, les variations sont négligeables à partir d'un certain rang $D \ll N$. Une troncature de la décomposition, définissant le sous-espace propre de représentation est alors effectuée en ne conservant que les D premiers

vecteurs propres R_j , rangés en colonne dans une matrice $N \times D$, que nous noterons R . Le taux de variance expliquée, défini comme le rapport de la somme des D premières valeurs propres et du total des valeurs propres, fournit un critère objectif de choix du rang de troncature.

8.1.2 Le modèle d'observation

Notre modèle d'observation suppose, comme dans [29] dont il est inspiré, une relation linéaire entre une image observée y et la variable latente f dans l'espace propre D -dimensionnel :

$$y = \mu + \sum_{j=1}^D f_j R_j + \eta = \mu + Rf + \eta \quad (8.2)$$

où η représente le bruit d'observation (voir Fig. 8.1). En notant $p = y - \mu$ la variable centrée, on retrouve le modèle génératif linéaire classique de l'analyse factorielle. L'intérêt de ce type de modèle est que D est relativement faible (typiquement, quelques dizaines de coefficients, à comparer aux centaines de milliers de pixels des images traitées). Les calculs nécessaires à la détection et à la reconnaissance s'en trouvent donc fortement simplifiés.

8.1.3 Détection et reconnaissance dans le cadre bayésien

Le principe de l'algorithme de détection d'objets d'intérêt dans une scène est illustré figure 8.2. En chaque position de l'image, on extrait un vecteur d'observation y dont on évalue la vraisemblance $\mathcal{P}(y|\mathcal{B})$, par rapport au modèle appris sur la base \mathcal{B} . Cette valeur est affectée à la position correspondant au centre de l'imagette extraite, y . Lorsque toute la scène a été parcourue, on dispose d'une carte de vraisemblance. La position des objets d'intérêt éventuellement présents dans la scène est alors obtenue par seuillage.

La vraisemblance de l'observation est, en principe, obtenue par intégration de la loi conjointe de y et de f :

$$\mathcal{P}(y|\mathcal{B}) = \int \mathcal{P}(y, f|\mathcal{B}) df = \int \mathcal{P}(y|f, \mathcal{B}) \mathcal{P}(f|\mathcal{B}) df, \quad (8.3)$$

où l'on reconnaît la densité de probabilité du bruit, $\mathcal{P}(y|f, \mathcal{B}) = \mathcal{P}(\eta|\mathcal{B})$ et la densité *a priori* de la variable latente, $\mathcal{P}(f|\mathcal{B})$. En l'absence d'expression analytique simple (sauf dans certains cas particuliers), et pour éviter les lourdeurs calculatoires liées aux méthodes de simulation numérique, nous avons recours à une approximation de la distribution $\mathcal{P}(y|\mathcal{B})$. Celle-ci, classique en inférence bayésienne (cf. [9], p. 92), suppose que la vraisemblance est suffisamment « pointue » pour qu'on puisse l'approcher par la valeur de son mode multipliée par la largeur du pic. L'expression que nous utilisons donc est :

$$\mathcal{P}(y|\mathcal{B}) \propto \mathcal{P}(y|\hat{f}, \mathcal{B}) \mathcal{P}(\hat{f}|\mathcal{B}), \quad (8.4)$$

où $\hat{f} = \arg \max_f \mathcal{P}(y, f|\mathcal{B})$ est l'estimée au sens du Maximum A Posteriori (MAP) des paramètres du modèle. Cette approximation conduit, en général, à des résultats de détection très satisfaisants, comme l'ont montré nos expériences.

8. MODÈLES PROBABILISTES D'APPARENCE

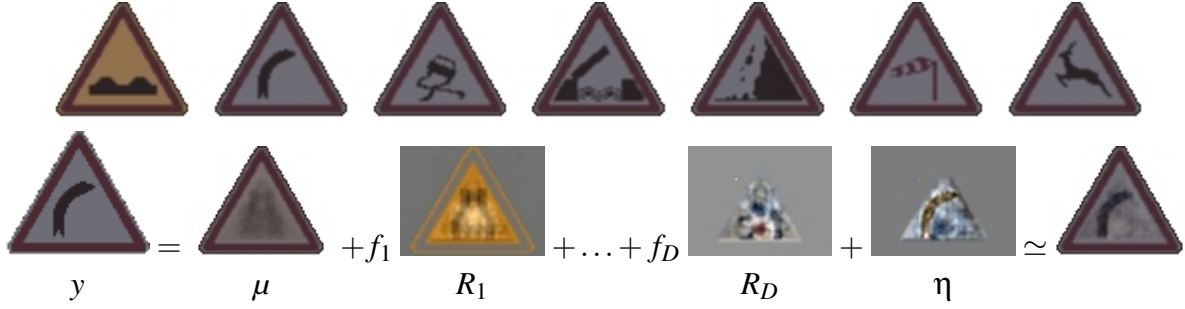


FIG. 8.1 – Ligne du haut : quelques-unes des $N_{\mathcal{B}} = 43$ images d'apprentissage en couleurs considérées dans cette expérience. Ligne du bas : décomposition d'une observation selon le modèle (8.2). Les coordonnées sont obtenues par projection orthogonale (hypothèses statistiques : bruit gaussien, *a priori* uniforme) sur les $D = 10$ premiers vecteurs propres (taux de variance expliquée : environ 75%). Les vecteurs propres et le vecteur d'erreur sont normalisés entre 0 et 255 pour la visualisation. L'erreur quadratique moyenne est de $1,6 \cdot 10^{-4}$ sur cette expérience.

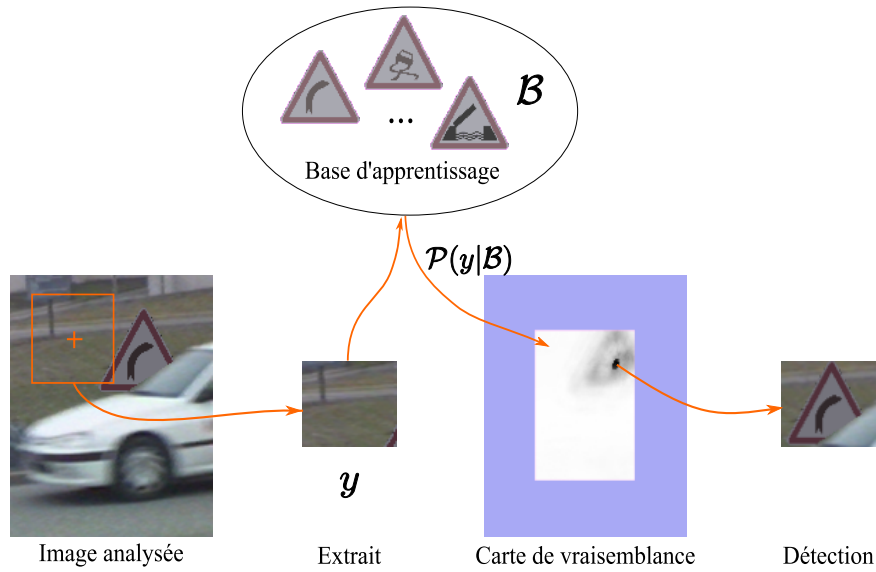


FIG. 8.2 – Principe de l'algorithme de détection (voir texte). Les fortes valeurs de vraisemblance apparaissent en sombre. D'après [8].

La reconnaissance, quant à elle, est effectuée en comparant les coordonnées dans l'espace propre estimées au sens du MAP ou du Maximum de Vraisemblance (MV) [7], \hat{f} , à celles des panneaux d'apprentissage. Le modèle le plus proche est choisi comme modèle reconnu, dans l'esprit de la classification au sens du plus proche voisin (*Nearest Neighbor classification*).

8.1.4 Estimation robuste au sens du MAP

L'estimation de la variable latente f , au sens du MAP, joue un rôle central pour la détection et la reconnaissance. Le cadre proposé au paragraphe précédant permet, en premier lieu, de décrire de manière unifiée plusieurs détecteurs proposés dans la littérature. Ainsi, en supposant gaussien le bruit $\mathcal{P}(y|f, \mathcal{B})$ et uniforme l'*a priori* sur la variable latente $\mathcal{P}(f|\mathcal{B})$, on retrouve l'estimation au sens des moindres carrés, dont la solution est la projection orthogonale de p sur le sous-espace propre. La mesure de similarité associée est appelée *Distance From Feature Space* (DFFS) dans [20]. Dans le cas d'un bruit et d'un *a priori* gaussiens, l'estimée s'écrit $\hat{f} = (I_D + \sigma^2 \Lambda^{-1})^{-1} R^T p$ (où σ est la variance du bruit et $p = y - \mu$, la variable centrée) et l'on retrouve la distance quadratique proposée indépendamment dans [29] et [20].

Au-delà du cadre classique des hypothèses gaussiennes ou uniformes, nous avons montré [8] que l'association d'un modèle robuste de distribution du bruit et d'un modèle *a priori* bien adapté permettait d'améliorer de manière significative les performances des systèmes de détection. En utilisant la théorie semi-quadratique, nous avons fourni des algorithmes déterministes permettant de réaliser la régression dans la plupart des situations. Cette théorie, que l'on a vu resurgir à des époques différentes, dans des domaines variés tels que l'estimation robuste [17], la régularisation [12, 13, 3] ou encore l'optimisation [19], a fait l'objet d'un grand nombre de travaux et nous n'en faisons qu'une description succincte, renvoyant le lecteur à [28], par exemple, pour plus de détails et de références.

Théorie semi-quadratique

On s'intéresse ici au cas où les densités de probabilités entrant en jeu dans l'estimation sont de forme exponentielle, plus précisément, gibbsienne. Cette classe de fonction inclut la gaussienne mais ne se limite pas à celle-ci. En passant au co-logarithme, le problème d'estimation se ramène à la minimisation de fonctionnelles de la forme générale :

$$\mathcal{J}(f) = \frac{1}{2} \sum_{i=1}^N \varphi(u_i). \quad (8.5)$$

Nous utilisons volontairement la variable muette u par souci de généralité. Dans le contexte de l'estimation, elle représente un *résidu* (écart entre l'observation et sa prédiction par le modèle génératif à partir de la valeur de f) et ce type d'estimateur est appelé M-estimateur [17]. Dans le domaine de la régularisation, u représente le gradient de l'image f et on parle de modèle à phi-fonction [12, 13, 3].

Lorsque la distribution optimisée est gaussienne, la fonction φ est quadratique. Les fortes valeurs de u ont donc une influence importante sur le résultat de l'estimation. Une seule donnée erronée (*outlier*), produisant une forte valeur de résidu, suffit à fausser considérablement le résultat : on dit que le point de rupture de l'estimateur est 0. Pour diminuer cette influence et rendre l'estimation robuste à la présence de données erronées (certains estimateurs supportent

8. MODÈLES PROBABILISTES D'APPARENCE

jusqu'à 50% d'*outliers*), il est nécessaire de modifier la forme de la fonction φ . De nombreuses variantes ont été proposées dans la littérature [3]. Un examen des équations normales associées à la minimisation de (8.5) montre [3, 28] que les conditions à vérifier pour assurer la robustesse de l'estimation sont les suivantes :

$$\lim_{u \rightarrow 0} \frac{\varphi'(u)}{2u} = M, \quad 0 < M < +\infty \quad (8.6)$$

$$\lim_{u \rightarrow \infty} \frac{\varphi'(u)}{2u} = 0 \quad (8.7)$$

$$\frac{\varphi'(u)}{2u} \text{ strictement décroissante.} \quad (8.8)$$

En général, M est choisi égal à 1. Ces conditions assurent que φ soit quadratique près de l'origine (8.6) et sous-quadratique à l'infini (8.7). La condition (8.8) assure la cohérence du modèle.

On peut montrer [3] que les fonctions réelles φ paires, croissantes, vérifiant les conditions (8.6-8.8) avec $M = 1$, sont telles que :

$$\varphi(u) = \inf_{b \in [0,1]} (bu^2 + \Psi(b)), \forall u \quad \text{et} \quad b_{inf} = \frac{\varphi'(u)}{2u}, \quad (8.9)$$

où Ψ est une fonction strictement convexe définie à partir de φ et b est la variable auxiliaire. On parle d'extension multiplicative. Ce résultat généralise un théorème initialement limité aux fonctions φ à asymptote horizontale, introduit dans [12] dans le contexte de la régularisation des problèmes inverses en reconstruction d'images. On a également, sous les mêmes conditions [3], une extension additive :

$$\varphi(u) = \inf_{b \in \mathbb{R}^+} ((b-u)^2 + \xi(b)), \forall u \quad \text{et} \quad b_{inf} = \left(1 - \frac{\varphi'(u)}{2u}\right)u, \quad (8.10)$$

où ξ est une fonction strictement convexe définie à partir de φ . En reconstruction régularisée, ce résultat a été formalisé à l'origine dans [13]. Ces expressions sont des approximations locales paraboliques du critère initial, ce qui correspond à l'approche suivie dans le domaine de l'estimation robuste par Huber dès les années 1970 [17], comme remarqué dans [2, 6].

En reportant les équations (8.9) et (8.10) dans (8.5), on obtient respectivement les critères *augmentés* d'une variable auxiliaire, b :

$$\mathcal{J}^*(f, b) = \frac{1}{2} \sum_{i=1}^N b_i (u_i)^2 + \Psi(b_i) = \frac{1}{2} \|u\|_B^2 + \frac{1}{2} \sum_{i=1}^N \Psi(b_i), \quad (8.11)$$

avec $B = \text{diag}(b_i)$ et :

$$\mathcal{J}^\#(f, b) = \frac{1}{2} \sum_{i=1}^N (b_i - u_i)^2 + \xi(b_i) = \frac{1}{2} \|u - b\|^2 + \frac{1}{2} \sum_{i=1}^N \xi(b_i). \quad (8.12)$$

Contrairement à ce que l'on pourrait penser à première vue, ces critères sont plus simples à minimiser que le critère d'origine. En effet, la minimisation directe de \mathcal{J} conduit à des équations non linéaires. Par contre, à b fixée, les critères augmentés sont quadratiques et conduisent donc à la résolution d'équations normales linéaires. De plus, lorsque la variable f (donc, u) est fixée le critère augmenté est convexe en b et l'expression du minimum, b_{inf} , est donnée dans (8.9) et (8.10). Cela suggère une stratégie de *minimisations alternées* par rapport à chacune des variables, en fixant l'autre, dont on peut montrer la convergence vers un minimum (éventuellement local) du critère (8.5).

8.1. DÉTECTION D'OBJETS PAR MODÈLES PROBABILISTES D'APPARENCE

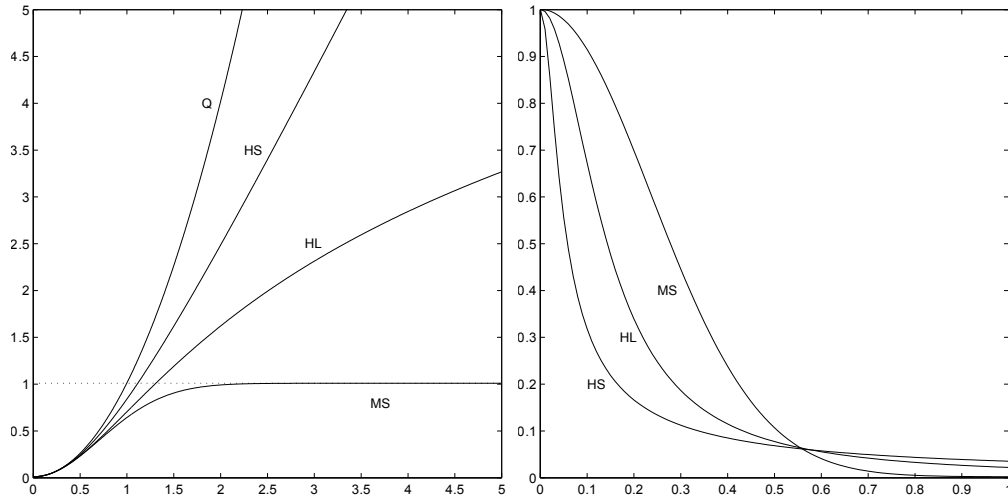


FIG. 8.3 – A gauche, exemples de fonctions ϕ , où Q est la fonction quadratique : $\phi(u) = u^2$; HS est la fonction dite *des hyper-surfaces* [3] : $\phi(u) = 2\sqrt{1+u^2} - 2$; HL est la fonction de Hebert et Leahy [15] : $\phi(u) = \log(1+u^2)$; MS est la fonction correspondant au Mean-Shift gaussien (cf. §8.2.1) : $\phi(u) = 1 - \exp(-u^2)$. A droite, fonctions de pondérations associées, $\frac{\phi'(u)}{2u}$ (un coefficient d'échelle arbitraire est appliqué aux abscisses u , cf. [3]).

Algorithmes d'estimation

Considérons d'abord le cas d'une hypothèse robuste pour le bruit et d'un *a priori* uniforme (MV robuste). Le critère à minimiser s'écrit alors directement sous la forme (8.5), la variable u représentant le résidu, $(p - Rf)/s$. Le paramètre s est un facteur d'échelle dont le réglage peut être laissé à l'utilisateur ou bien déterminé lors de la phase d'apprentissage [6]. En augmentant celui-ci selon (8.11) et en appliquant le principe des minimisations alternées en f et en b , on obtient un algorithme de *moindres carrés re-pondérés itératifs* (ou *Iterated Re-weighted Least Squares*, IRLS) :

1. Choisir une estimée initiale f^0 , et initialiser l'indice de boucle à $k = 1$,
2. Calculer $u^k = \frac{1}{s}(p - Rf^{k-1})$ et $B^k = \text{diag}\{\phi'(u_i^k)/(2u_i^k)\}_{i=1\dots N}$,
3. Résoudre $R^T B^k R f^k = R^T B^k p$
4. Si $\|f^k - f^{k-1}\| > \epsilon$, incrémenter k , et aller en 2, sinon $f_{IRLS} = f^k$.

On peut également utiliser son équivalent pour la forme additive (8.12) : l'algorithme des *résidus modifiés* (ou *Residual Steepest Descent*, RSD). Celui-ci s'écrit :

1. Choisir une estimée initiale f^0 , et initialiser l'indice de boucle à $k = 1$,
2. $u^k = \frac{1}{s}(p - Rf^{k-1})$ et $b^k = \{u_i^k(1 - \phi'(u_i^k)/(2u_i^k))\}_{i=1\dots N}$,
3. Calculer $f^k = R^T(p - sb^k)$
4. Si $\|f^k - f^{k-1}\| > \epsilon$, incrémenter k , et aller en 2, sinon $f_{RSD} = f^k$.

Le second algorithme est, à notre connaissance, généralement peu employé. En effet, son taux de convergence est plus faible que celui de l'IRLS et il nécessite un plus grand nombre d'itérations pour converger. Par contre, on peut remarquer que chaque itération est beaucoup moins

8. MODÈLES PROBABILISTES D'APPARENCE

coûteuse, car elle n'implique pas d'inversion de système linéaire. Au total, notre expérience a montré que l'algorithme des résidus modifiés est plus rapide pour ce type de problème de régression [6].

Dans le cas d'une hypothèse robuste pour le bruit et d'un *a priori* gaussien, un terme quadratique est ajouté à (8.5). Les algorithmes dérivés conservent la même architecture, mais le calcul de la nouvelle estimée f_k devient, respectivement : $f_{IRLS}^k = (R^T B^k R + s^2 \Lambda^{-1})^{-1} R^T B^k p$ pour l'IRLS et : $f_{RSD}^k = (I_D + s^2 \Lambda^{-1})^{-1} R^T (p - s b^k)$ pour l'algorithme des résidus modifiés (le système à résoudre est diagonal).

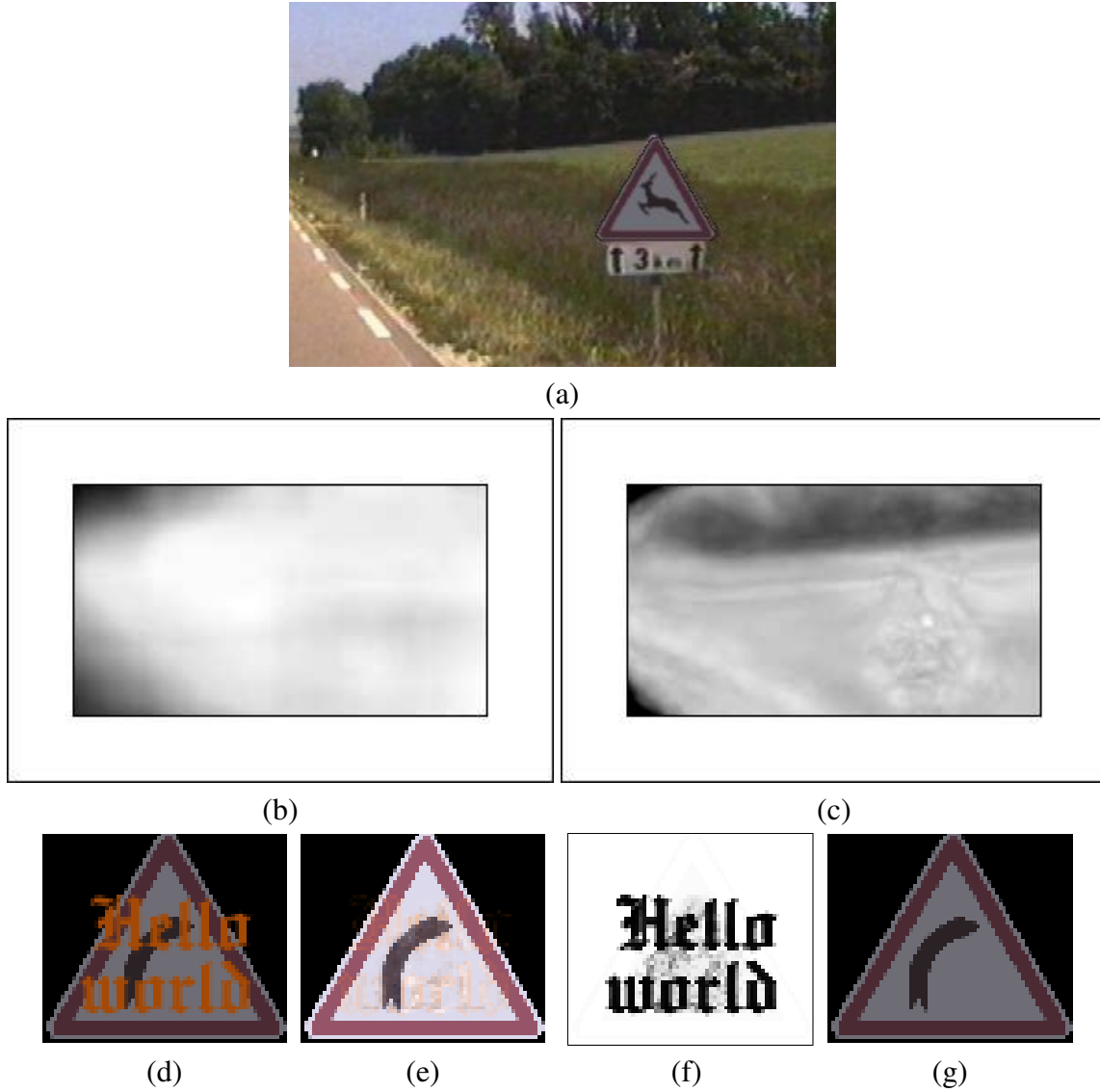


FIG. 8.4 – Lignes supérieures : expérience de détection. Image synthétique analysée (a). Cartes de vraisemblance pour des hypothèses de bruit et d'*a priori* (respectivement) : (b) gaussien-gaussien et (c) robuste-non gaussien. Les plus fortes vraisemblances apparaissent en clair. Ligne du bas : expérience de reconnaissance. (d) image analysée, y . (e) images reconstruite $R\hat{f}$. (f) carte des données erronées (*outliers*), b . (g) panneau reconnu.

Nous avons également envisagé le cas d'une hypothèse robuste sur le bruit et d'un modèle statistique non gaussien sur la variable latente. Ce dernier est défini ou estimé à partir des coordonnées des échantillons dans l'espace propre. Par exemple, si l'apprentissage porte sur des rotations de l'objet dans le plan image [8], on dispose d'une expression analytique de $P(f|\mathcal{B})$ [22]. Dans une approche plus générique, nous avons proposé de considérer une formulation non paramétrique de l'*a priori*, par fenêtres de Parzen. Dans un premier temps, nous avons recouru à une approximation pour faciliter l'optimisation du critère ainsi bâti. Celle-ci consiste à effectuer la régression au sens du MV, c'est-à-dire sans *a priori*, puis à injecter l'estimée \hat{f}_{MV} ainsi obtenue dans l'expression du MAP (8.4) pour le calcul des cartes de vraisemblance. Bien qu'une telle approximation ne soit théoriquement pas justifiée en général, les résultats obtenus [8] ont montré l'intérêt de la modélisation, comme l'illustre la figure 8.4. Dans cet exemple, la base d'images d'apprentissage contient 43 signaux de danger, chacun étant vu sous 36 angles de rotation dans le plan. On ne retient que 30 vecteurs propres. La méthode de référence, fondée sur des hypothèses gaussiennes [20], est mise en défaut, tandis qu'une bonne discrimination est obtenue en associant l'estimation robuste à une modélisation *a priori* adaptée aux données.

En reconnaissance, l'utilisation d'un estimateur robuste améliore également aussi les performances du système [6], comme le montre la figure 8.4. Dans cette expérience, le panneau est en partie occulté par une inscription jaune. Le panneau reconnu à partir de l'estimée au sens du MV gaussien est le panneau temporaire à fond jaune, alors que le panneau reconnu dans le cas du MV robuste est correct. On note que l'estimation fournit, de plus, une carte des données erronées (*outliers*).

8.2 Modèles non paramétriques dans l'espace propre et optimisation

Nous décrivons maintenant un algorithme permettant de résoudre le problème d'estimation de la variable latente au sens du MAP dans le cas d'un terme d'attache aux données robuste et d'un *a priori* non paramétrique, sans recours à aucune approximation [34, 32]. L'optimisation est réalisée grâce à une version modifiée de l'algorithme *Mean Shift*, dont nous expliquons d'abord le principe et donnons une interprétation en termes d'algorithme semi-quadratique.

8.2.1 Le *Mean Shift* : un algorithme semi-quadratique

L'algorithme *Mean Shift* a été proposé par Fukunaga en 1975 [11], remis au goût du jour vingt ans plus tard par Cheng [4] en classification et enfin, popularisé par Comaniciu et Meer [5] au début des années 2000 pour des applications de segmentation et de suivi. Il s'agit d'un algorithme de montée de gradient à pas adaptatif, dédié à des densités de probabilité décrites de manière non paramétrique. En particulier, on s'intéresse au modèle par fenêtres de Parzen [9], de noyau K à symétrie radiale, auquel on associe une fonction appelée profil, κ , telle que $K(u) = c\kappa(\|u\|^2)$. La densité de probabilité de la variable latente, estimée en f à partir de $N_{\mathcal{B}}$ échantillons f_n , s'écrit :

$$\hat{p}_K(f) = \frac{c}{N_{\mathcal{B}}h^D} \sum_{n=1}^{N_{\mathcal{B}}} \kappa\left(\left\|\frac{f-f_n}{h}\right\|^2\right), \quad (8.13)$$

8. MODÈLES PROBABILISTES D'APPARENCE

où h est le paramètre dit de *largeur de bande* qui règle le degré de lissage de l'estimée. Afin de définir un algorithme de montée de gradient, on dérive d'abord $\hat{p}(f)$. En divisant et multipliant l'expression obtenue par $\sum_{n=1}^{N_B} g\left(\left\|\frac{f-f_n}{h}\right\|^2\right)$, où $g(u) = -\kappa'(u)$, on arrive à :

$$\nabla \hat{p}_K(f) = \underbrace{\frac{2c}{N_B h^{D+2}} \left[\sum_{n=1}^{N_B} g\left(\left\|\frac{f-f_n}{h}\right\|^2\right) \right]}_{\hat{p}_G(f)} \times \underbrace{\left[\frac{\sum_{n=1}^{N_B} f_n g\left(\left\|\frac{f-f_n}{h}\right\|^2\right)}{\sum_{n=1}^{N_B} g\left(\left\|\frac{f-f_n}{h}\right\|^2\right)} - f \right]}_{ms(f)}, \quad (8.14)$$

où G est le noyau radial de profil g . Le premier facteur est la densité de probabilité évaluée au point f à l'aide du modèle non-paramétrique de noyau G . Le second facteur est appelé *mean shift*. En ré-arrangeant cette expression, on voit qu'il s'agit d'une approximation du gradient normalisé de la densité de probabilité, à une constante multiplicative près : $ms(f) \propto \frac{\nabla \hat{p}_K(f)}{\hat{p}_G(f)}$. Il est intéressant d'utiliser ce gradient normalisé dans un algorithme de montée de gradient car le pas de montée s'adapte automatiquement à la valeur locale de densité de probabilité. L'itération de cet algorithme de gradient à pas adaptatif s'écrit naturellement :

$$f^{k+1} = f^k + ms(f^k) = \frac{\sum_{n=1}^{N_B} f_n g\left(\left\|\frac{f^k-f_n}{h}\right\|^2\right)}{\sum_{n=1}^{N_B} g\left(\left\|\frac{f^k-f_n}{h}\right\|^2\right)}. \quad (8.15)$$

Nous pouvons re-formuler cette itération en deux demi-étapes, pour mieux faire apparaître son caractère d'algorithme semi-quadratique (algorithme IRLS) :

$$b_n^{k+1} = g\left(\left\|\frac{f^k-f_n}{h}\right\|^2\right) \quad \forall n = 1 \dots N_B, \quad (8.16)$$

$$f^{k+1} = \frac{\sum_{n=1}^{N_B} b_n^{k+1} f_n^k}{\sum_{n=1}^{N_B} b_n^{k+1}}. \quad (8.17)$$

De fait, le problème de maximisation de $\hat{p}_K(f)$ défini dans (8.13) peut s'écrire comme le problème de minimisation de $\mathcal{J}(f) \propto \sum_{n=1}^{N_B} \varphi\left(\frac{f-f_n}{h}\right)$, où $\varphi(u) = cte - \kappa(\|u\|^2)$, où cte est une constante. Ce problème d'estimation robuste de la moyenne au sens du maximum de vraisemblance se résout aisément dans le cadre semi-quadratique. En effet, φ (tracée, dans le cas du noyau gaussien, sur la Fig. 8.3, courbe MS) satisfait les conditions d'application de la théorie semi-quadratique et le critère augmenté correspondant s'écrit :

$$\mathcal{J}^*(f, b) = \sum_{n=1}^{N_B} \left\{ b_n \left\| \frac{f-f_n}{h} \right\|^2 + \Psi(b_n) \right\}. \quad (8.18)$$

Le minimum en b à f fixé est obtenu pour $b_n = \varphi'(u_n)/2u_n = -\kappa'(\|u_n\|^2) = g(\|u_n\|^2)$, et l'on retrouve (8.16). D'autre part, la minimisation en f à b fixé conduit à (8.17). En procédant

par minimisations alternées, on aboutit bien à l'algorithme IRLS (8.16)-(8.17). La preuve de convergence proposée dans [5] utilise, d'ailleurs, les mêmes propriétés que celles des théorèmes semi-quadratiques. Notons que le parallèle que nous venons de faire a également été proposé de manière indépendante, dans [10].

8.2.2 Mean Shift et estimation du MAP gaussien

L'algorithme *Mean Shift* à noyau gaussien peut facilement être modifié pour s'accommoder d'un facteur gaussien de vraisemblance [34]. Le produit de gaussiennes demeure, en effet, gaussien. Si on note f_n les coordonnées dans l'espace propre de l'échantillon \mathbf{x}_n , Σ_η la matrice de covariance (diagonale) du bruit et Σ_f , celle associée aux noyaux de Parzen, la densité *a posteriori* s'écrit :

$$\mathcal{P}(f|p) = \frac{c}{N_B} \sum_{n=1}^{N_B} \exp \left[-\frac{1}{2} \left(\|p - Rf\|_{\Sigma_\eta^{-1}}^2 + \|f - f_n\|_{\Sigma_f^{-1}}^2 \right) \right], \quad (8.19)$$

où c est une constante. Cette expression peut se factoriser [32] sous forme d'une densité à noyau pondérée :

$$\mathcal{P}(f|p) = \frac{c}{N_B} \sum_{n=1}^{N_B} c_n \exp \left[-\frac{1}{2} \|f - \mu_n\|_{\Sigma^{-1}}^2 \right] \triangleq \frac{c}{N_B} \sum_{n=1}^{N_B} c_n \Gamma_n(f), \quad (8.20)$$

où $\Sigma = (R^T \Sigma_\eta^{-1} R + \Sigma_f^{-1})^{-1}$ et $\mu_n = \Sigma(R^T \Sigma_\eta^{-1} p + \Sigma_f^{-1} f_n)$. L'expression des pondérations c_n est donnée dans [34]. Évidemment, l'expression du *mean shift* est modifiée en conséquence :

$$ms(f) = \frac{\sum_{n=1}^{N_B} c_n \Gamma(f) \mu_n}{\sum_{n=1}^{N_B} c_n \Gamma(f)} - f. \quad (8.21)$$

Notons qu'une expression simplifiée, n'impliquant pas le calcul de μ_n (et donc plus facile à implanter), peut être dérivée de (8.21) [32, p. 177].

8.2.3 Modélisation non gaussienne dans l'espace propre et estimation robuste de l'apparence

L'algorithme *Mean Shift* étant semi-quadratique, il paraît intéressant de l'utiliser pour optimiser le critère d'estimation combinant vraisemblance robuste et *a priori* non paramétrique, dans un formalisme unifié. Une alternative consiste à passer au log dans l'expression de la loi *a posteriori* et à développer chaque terme séparément en utilisant (8.11), comme cela a été fait dans [18, 26].

Nous avons, pour notre part, choisi de conserver le formalisme du *Mean Shift* gaussien. Le principe général est le suivant : l'extension semi-quadratique (8.11) de la vraisemblance robuste $\mathcal{P}(p|f) \propto \exp \left[-\frac{1}{2} \sum_{i=1}^N \varphi(u_i) \right]$, où $u = \frac{1}{s} \Sigma_\eta^{-1/2} (p - Rf)$, conduit, en prenant son exponentielle négative, à :

$$\mathcal{P}^*(p|f, b) \propto \exp \left[-\frac{1}{2} \|p - Rf\|_{\frac{1}{s^2} \Sigma_\eta^{-1} B}^2 \right], \quad (8.22)$$

où $B = \text{diag}\{b_i\}_{i=1\dots N}$. Autrement dit, la vraisemblance *augmentée* $\mathcal{P}^*(p|f, b)$ est gaussienne. En lui adjoignant un *a priori* $\mathcal{P}(f)$ non paramétrique à noyau gaussien, on retrouve à B fixée un problème d'optimisation du MAP gaussien, similaire à (8.19), mais où Σ_η^{-1} est remplacé par $\frac{1}{s^2}\Sigma_\eta^{-1}B$. Or, comme nous l'avons vu au paragraphe précédent, celui-ci peut être résolu à l'aide d'un *Mean Shift* pondéré. Le principe des maximisations alternées nous conduit [33, 34, 32] à un algorithme de *Mean Shift* re-pondéré itératif, dont la structure globale est formée de deux boucles semi-quadratiques imbriquées :

1. Choisir une estimée initiale f^0 , et poser $B^0 = I_N$,
2. A partir de $f^{k,0} = f^{k-1}$, estimer f^k en itérant $f^{k,l} = f^{k,l-1} + ms(f^{k,l-1})$ jusqu'à convergence, avec ms défini comme en (8.21) où Σ_η^{-1} est remplacé par $\frac{1}{s^2}\Sigma_\eta^{-1}B$.
3. Calculer $u^k = \frac{1}{s}\Sigma_\eta^{-1/2}(p - Rf^{k-1})$ et $B^k = \text{diag}\{\varphi'(u_i^k)/(2u_i^k)\}_{i=1\dots N}$,
4. Si $\|f^k - f^{k-1}\| > \varepsilon$, incrémenter k , et aller en 2, sinon $f_{RMS} = f^k$.

Cet algorithme permet de gérer le modèle associant terme d'attache aux données robuste et *a priori* non paramétrique, sans recourir à aucune approximation. Les résultats expérimentaux obtenus sur la base COIL de l'Université de Colombia, ont permis de confirmer l'intérêt de notre modèle pour une application de reconnaissance [34]. En particulier, l'algorithme s'est montré plus performant que la méthode de Murase et Nayar [21]. Ce travail ouvre donc la voie à l'utilisation de modèles d'apparence arbitrairement complexes.

8.3 Conclusion

L'apparence des objets représente une information suffisante pour assurer leur détection et leur reconnaissance, comme le démontre le fonctionnement du système visuel humain. Identifier des formes dans des images réelles à l'aide de modèles d'apparence tels que nous les utilisons est, cependant, une tâche difficile de vision artificielle. Cela tient d'abord au fait qu'il s'agit essentiellement d'un problème à une seule classe. La modélisation se concentre sur la représentation de la classe des objets d'intérêt, tandis que la classe « autres » n'est pas prise en compte. Dans ce contexte, utiliser des modèles linéaires pour représenter des variétés arbitrairement complexes peut paraître voué à l'échec. Les méthodes que nous avons proposées mettent à profit un modèle *a priori* non paramétrique de la variable latente pour compenser ce point faible. Comme nous l'écrivions en introduction de [34], il s'agissait en quelque sorte de « probabiliser » l'idée de Murase et Nayar [21] : modéliser de façon précise la variété décrite par la variable latente dans l'espace propre (ce qui conduit à de meilleurs résultats en détection). Une autre difficulté tient à la complexité des scènes observées, entraînant des perturbations importantes des observations. L'utilisation de techniques robustes, semi-quadratiques, permet de s'affranchir de la présence de données erronées. Les méthodes présentées dans ce chapitre ont, pour l'instant, été validées sur des images synthétiques [8] ou acquises en conditions contrôlées [34]. Préalable indispensable à toute application pratique, une phase de validation sur images réelles, dans une application de reconnaissance des panneaux de signalisation, est prévue à court terme.

Bibliographie

- [1] P. BELHUMEUR, J. HESPAÑA, et D. KRIEGMAN. Eigenfaces vs. Fisherfaces : Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7) :711–720, juillet 1997.
- [2] A. BLAKE et A. ZISSERMAN. *Visual reconstruction*. MIT Press, Cambridge, USA, 1987.
- [3] P. CHARBONNIER. *Reconstruction d'image : régularisation avec prise en compte des discontinuités*. Thèse de doctorat, Université de Nice-Sophia Antipolis, France, septembre 1994.
- [4] Y. CHENG. Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8) :790–799, août 1995.
- [5] D. COMANICIU et P. MEER. Mean shift : A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5) :603–619, mai 2002.
- [6] R. DAHYOT. *Analyse d'images séquentielles de scènes routières par modèles d'apparence pour la gestion du réseau routier*. Thèse de doctorat, Université Louis Pasteur, Strasbourg I, 2001. Collections Etudes et Recherches des Laboratoires des Ponts et Chaussées, CR 31, LCPC.
- [7] R. DAHYOT, P. CHARBONNIER, et F. HEITZ. Robust visual recognition of colour images. Dans *IEEE Int. Conf. Computer Vision and Pattern Recognition*, volume 1, pages 685–690, juin 2000. CVPR 2000, Hilton Head Island, USA.
- [8] R. DAHYOT, P. CHARBONNIER, et F. HEITZ. Robust Bayesian detection using appearance-based models. *Pattern Analysis and Applications*, 7 :317–332, 2004.
- [9] R. DUDA, P. HART, et D. STORK. *Pattern classification*. Wiley-Interscience, seconde édition, 2001.
- [10] M. FASHING et C. TOMASI. Mean shift is a bound optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3) :471–474, mars 2005.
- [11] K. FUKUNAGA et L. HOSTETLER. The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21 :32–40, janvier 1975.
- [12] D. GEMAN et G. REYNOLDS. Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(3) :367–383, mars 1992.
- [13] D. GEMAN et C. YANG. Nonlinear image recovery with half-quadratic regularization and FFT's. *IEEE Transactions on Image Processing*, 4(7) :932–946, juillet 1995.
- [14] T. HASTIE et W. STUETZLE. Principal curves. *Journal of the American Statistical Association*, 84 :502–516, 1989.
- [15] T. HEBERT et R. LEAHY. A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Transactions on Medical Imaging*, 8(2) :194–202, juin 1989.

8. MODÈLES PROBABILISTES D'APPARENCE

- [16] C. HEINRICH. *Problèmes inverses, analyse statistique de formes et problèmes connexes*. Mémoire d'habilitation à diriger des recherches, Université Louis Pasteur de Strasbourg - Strasbourg I, octobre 2008.
- [17] P.J. HUBER. *Robust statistics*. John Wiley and Sons, New York, 1981.
- [18] A. JAGMOHAN, M. SINGH, et N. AHUJA. Dense stereo matching using kernel maximum likelihood estimation. Dans *17th International Conference on Pattern Recognition (ICPR'04)*, volume 3, pages 28–31, août 2004.
- [19] K. LANGE, D. R. HUNTER, et I. YANG. Optimization transfer using surrogate objective functions. *Journal of Computational and Graphical Statistics*, 9(1) :1–20, mars 2000.
- [20] B. MOGHADDAM et A. PENTLAND. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7) :696–710, juillet 1997.
- [21] H. MURASE et S. K. NAYAR. Visual learning and recognition of 3-D objects from appearance. *International Journal of Computer Vision*, 14(1) :5–24, janvier 1995.
- [22] R. PARK. Comments on optimal approximation of uniformly rotated images : Relationship between Karhunen-Loeve expansion and discrete cosine transform". *IEEE Transactions on Image Processing*, 11(3) :332–334, mars 2002.
- [23] S. ROWEIS et Z. GHAHRAMANI. A unifying review of linear Gaussian models. *Neural Computation*, 11(2) :305–345, 1999.
- [24] S. ROWEIS et L. SAUL. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500) :2323–2326, 2000.
- [25] B. SCHÖLKOPF, A. SMOLA, et K.-R. MÜLLER. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5) :1299–1319, juillet 1998.
- [26] M. SINGH, H. ARORA, et N. AHUJA. A robust probabilistic estimation framework for parametric image models. Dans T. PAJDLA et J. MATAS, éditeurs, *European Conference on Computer Vision, Lecture Notes in Computer Science*, volume 3021, pages 508–522. Springer, 2004.
- [27] L. SIROVICH et M. KIRBY. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4(3) :519–524, 1987.
- [28] J.-P. TAREL, S.-S. IENG, et P. CHARBONNIER. *Robust Lane Marking Detection by the Half Quadratic Approach*. Collections Etudes et Recherches des Laboratoires des Ponts et Chaussées, CR 49, LCPC, novembre 2007.
- [29] M. TIPPING et C. BISHOP. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B*, 61(3) :611–622, 1999.
- [30] M. E. TIPPING et C. M. BISHOP. Mixtures of probabilistic principal component analysers. *Neural Computation*, 11(2) :443–482, 1999.
- [31] M. TURK et A. PENTLAND. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1) :71–86, 1991.
- [32] T. VIK. *Non-Gaussian Statistical Appearance Models. Application to the Creation of a Probabilistic Atlas of Brain Perfusion in Medical Imaging*. Thèse de doctorat, Université Louis Pasteur - Strasbourg I, 2004. En Anglais.

BIBLIOGRAPHIE

- [33] T. VIK, F. HEITZ, et P. CHARBONNIER. Mean shift-based Bayesian image reconstruction into visual subspace. Dans *Proceedings of the 2003 IEEE International Conference on Image Processing (ICIP 2003)*, volume 1, pages 697–700, Barcelone, Espagne, septembre 2003.
- [34] T. VIK, F. HEITZ, et P. CHARBONNIER. Robust pose estimation and recognition using non-Gaussian modeling of appearance subspaces. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 29(5) :901–905, mai 2007.

8. MODÈLES PROBABILISTES D'APPARENCE

Chapitre 9

Estimation de la saillance visuelle

*Ludovic Simon,
Jean-Philippe Tarel,
Roland Brémond*

La signalisation routière et plus particulièrement celle verticale est le principal média de communication en direction des conducteurs. Malheureusement, tous les panneaux ne sont pas vus par tous les conducteurs, ce qui peut conduire à des situations accidentogènes. Il est donc important de contrôler que les panneaux de signalisation sont suffisamment saillants pour attirer l'attention.

La saillance d'un objet peut être définie comme le degré avec lequel cet objet attire l'attention visuelle d'un observateur. Ce qui attire l'attention du conducteur est lié à la fois à des facteurs psychologiques (motivations, tâche de conduite, etc) et aux caractéristiques photométriques, géométriques et sémantiques des objets de l'environnement routier (forme, couleurs, fond, signification, pertinence, etc). La perception des panneaux dépend des deux composantes principales de l'attention visuelle : la capture de l'attention (mécanisme ascendant ou *bottom-up*) et la recherche visuelle (mécanisme descendant, ou *top-down*) [5]. La capture de l'attention par un objet intrinsèquement très saillant (phénomène dit de *pop-out*) perd de son importance lorsque l'observateur est engagé dans une activité cognitive, comme l'est la tâche de conduite. Dans une tâche de recherche visuelle complexe, les phénomènes descendants, volontaires et liés aux connaissances a priori, sont prépondérants.

Comme la plupart des modèles actuels de calcul de la saillance de recherche sont limités à des situations de laboratoire, nous proposons un nouveau modèle qui s'applique à des images de scènes naturelles. Notre paradigme consiste dans un premier temps à construire un détecteur de l'objet d'intérêt pour trouver les zones saillantes qui sont susceptibles d'attirer l'attention du conducteur lors de la recherche d'un panneau. Ce détecteur donne donc des zones saillantes différentes selon la tâche qui est ciblée (par exemple recherche d'un sens interdit, recherche d'un panneau de sens unique...). Ce détecteur doit être conçu de façon assez différente de ceux présentés dans les autres chapitres, comme nous l'expliquons dans la partie 9.2. Dans un deuxième temps, notre paradigme consiste à lier la confiance en la détection à la saillance de recherche, comme nous le détaillons dans la partie 9.3. Le modèle proposé a été testé lors d'expériences psycho-visuelles telle celle décrite dans la partie 9.4. Cela ouvre la possibilité de concevoir des

systèmes d'aide au diagnostic de la saillance de la signalisation le long d'un réseau routier, à partir d'images prises avec un appareil numérique à bord d'un véhicule d'inspection.

9.1 La saillance visuelle

L'attention visuelle peut être considérée comme un processus à deux composantes : la capture (involontaire) de l'attention et la recherche visuelle [5]. Ces deux composantes sont en interaction complexe lors de la conduite automobile. La capture de l'attention dépend des caractéristiques visuelles propres de l'objet qui se distinguent fortement de son arrière-plan (*feature contrast*), et qui attirent l'attention de l'observateur indépendamment de la tâche en cours. Il s'agit d'un processus ascendant, qui est prépondérant dans une situation de laboratoire en recherche libre, par exemple, lorsqu'un observateur regarde une peinture abstraite. Au contraire, la recherche visuelle est un processus descendant. Elle est dirigée par une attention volontaire, qui dépend de l'expérience du conducteur et de sa connaissance de l'objet recherché. La recherche d'un détail spécifique sur une photo est un exemple de recherche visuelle pure.

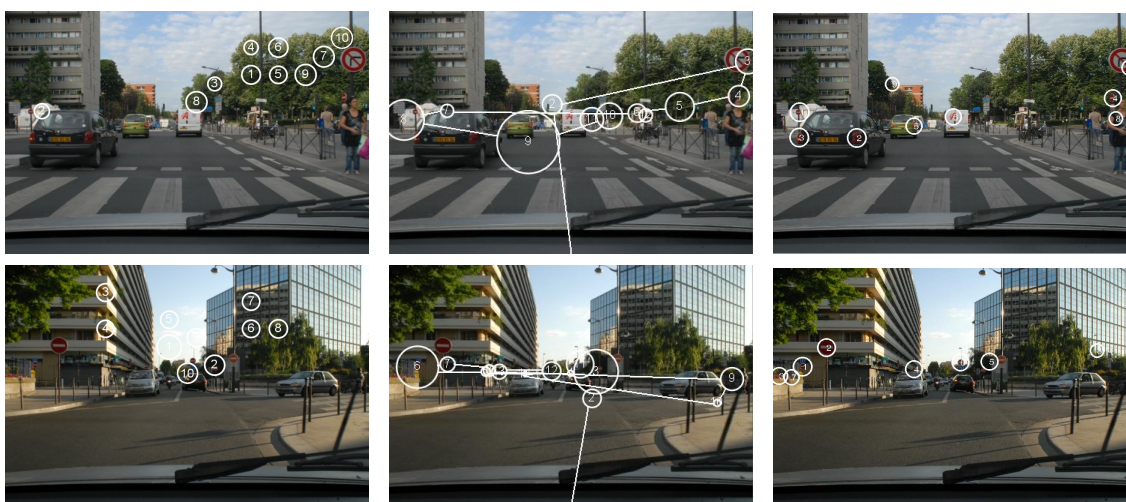


FIG. 9.1 – (à gauche) focalisations du regard prédites par le modèle standard de saillance *bottom-up* [3]. (au milieu) parcours oculaire d'un sujet lors de la recherche d'un panneau. (à droite) focalisations prédites par notre modèle de saillance de recherche des panneaux de sens interdit. Les focalisations de gauche et droite sont étiquetées dans l'ordre décroissant de leur saillance. Les fixations de l'image du milieu sont étiquetées dans l'ordre du chemin oculaire.

Bien que plusieurs modèles de calcul de la saillance *bottom-up* ont été proposés dans la dernière décennie, en raison de la complexité du comportement humain, ce n'est que très récemment que quelques modèles de calcul de la saillance de recherche ont été proposés [4]. Le modèle de saillance *bottom-up* le plus populaire a été proposé par Itti *et al.* [3, 2]. Cet algorithme calcule la carte de saillance, selon une approche ascendante, en se fondant sur la modélisation bas niveau du Système Visuel Humain (SVH) [6]. Ce modèle a été testé lors d'une expérience psycho-visuelle utilisant des données oculométriques pour connaître les zones de focalisation des sujets [13]. Leur conclusion est que ce modèle a une certaine pertinence lorsque les observateurs ont pour tâche de mémoriser les images présentées, mais qu'il échoue lorsque la tâche

9.2. DÉTECTION DES ZONES SAILLANTES

est de rechercher un objet particulier. La figure 9.1 illustre cette limite du modèle strictement *bottom-up* sur deux scènes routières : à gauche les focalisations prédites par le modèle [3] sont incorrectement réparties dans le ciel et les arbres, alors que le chemin oculaire observée parcourt essentiellement la ligne d'horizon, - cf. milieu de la figure 9.1. L'étude décrite dans [1] dans le cadre d'images routières va dans le même sens.

Afin de permettre la conception d'un système automatique d'estimation de la saillance des panneaux le long d'un réseau routier, un modèle de calcul de la saillance de recherche est donc nécessaire. Jusqu'à récemment, il n'y avait pas de modèle de calcul complet, seulement des modèles théoriques ou informatiques pouvant travailler uniquement sur des situations simples de laboratoire. Nous avons proposé un premier modèle du calcul de la saillance de recherche, basé sur l'apprentissage de l'aspect de l'objet d'intérêt dans les images. Dans un premier temps, nous avons abordé la détection des zones saillantes dans l'image [9, 10]. En nous appuyant sur un algorithme d'apprentissage statistique, l'algorithme proposé construit un modèle de l'apparence de chaque type de panneau. Cet apprentissage peut être mis en parallèle avec les connaissances sur la signalisation routière acquises par un conducteur lors de son apprentissage de la conduite. A partir de ce modèle d'apparence, un algorithme de détection est construit pour localiser les zones saillantes dans l'image lors d'une recherche visuelle. Ces zones sont définies comme des régions pouvant contenir un panneau. Dans un deuxième temps, la valeur de la saillance de chaque zone détectée est estimée en utilisant la confiance dans la détection et la taille de l'objet [12, 8]. La fonction qui lie la saillance, la confiance dans la détection et la taille de l'objet a été obtenue à partir des données issues d'une expérimentation psycho-visuelle.

9.2 Détection des zones saillantes

La recherche d'un panneau routier peut se traduire comme un problème de détection dans une image en termes de reconnaissance de formes.

L'aspect d'un panneau peut varier de façon importante du fait des variations d'éclairage, des conditions météorologiques et de l'usure des panneaux. Pour modéliser cette variabilité d'aspect, nous avons utilisé une technique d'apprentissage statistique. L'apprentissage est effectué à partir d'une base d'exemples positifs et négatifs, nommée base d'apprentissage. Chaque exemple positif est un échantillon de l'aspect du panneau, tandis que chaque exemple négatif est un échantillon de fond. La figure 9.2 montre quelques échantillons d'exemples positifs et négatifs pour le panneau de sens interdit. Chaque image de la base d'apprentissage est résumée par un vecteur de caractéristiques, nommé signature. A partir de cette base de signatures positives et négatives, l'algorithme d'apprentissage statistique nommé *Support Vector Machine* (SVM) [7] est en mesure de construire une frontière qui divise l'espace des signatures en parties positives et négatives. Le fait de travailler sur les signatures plutôt que sur les images permet d'accélérer la phase d'apprentissage. Après cette phase d'apprentissage, le classificateur qui en résulte est en mesure d'évaluer le label positif ou négatif de n'importe quelle nouvelle image. Le principal avantage du SVM, outre sa justification par la théorie de l'apprentissage statistique, est que la fonction de classification obtenue donne des valeurs continues et non des valeurs binaires comme d'autres algorithmes de classification par apprentissage (Adaboost). Cette valeur de classification peut-être associée à une confiance dans la classification : quand elle est supérieure à 1, la classification est positive avec une bonne confiance, quand elle est inférieure à -1,

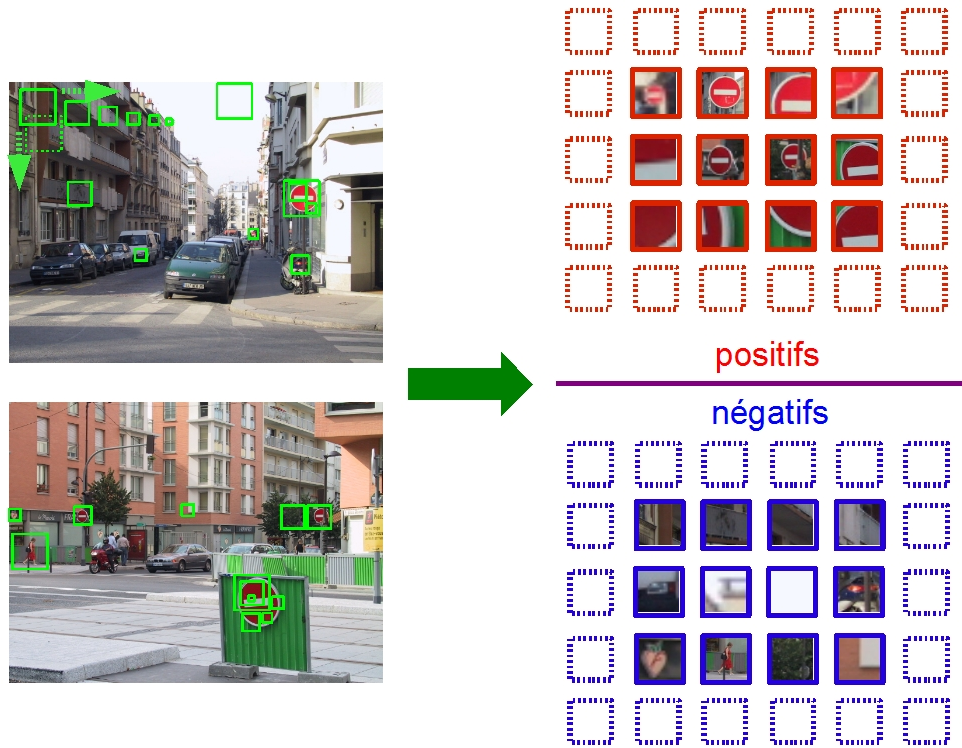


FIG. 9.2 – A partir d’images où les panneaux sont détournés manuellement, des imagerie contenant des sens interdits (en entier ou des portions importantes) sont extraites à différentes échelles. Une base d’imagerie ne contenant aucun sens interdit est aussi construite en parallèle.

la classification est négative avec une bonne confiance, et enfin entre -1 et 1 la classification est plus incertaine.

Ce classificateur est utilisé pour effectuer la détection des panneaux de signalisation dans chaque image. La détection consiste à parcourir l’image avec des fenêtres glissantes de différentes tailles, puis pour chaque imagerie extraite, d’utiliser les classifieurs pour évaluer si l’imagerie correspond à l’aspect du panneau à détecter ou non. Cela permet de construire une carte de possible présence des objets recherchés (ici, des panneaux de sens interdit). Ainsi, sur l’image originale 640×480 - cf. Fig. 9.3, six cartes sont produites aux échelles 60×60 , 40×40 , 30×30 , 20×20 , 16×16 et 10×10 pour la détection des sens interdits. Ces cartes sont fusionnées en calculant le maximum des six cartes après pondération par $1 + \log(\frac{s}{10})$, où s est la taille de la fenêtre glissante. Le résultat est présenté en bas à gauche de la figure 9.3. Après seuillage, une carte binaire de détection est construite - cf. Fig. 9.3 (en bas à droite). Le résultat obtenu est une carte des zones qui contiennent potentiellement le panneau recherché. Mais ce n’est pas un détecteur au sens classique, car il sélectionne également des régions qui attirent l’attention dans une recherche visuelle de panneaux sans être des panneaux. Ces zones sont appelées distracteurs, par exemple la personne avec une chemise rouge et blanche dans la figure 9.3. Nous avons considéré la carte binaire obtenue comme la carte des zones saillantes lors de la recherche.

Pour faire l’apprentissage avec le SVM, il est nécessaire de choisir une signature et la mesure de similarité entre signature nommée noyau. Ce choix est effectué de façon à obtenir la

9.2. DÉTECTION DES ZONES SAILLANTES

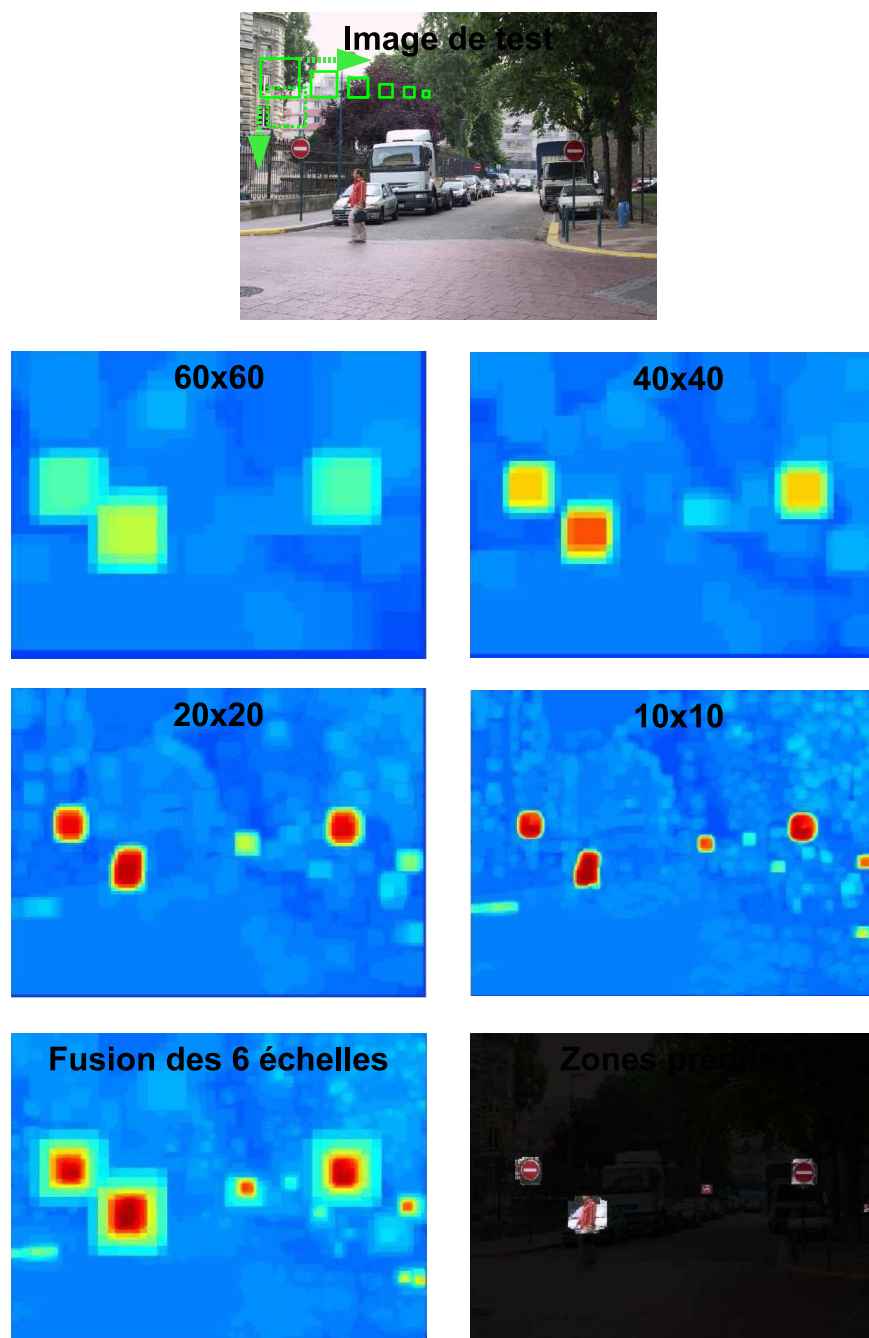


FIG. 9.3 – Détection de sens interdit utilisant une fenêtre glissante à différentes échelles. Les six cartes à différentes échelles sont fusionnées (en bas à gauche) et le résultat est seuillé (en bas à droite) pour obtenir la carte des zones considérées comme saillantes lors de la recherche d'un sens interdit.

meilleure adéquation possible entre les zones saillantes pour les observateurs et les cartes générées comme expliqué précédemment. Dans [10], les expériences qui nous ont permis de sélectionner le meilleur couple noyau et signature sont décrites. Le principal résultat est que l'histogramme des couleurs normalisées est assez pertinent lorsque l'objectif est de détecter les sens interdits. La forme semble moins importer pour ce panneau, mais peut-être est-elle importante pour d'autres types de panneaux. Le meilleur choix de noyau semble être le noyau de puissance, $k(x, x') = \|x - x'\|^\alpha$, qui n'a pas besoin d'un paramètre d'échelle car il permet une adaptation implicite à la densité locale des signatures, contrairement à la plupart des noyaux classiques. Dans la suite, nous avons utilisé $\alpha = 1$.

9.3 Estimation de la saillance des panneaux

Notre paradigme consiste à s'appuyer sur un détecteur de panneau fondé sur un algorithme d'apprentissage pour modéliser la variabilité de son apparence, et de définir la saillance de recherche comme une fonction de la confiance dans la présence de ce panneau dans une image. Ainsi, si un panneau est difficile à détecter par son aspect, la confiance dans la détection sera faible et sa saillance aussi. La saillance de recherche d'un panneau est une mesure de la facilité avec laquelle on le détecte dans une image.

La carte de confiance fusionnée est nommée Carte de Saillance Intrinsèque (CSI). Elle quantifie la confiance dans la détection du panneau mais elle tient peu compte de l'arrière plan autour de l'objet détecté. Afin d'en tenir compte, cette carte de confiance est corrigée en soustrayant sa moyenne locale calculée sur un disque d'angle 2° . Cette valeur a été obtenue expérimentalement [12]. La carte ainsi obtenue est appelée Carte de Saillance avec prise en compte du Fond (d'acronyme CSF dans la suite). Notons que la CSF n'est que peu dépendante de la taille de l'objet détecté, alors qu'il est connu que la taille joue un rôle important dans la saillance. Plus l'objet est petit, moins il est saillant lors d'une recherche visuelle. Les composantes connexes de pixels positifs sont donc extraites de la carte CSF. Pour chaque composante i , la valeur de la Carte de Saillance de Recherche (d'acronyme CSR dans la suite) est calculé comme suit :

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

où $CSF(i)$ est la valeur de la carte CSF moyennée sur la i -ème composante connexe et $A(i)$ est son aire. Il aurait été possible de définir la saillance de recherche sans appliquer de racine quatrième dans l'équation précédente. Nous avons préféré introduire cette racine afin d'avoir un lien linéaire entre notre définition de la CSR et les notes de saillance recueillies lors des expérimentations avec sujets. Le méthode de calcul de la saillance de recherche peut être résumée par la figure 9.4.

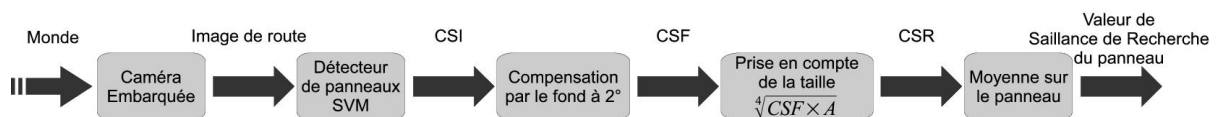


FIG. 9.4 – Schéma de calcul de la saillance de recherche CSR.

9.4. EXPÉRIMENTATIONS

Le domaine d'application du modèle proposé de saillance ne se limite pas à l'inspection de la saillance de la signalisation routière par un véhicule dédié, il peut également servir dans le cadre d'aides à la conduite, comme nous l'avons proposé dans [11].

9.4 Expérimentations

Afin de tester le modèle proposé, nous avons mené une expérimentation avec sujets dans une salle qui est contrôlée photométriquement. La figure 9.5 montre un sujet et au premier plan l'opérateur qui contrôle l'expérience. Le dispositif d'affichage du sujet est équipé d'un système de suivi du regard (de la marque SMI) pour enregistrer la position des fixations oculaires et leur durée. Le système a une précision de $0,5^\circ$ et une fréquence de 50Hz. L'écran de 19" est observé à une distance de 70 cm. Ainsi, les sujets voient la scène routière avec un angle visuel de 20° . Pour chaque sujet, 40 images routières ont été présentées aléatoirement, contenant un total de 76 panneaux de sens interdit. Ces images ont été sélectionnées de façon à contenir des panneaux d'apparences diverses dans des contextes différents, afin d'obtenir une grande variété des niveaux de saillance.

Un total de 40 sujets a effectué l'expérience, et 32 ont été retenus. Il leur a été demandé de se mettre mentalement en situation de conduite devant chaque image qui a été prise avec un point de vue proche de celui d'un conducteur. L'expérience a consisté en deux phases. Dans la première phase, les sujets ont été invités à compter les panneaux de sens interdit, sachant que les images disparaissaient au bout de 5 secondes. Dans la deuxième phase, les sujets étaient invités à noter la facilité de détection de chaque panneau entre 0 et 10. L'analyse combinée des fixations du regard et des réponses des sujets nous a permis de savoir, pour chaque sujet, quelles étaient les panneaux vus / non vus. Pour un panneau donné, le pourcentage moyen de détection par les sujets donne son Taux de Détection lors de la Recherche (d'acronyme TDR dans la suite). Dans la deuxième phase, les notes de chaque sujet ont été centrées et normalisées en variance en raison de la variabilité inter-sujets dans l'utilisation de l'échelle de score. Plus précisément, le Score Subjectif Standardisé (d'acronyme SSS dans la suite) est défini par :

$$SSS(i, j) = \frac{score_{i,j} - E_i(score_{i,j})}{\sqrt{E_i((score_{i,j} - E_i(score_{i,j}))^2)}} + 5$$

où i est l'index du sujet et j celui du panneau.

Nous avons d'abord vérifié que le taux de détection par les sujets TDR est relié au score subjectif SSS. Comme illustré dans la partie gauche de la figure 9.6, plus le SSS est grand, plus le TDR est proche de 1. Le lien fort entre le score subjectif SSS et le taux de détection par les sujets est cohérent avec le paradigme proposé : la saillance de recherche d'un objet est une évaluation de sa facilité de détection. De plus, la forme en sigmoïde de cette relation est cohérente avec les courbes psychologiques classiques de performance. Ensuite, nous avons étudié le lien entre le SSS et saillance de recherche estimée CSR. Comme le montre la figure 9.6 à droite, une relation linéaire peut être observée entre SSS et CSR. L'analyse statistique [12, 8] a montré que CSR explique 56% de la variance entre les panneaux, et 39% de la variance globale. La même analyse en utilisant simplement la racine quatrième de l'aire du panneau explique 46% de la variance entre les panneaux et 32% de la variance globale. Ainsi, le modèle proposé améliore de 18% l'estimation de la saillance qui serait uniquement basée sur la taille.

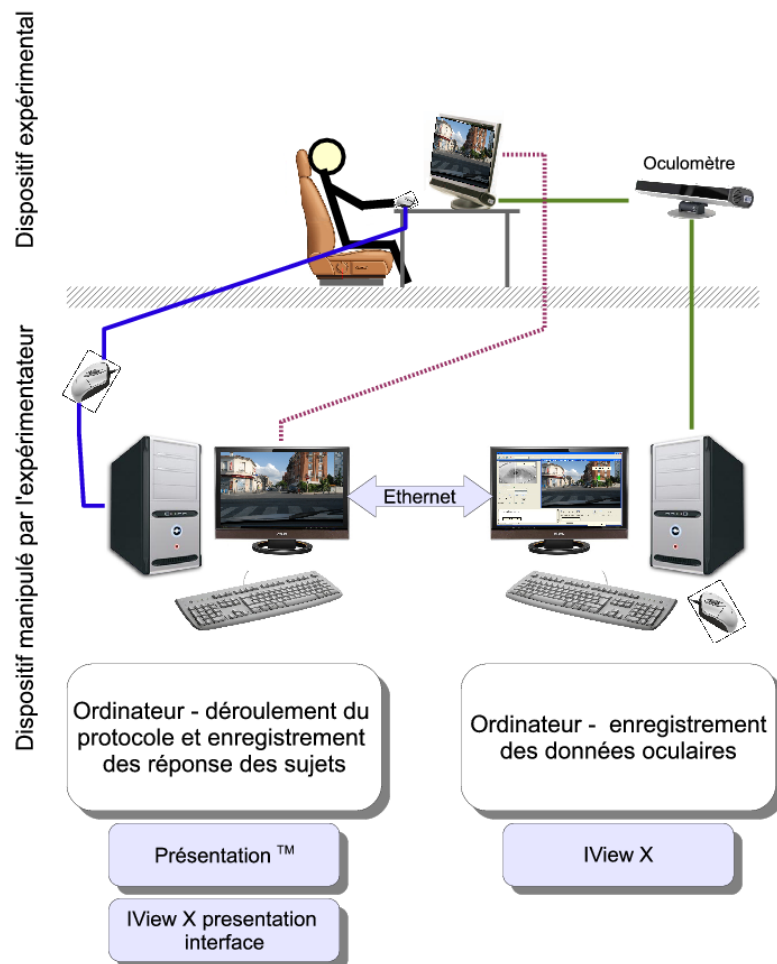


FIG. 9.5 – Salle d'expérience avec le système de suivi du regard et le schéma descriptif de la configuration matérielle détaillée dans [8].

9.4. EXPÉRIMENTATIONS

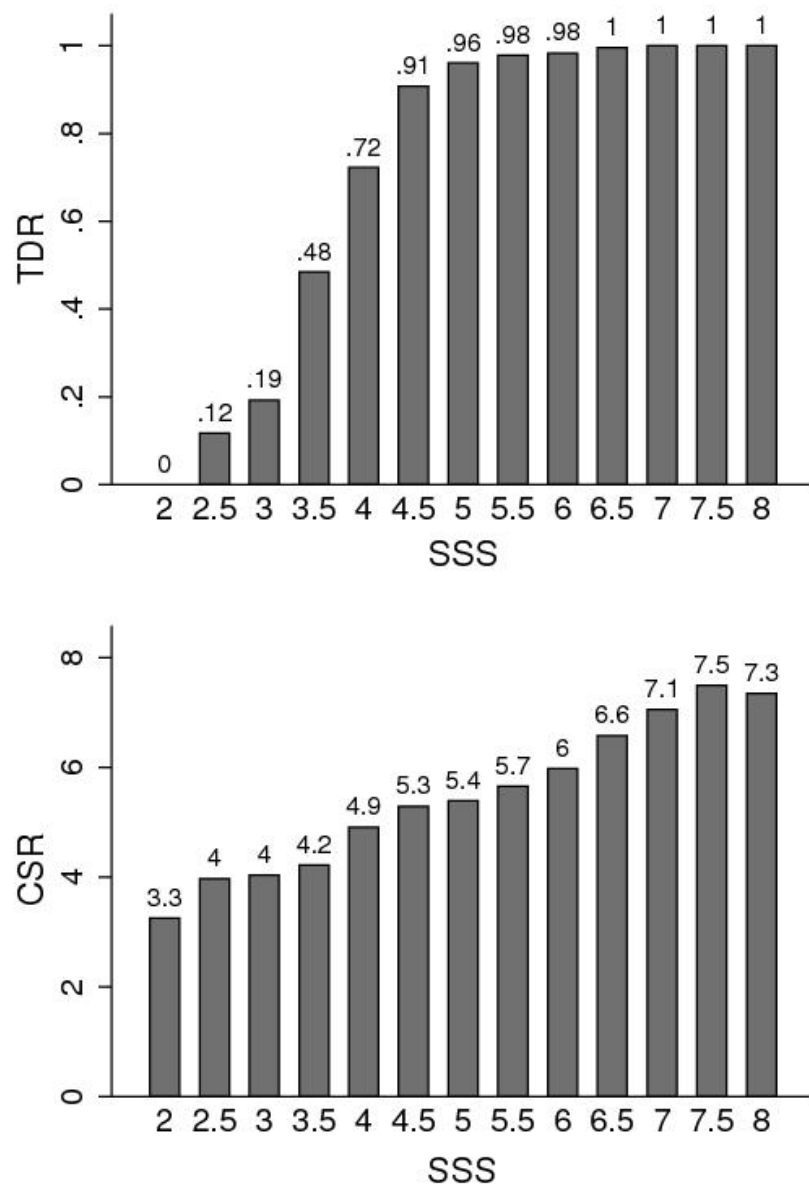


FIG. 9.6 – En haut, le lien entre le Score Subjectif Standardisé (SSS) et le Taux de Détection lors de la Recherche des panneaux (TDR). En bas, la relation entre le SSS et la saillance de recherche CSR estimée par notre modèle.

Le lien entre TDR et SSS implique que CSR est également corrélée au taux de détection TDR. Le TDR peut donc aussi être utilisé comme un indicateur objectif de la saillance de chaque panneau.

9.5 Conclusion

La plupart des modèles de calcul de la saillance visuelle sont limités aux processus ascendants (*bottom-up*), alors que la conception d'un système de diagnostic automatique de la saillance de la signalisation routière implique d'avoir un modèle de calcul de la saillance pendant la tâche de recherche visuelle. Nous avons donc proposé un paradigme original permettant de définir la saillance de recherche de façon objective. Cela nous a conduit à une méthode d'estimation de la saillance de recherche dans une image en utilisant un apprentissage par SVM de la variabilité des aspects possibles de l'objet recherché. On peut dire, sur la base de nos expériences psycho-visuelles, que ce modèle de calcul s'avère relativement bien corrélé au score donné par les sujets sur la facilité à les détecter. L'essentiel de ce travail a pris les panneaux de sens interdit comme exemple, il est donc nécessaire de tester nos conclusions et notre modèle sur d'autres types de panneaux routiers.

Bibliographie

- [1] R. BRÉMOND, J.-P. TAREL, H. CHOUKOUR, et M. DEUGNIER. La saillance visuelle des objets routiers, un indicateur de la visibilité routière. Dans *Proceedings of Journées des Sciences de l'Ingénieur (JSI'06)*, Marne la Vallée, France, 2006.
- [2] L. ITTI et C. KOCH. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40 :1489–1506, 2000.
- [3] L. ITTI, C. KOCH, et E. NIEBUR. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11) :1254–1259, 1998.
- [4] L. ITTI, G. REES, et J. TSOTSOS. *Neurobiology of Attention*. Elsevier, San Diego, USA, 2005.
- [5] E. I. KNUDSEN. Fundamental components of attention. *Annual Review Neuroscience*, 30 :57–78, 2007.
- [6] C. KOCH et S. ULLMAN. Shifts in selective visual attention : towards the underlying neural circuitry. *Human neurobiology*, 4(4) :219–227, 1985.
- [7] B. SCHÖLKOPF et A. SMOLA. *Learning with Kernels*. MIT Press, 2002.
- [8] L. SIMON. *Saillance de la signalisation verticale dans les images routières : étude de la faisabilité d'un outil de diagnostic*. Thèse de doctorat, Université Paris-VI, 2009.
- [9] L. SIMON, J.-P. TAREL, et R. BRÉMOND. A new paradigm for the computation of conspicuity of traffic signs in road images. Dans *Proceedings of the International Conference of the 26th session of the CIE (CIE'07)*, volume 2, pages D4–38–D4–41, Beijing, Chine, 2007.
- [10] L. SIMON, J.-P. TAREL, et R. BRÉMOND. Towards the estimation of conspicuity with visual priors. Dans *Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP'08)*, pages 323–328, Funchal, Portugal, 2008.
- [11] L. SIMON, J.-P. TAREL, et R. BRÉMOND. Alerting the drivers about road signs with poor visual saliency. Dans *Proceedings of IEEE Intelligent Vehicle Symposium (IV'2009)*, pages 48–53, Xian, Chine, 2009.
- [12] L. SIMON, J.-P. TAREL, et R. BRÉMOND. A computational model of visual search saliency for road signs diagnostic. Dans *Proceedings of European Conference on Visual Perception (ECVP'09)*, Regensburg, Allemagne, 2009.
- [13] G. UNDERWOOD et T. FOULSHAM. Visual saliency and semantic incongruity influence eye movements when inspecting pictures. *Quarterly Journal of Experimental Psychology*, 59(11) :1931–1949, 2006.

Chapitre 10

Conclusion générale

Dans ce recueil, nous avons présenté six méthodes de détection et de reconnaissance de la signalisation verticale « de police », issues des travaux de recherche de quatre équipes du RST dans ce domaine. Ces approches couvrent les différentes catégories de méthodes proposées dans la littérature récente du domaine, dont une vision synthétique a été proposée au **chapitre 2**. Ainsi, les algorithmes des chapitres 3, 4 et 7, se fondent sur l'exploitation combinée des informations de couleur et de forme, soit par une technique de vote univarié (**chapitre 3**), soit par un ajustement géométrique robuste suivi d'une étape de rectification perspective et de reconnaissance par corrélation (**chapitre 4**), soit encore par ajustement d'un modèle déformable selon des critères de gradient et de couleur (**chapitre 7**). La méthode décrite **chapitre 5** exploite principalement l'information de forme, à l'aide d'un algorithme de vote bivarié. Enfin, les deux derniers chapitres reposent sur des méthodes statistiques, soit à partir de modèles globaux de l'apparence des objets (**chapitre 8**), soit en procédant à une classification grâce à un *Support Vector Machine (SVM)* (**chapitre 9**).

Chacune des méthodes proposées a fait l'objet d'évaluations individuelles sur des bases d'images différentes. Cette étape est nécessaire, à la fois pour le réglage des paramètres des méthodes et pour la quantification des performances de celles-ci. Cet aspect n'est pas toujours très bien détaillé dans les publications que l'on trouve dans le domaine. Par ailleurs, les évaluations sont bien souvent menées sur des bases contenant au moins un panneau par image, ce qui ne correspond pas à la réalité des applications. Au contraire, nous avons mis en place des séquences d'images de test, accompagnées de références établies manuellement, ou « vérité-terrains ». Ces dernières couvrent des itinéraires représentatifs, en zone urbaine aussi bien qu'en rase campagne. A l'aide de ces bases d'images, nous avons effectué des comparaisons quantitatives entre les trois premières méthodes proposées, selon des critères objectifs. La prochaine mise en ligne de ces bases sur les sites du projet iTOWNS, du LCPC et de l'IGN constituera donc une contribution à part entière aux travaux du domaine, en offrant une référence à disposition de la communauté.

D'une manière générale, l'ensemble des méthodes proposées atteint de très bonnes performances de détection et de reconnaissance et ces travaux de recherche ont conduit à plusieurs publications. Ces résultats permettent d'envisager dès à présent des applications pratiques, que ce soit dans le cadre du projet ANR iTOWNS, ou dans le développement de logiciels d'ex-

ploitation des images numériques tels que le logiciel LPC IREVE¹ ou la chaîne d'exploitation ROUTE3D développée par le laboratoire MATIS. Récemment, l'entreprise VIAMETRIS, qui développe ses propres outils de recueil de données routières, a repris et amélioré la méthode décrite au **chapitre 7**. L'utilisation de ces algorithmes, avec une contrainte temps réel dans des systèmes embarqués d'aide à la conduite est également envisageable. L'analyse de ces résultats suggère également un certain nombre de pistes d'amélioration et de perspectives de recherche, que nous allons maintenant détailler.

Auparavant, notons que l'intérêt des méthodes proposées dépasse souvent le cadre strict de l'inventaire de la signalisation. Ainsi, les algorithmes d'optimisation par méta-heuristiques biologiques (**chapitre 7**) sont également appliqués dans le domaine de l'acoustique par l'ERA 32. Les méthodes de détection par modèles d'apparence décrites au **chapitre 8** ont été employées par le LSIIT en imagerie médicale. Par ailleurs, les travaux sur l'estimation robustes initiés dans ce cadre ont débouché sur des algorithmes efficaces de détection de la signalisation horizontale proposés par le LEPSiS en collaboration avec l'ERA 27. Enfin, les outils décrits au **chapitre 9** sont utilisés pour mieux comprendre les mécanismes de perception visuelle en situation de conduite.

L'expérimentation croisée des trois premières méthodes proposées a montré qu'elles atteignent des performances relativement similaires, en termes de bonnes détection et de temps de calcul. Le traitement d'une image prend de une à deux secondes. Certains algorithmes, implantés dans l'environnement Matlab®, peuvent encore être optimisés. Par contre, les méthodes se différencient par leur principe, chacune possédant des atouts et des limitations propres, ce qui se traduit par des natures et des taux de fausses alarmes différents. Une première piste d'amélioration consiste donc à mettre à profit la complémentarité des méthodes, afin d'obtenir un outil de détection le plus efficace possible.

Afin d'aller vers des applications pratiques de ces algorithmes, il est nécessaire de mettre en place et de valider la phase d'identifications de la signalisation. Pour l'instant, cette problématique n'a été traitée que dans deux des méthodes proposées (cf. **chapitre 4** et **chapitre 8**). On peut décomposer l'identification en deux étapes : la classification des objets détectés selon leur forme et la reconnaissance du pictogramme. L'identification du type de forme peut être effectuée à partir de critères tels que le degré de compatibilité décrit au **chapitre 4**. Dans la littérature, des algorithmes de classification, fondés par exemple sur les réseaux de neurones ou les machines à vecteur support (SVM) ont également été proposés. Il semble envisageable d'étendre les méthodes de détection par SVM décrits au **chapitre 9**, pour construire une chaîne complète de détection et de reconnaissance. En ce qui concerne la reconnaissance du pictogramme, il est nécessaire de prendre en compte les distorsions perspectives, ou de les corriger comme cela est proposé au **chapitre 4** avant d'appliquer, par exemple, des méthodes de corrélation. Ces dernières étant relativement sensibles aux occultations et aux dégradations, l'usage de méthodes robustes fondées sur l'utilisation de l'apparence globale du **chapitre 8** serait particulièrement bien indiqué. Un autre axe d'amélioration à court terme consistera à effectuer un suivi temporel de l'objet détecté dans la séquence d'images de scènes routières. Il peut être intéressant de confirmer ou d'infirmer la présence et le type d'un panneau en fonction des résultats obtenus sur les images précédentes ou suivantes. Enfin, il est nécessaire de reconstruire la localisation 3D du panneau dans le référentiel de la route par triangulation photogrammétrique à

¹IREVE : logiciel d'Imagerie Routière, Étalonnage, Visualisation et Exploitation, commercialisé par le LCPC

partir de la base stéréoscopique et/ou de la séquence d'images. Cette reconstruction fait l'objet de travaux en cours au laboratoire MATIS.

Notons que, toujours sur le plan pratique, il sera nécessaire à terme d'étendre et de valider les algorithmes de détection et de reconnaissance aux panneaux monochromatiques de fin de prescription (des travaux sont actuellement en cours à l'ERA 27 sur ce sujet), aux panneaux d'entrée et de sortie d'agglomération, aux panneaux directionnels, aux panneaux ainsi qu'aux autres types de signalisation verticale (signaux lumineux de circulation par exemple).

A moyen terme, nous souhaitons étudier la prise en compte le contexte de l'objet dans la scène routière. A ce sujet deux pistes peuvent être envisagées. La première est de considérer l'environnement spatial interne dans l'image en couplant la détection du panneau à la détection d'autres objets (chaussées, marquages...) pour valider ou non le résultat de l'algorithme. La seconde consiste à apporter une information sur la nature même de la scène routière (urbain/rural, 2×2 voies/chaussée bidirectionnelle) permettant ainsi de donner une indication sur le type de panneaux potentiellement présents. On peut également envisager d'adapter les paramètres des détecteurs au type de scène, le contexte urbain étant, par exemple, plus difficile que l'analyse en rase campagne du fait de la présence de nombreux distracteurs. On pourra pour cela exploiter les informations extraites des séquences d'images ou bien utiliser l'ensemble des données disponibles dans le Référentiel à Grande Échelle de l'IGN. Enfin, l'information tridimensionnelle n'est actuellement pas prise en compte dans l'étape de détection. Lorsqu'elle est disponible, comme c'est le cas dans le projet iTOWNS, où les données sont fournies par le véhicule Stéréopolis de l'IGN, équipé de capteurs LASER et de caméras de stéréovision, cette information pourrait s'avérer discriminante pour la détection. Elle est également intéressante pour l'étude du contexte des objets.

Ces deux dernières perspectives s'inscrivent dans une orientation commune à nos équipes, qui souhaitent travailler à une analyse plus complète des scènes routières, à partir de données 3D et d'images. Au-delà de la détection de la chaussée et de l'analyse de la signalisation horizontale et verticale, il s'agit d'extraire les éléments de guidage, de balisage, d'éventuels obstacles en rives, mais aussi de découper les itinéraires en unités paysagères et sémantiques. L'objectif est de fournir des outils pour une meilleure connaissance de l'infrastructure, en termes de cartographie et d'aménagement mais aussi de visibilité et de lisibilité, contribuant ainsi à une meilleure gestion des infrastructures et à l'amélioration de la sécurité routière.

Liste des auteurs

| Nom | organisme | chapitre |
|---------------------|--|-----------|
| Rachid Belaroussi | LCPC Paris, LEPSIS rachid.belaroussi@lcpc.fr | 5,6 |
| Roland Brémond | LCPC Paris, LEPSIS roland.bremond@lcpc.fr | 9 |
| Pierre Charbonnier | LRS, ERA 27 « Imagerie - Méthodes optiques » pierre.charbonnier@developpement-durable.gouv.fr | 2,3,6,7,8 |
| Rozenn Dahyot | School of Computer Science and Statistics dahyotr@tcd.ie | 8 |
| Guillaume Dutilleux | LRS, ERA 32 « Acoustique » guillaume.dutilleux@developpement-durable.gouv.fr | 7 |
| Philippe Foucher | LRS, ERA 27 « Imagerie - Méthodes optiques » philippe.foucher@developpement-durable.gouv.fr | 2,3,6 |
| Fabrice Heitz | Université de Strasbourg, LSIIT fabrice.heitz@ensps.u-strasbg.fr | 8 |
| Nicolas Paparoditis | IGN, MATIS nicolas.paparoditis@ign.fr | 4,6 |
| Ludovic Simon | CETE Ile-de-France lsimon@developpement-durable.gouv.fr | 9 |
| Bahman Soheilian | IGN, MATIS bahman.soheilian@ign.fr | 4,6 |
| Jean-Philippe Tarel | LCPC Paris, LEPSIS jean-philippe.tarel@lcpc.fr | 2,5,6,9 |
| Torbjørn Vik | Philips Technologie torbjoern.vik@philips.com | 8 |

Laboratoire Central des Ponts et Chaussées (LCPC)

Université Paris-Est
LEPSiS, UMR LCPC-INRETS
58, Boulevard Lefebvre
75732 Paris Cedex 15
France

Laboratoire Régional des Ponts et Chaussées de Strasbourg (LRS)

CETE de l'Est
11, rue Jean Mentelin - BP 9
67035 Strasbourg cedex
France

Institut Géographique National (IGN)

SR/Laboratoire MATIS
73, avenue de Paris
94165 Saint Mandé
France

CETE Ile-de-France

DREIF/CETE Ile-de-France/MO/TS
Site de Trappes en Yvelines (78)
12 Rue Teisserenc de Bort
78190 Trappes cedex
France

Université de Strasbourg

Laboratoire des Sciences de l'Image, de l'Informatique et de la Télédétection (LSIIT)
Pôle API
Boulevard Sebastien Brant - BP 10413
67412 Illkirch cedex
France

School of Computer Science and Statistics

Lloyd Institute
Trinity College Dublin
College Green, Dublin 2,
Irlande

Philips Technologie GmbH Forschungslaboratorien

Roengenstr. 24-26,
22335 Hambourg,
Allemagne

