

**Information visuelle multirésolution pour l'estimation de la  
vitesse du trafic routier**

par

Olivier Godin

Mémoire présenté au Département d'informatique  
en vue de l'obtention du grade de maître ès sciences (M.Sc.)

**FACULTÉ DES SCIENCES  
UNIVERSITÉ DE SHERBROOKE**

Sherbrooke, Québec, Canada, mai 2013



Library and Archives  
Canada

Published Heritage  
Branch

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque et  
Archives Canada

Direction du  
Patrimoine de l'édition

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file Votre référence*

*ISBN: 978-0-494-95156-9*

*Our file Notre référence*

*ISBN: 978-0-494-95156-9*

#### NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

#### AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Le 17 mai 2013

*le jury a accepté le mémoire de Monsieur Olivier Godin  
dans sa version finale.*

Membres du jury

Professeure Marié-Flavie Auclair-Fortier  
Directrice de recherche  
Département d'informatique

Professeur Nizar Bouguila  
Codirecteur de recherche  
Université Concordia

Professeur Pierre-Marc Jodoin  
Évaluateur interne  
Département d'informatique

Professeur Richard Egli  
Président rapporteur  
Département d'informatique



*Pas le temps de niaiser.*

# SOMMAIRE

La hausse du taux d'occupation des routes rend nécessaire l'utilisation de dispositifs visant à surveiller et analyser la circulation des utilisateurs du réseau routier. Bien souvent, des réseaux de surveillance par caméras sont déjà en place sur les routes. L'information visuelle est donc déjà disponible, mais est peu utilisée pour l'analyse automatisée du trafic.

L'objectif de ce travail est de mettre au point une méthode automatisée basée sur l'information visuelle multirésolution permettant d'évaluer la fluidité de la circulation à plusieurs niveaux de précision. En effet, si une analyse manuelle de l'information visuelle peut permettre de dégager certaines tendances sur l'état du trafic routier, une approche automatisée permettra d'évaluer plus précisément plusieurs données sur le mouvement des véhicules.

En tout temps, le déplacement moyen dans la scène sera évalué, afin d'obtenir un indice de fluidité global. Tant que la circulation demeure fluide, il n'est pas nécessaire d'estimer la vitesse du trafic avec plus de précision. Toutefois, si cet indice descend sous un certain seuil, l'analyse sera raffinée pour fournir le déplacement moyen par direction, en utilisant une résolution deux fois supérieure à la précédente. Finalement, à partir des images à pleine résolution, il sera possible d'obtenir une mesure du déplacement de chaque véhicule.

**Mots-clés :** vision par ordinateur, analyse de la vidéo, soustraction de fond, flux optique, analyse multirésolution, estimation du mouvement, vitesse du trafic routier.

# REMERCIEMENTS

Je tiens d'abord à remercier ma directrice de recherche, Marie-Flavie Auclair-Fortier, pour son support moral et financier, ses conseils, et surtout la patience infinie<sup>8</sup> dont elle a fait preuve tout au long de mes études supérieures. Je souhaite aussi remercier mon codirecteur Nizar Bouguila pour les idées amenées au projet et les commentaires constructifs. J'aimerais également remercier les autres membres du jury, M. Richard Egli et M. Pierre-Marc Jodoin pour leurs commentaires et critiques pertinentes.

Merci à mes collègues de travail, de même que ceux du laboratoire Moivre pour les conversations, sérieuses ou non, le midi à la cafétéria. Je destine aussi un merci bien spécial aux membres de la promotion FordTaurus du baccalauréat en imagerie et médias numériques. Vous êtes des amis spectaculaires.

Je souhaite aussi tout particulièrement remercier Francis Letarte-Lavoie pour son support moral et psychologique. Il ne faudrait pas que nos patrons sachent combien de temps on passe à s'écrire d'insignifiants courriels sur les heures de travail.

Finalement, le plus gros et le plus important des remerciements va à mes parents, Nicole et Alain. Merci de m'avoir toujours soutenu. Vous êtes les meilleurs parents de l'univers et du cosmos.



# TABLE DES MATIÈRES

<b>SOMMAIRE</b>	<b>v</b>
<b>REMERCIEMENTS</b>	<b>vii</b>
<b>TABLE DES MATIÈRES</b>	<b>viii</b>
<b>LISTE DES TABLEAUX</b>	<b>xi</b>
<b>LISTE DES FIGURES</b>	<b>xii</b>
<b>INTRODUCTION</b>	<b>1</b>
<b>CHAPITRE 1 — Mesure de la vitesse du trafic routier</b>	<b>4</b>
1.1 Mesure de la vitesse sans information visuelle . . . . .	6
1.1.1 Boucles à induction . . . . .	6
1.1.2 Mesures de temps de vol . . . . .	8
1.2 Mesure de la vitesse avec information visuelle . . . . .	10

1.2.1	Calibration de caméra pour l'estimation de la vitesse réelle . . . . .	12
1.2.2	Mesure de la vitesse par l'analyse globale du trafic routier . . . . .	14
1.2.3	Mesure de la vitesse par suivi de véhicules . . . . .	16
 <b>CHAPITRE 2 — Vitesse du trafic routier par information visuelle multi- résolution</b>		<b>20</b>
2.1	Détection des véhicules par segmentation de l'arrière-plan . . . . .	23
2.1.1	Détection des véhicules par modélisation statistique . . . . .	24
2.1.2	Détection des véhicules par analyse de la texture . . . . .	26
2.1.3	Évaluation des régions d'intérêt . . . . .	29
2.2	Estimation du mouvement . . . . .	31
2.2.1	Calcul du flux optique . . . . .	32
2.2.2	Flux optique et trafic routier . . . . .	36
2.2.3	Analyse multirésolution du flux optique . . . . .	38
 <b>CHAPITRE 3 — Résultats expérimentaux</b>		<b>52</b>
3.1	Détection des véhicules en mouvement . . . . .	52
3.1.1	Méthodologie d'évaluation . . . . .	52
3.1.2	Résultats et analyse . . . . .	57
3.2	Estimation du mouvement par une approche multirésolution du flux optique	73
3.2.1	Analyse du flux optique en résolution R4 . . . . .	75
3.2.2	Analyse du flux optique en résolution R2 . . . . .	87

3.2.3	Analyse du flux optique en résolution R1 . . . . .	93
	<b>CONCLUSION ET PERSPECTIVES</b>	<b>98</b>
	<b>BIBLIOGRAPHIE</b>	<b>100</b>

# LISTE DES TABLEAUX

3.1	Données d'évaluation de la performance d'un classificateur . . . . .	53
3.2	Performances moyennes pour la série <i>Base</i> [20] . . . . .	62
3.3	Influence du nombre de distributions gaussiennes sur les performances de GMM-Lee sur la série <i>Base</i> [20] . . . . .	68
3.4	Performances moyennes ( $K = 2$ ) pour la série <i>Arrière-plan dynamique</i> [20]	70
3.5	Performances moyennes ( $K = 2$ ) pour la série <i>Vibration de caméra</i> [20] .	71
3.6	Performances moyennes ( $K = 2$ ) pour la série <i>Mouvement intermittent</i> [20]	73

# LISTE DES FIGURES

1.1	Système de boucle à induction typique . . . . .	6
1.2	Mesure de la vitesse d'un véhicule avec deux boucles à induction . . . . .	7
1.3	Principe de fonctionnement d'un radar à micro-ondes . . . . .	9
1.4	Mesure de vitesse avec un laser . . . . .	10
1.5	Positions relatives des repères caméra et monde . . . . .	13
1.6	Profil d'intensité d'une voie . . . . .	15
1.7	Graphe associé aux véhicules en mouvement dans deux images . . . . .	17
2.1	Analyse temporelle de chaque pixel d'une séquence. . . . .	25
2.2	Construction du code LBP d'un pixel . . . . .	28
2.3	Évaluation des régions d'intérêt. . . . .	31
2.4	Code de couleurs pour la visualisation du flux optique . . . . .	37
2.5	Niveaux de résolution de la pyramide vidéo. . . . .	39
2.6	Processus d'estimation de la vitesse moyenne dans une scène. . . . .	41
2.7	Évaluation du nombre de directions de trafic. . . . .	42

2.8	Histogramme de mouvement . . . . .	43
2.9	Estimation de la vitesse moyenne par direction . . . . .	44
2.10	Processus d'estimation de la vitesse moyenne par direction. . . . .	45
2.11	Construction d'un histogramme pour chaque région en mouvement. . . . .	47
2.12	Nombre de véhicules par région sur la séquence <i>highway</i> . . . . .	48
2.13	Séquence <i>streetLight</i> . . . . .	49
2.14	Processus d'estimation de la vitesse par véhicule. . . . .	49
2.15	Organigramme de la méthode d'estimation de la vitesse du trafic automobile. . . . .	51
3.1	Influence du choix de $P$ sur LBP-Satpathy . . . . .	58
3.2	Changements d'illumination avec LBP-Satpathy (séquence <i>highway</i> ) . . . . .	60
3.3	Changements d'illumination avec LBP-Satpathy (séquence <i>office</i> ) . . . . .	60
3.4	LBP-Satpathy sur la séquence <i>highway</i> . . . . .	63
3.5	LBP-Satpathy sur la séquence <i>office</i> . . . . .	63
3.6	LBP-Satpathy sur la séquence <i>pedestrians</i> . . . . .	63
3.7	LBP-Satpathy sur la séquence <i>PETS2006</i> . . . . .	64
3.8	GMM-Lee pour deux valeurs de $\alpha$ sur la séquence <i>PETS2006</i> . . . . .	66
3.9	Changement brusque d'illumination sur GMM-Lee (séquence <i>highway</i> ) . . . . .	67
3.10	Changement brusque d'illumination sur GMM-Lee (séquence <i>office</i> ) . . . . .	67
3.11	Changement brusque d'illumination sur GMM-Lee (séquence <i>highway</i> ) . . . . .	67
3.12	Changement brusque d'illumination sur GMM-Lee (séquence <i>office</i> ) . . . . .	68

3.13	GMM-Lee sur la séquence <i>fall</i> . . . . .	71
3.14	GMM-Lee sur la séquence <i>traffic</i> . . . . .	72
3.15	GMM-Lee sur la séquence <i>sofa</i> . . . . .	74
3.16	Vitesse moyenne sur la séquence <i>streetLight</i> . . . . .	76
3.17	Vitesse moyenne sur la séquence <i>highway</i> . . . . .	76
3.18	Déplacement moyen pour la séquence <i>MIT Traffic</i> (R4) [53] . . . . .	77
3.19	Histogramme de mouvement de la figure 3.16b . . . . .	78
3.20	Nombre de directions de trafic sur la séquence <i>streetLight</i> . . . . .	80
3.21	Nombre de directions de trafic sur la séquence <i>MIT Traffic</i> . . . . .	81
3.22	Directions de trafic sur la séquence <i>MIT Traffic</i> . . . . .	82
3.23	Nombre de directions de trafic sur la séquence <i>MIT Traffic</i> . . . . .	83
3.24	Directions de trafic sur la séquence <i>MIT Traffic</i> . . . . .	84
3.25	Nombre de directions de trafic sur la séquence <i>MIT Traffic</i> . . . . .	85
3.26	Directions de trafic sur la séquence <i>MIT Traffic</i> . . . . .	86
3.27	Évaluation locale du flux optique sur la séquence <i>highway</i> . . . . .	87
3.28	Comparaison des flux optiques sur la séquence <i>highway</i> . . . . .	88
3.29	Comparaison des flux optiques sur la séquence <i>streetLight</i> . . . . .	88
3.30	Vecteurs de flux optique de la figure 3.28 . . . . .	89
3.31	Vecteurs de flux optique de la figure 3.29 . . . . .	90
3.32	Évaluation du déplacement moyen par direction . . . . .	90

3.33	Nombre de directions de trafic sur la séquence <i>MIT Traffic</i> . . . . .	91
3.34	Déplacement moyen pour la séquence <i>MIT Traffic</i> (R2) [53] . . . . .	92
3.35	Flux optique sur la séquence <i>streetLight</i> . . . . .	94
3.36	Extrait de la séquence <i>streetLight</i> . . . . .	94
3.37	Proportion des pixels éliminés par la soustraction de fond en résolution R1	97
3.38	Temps d'exécution en sec. pour le calcul du flux optique en résolution R1	97



# INTRODUCTION

L'augmentation rapide du nombre de véhicules sur les routes au cours des années a rendu nécessaire la mise en place de systèmes visant à assurer la circulation plus fluide des utilisateurs du réseau routier. En raison des coûts qu'ils engendrent, les problèmes de congestion routière sont à l'origine du souhait des autorités d'en savoir davantage sur l'état du trafic routier [45]. Un des paramètres que l'on souhaite évaluer est la vitesse de la circulation automobile afin d'être à l'affût, par exemple, d'embouteillages anormaux ou d'accidents.

De nombreuses approches existent pour arriver à produire une estimation de la vitesse du flot routier. Les méthodes existantes vont de l'utilisation de détecteurs magnétiques placés sous la route aux mesures de temps de vol, en passant par l'analyse de séquences vidéo ou d'ondes sonores. En plus des problèmes inhérents à la méthode utilisée, les différentes techniques existantes font face à des problèmes communs tels que la nécessité de maintenir de bas temps de calcul, afin de permettre un traitement de l'information en temps réel.

De plus, la tendance vers l'automatisation des services engendre de nouveaux besoins dans tous les domaines. Dans le secteur routier, il suffit de penser à la démocratisation des systèmes d'assistance à la navigation (GPS) ou aux radars photo qui distribuent des contraventions pour excès de vitesse sans intervention humaine. Ces deux applications

pourraient grandement bénéficier d'une meilleure évaluation automatisée de la circulation. Dans cette optique, la capacité à évaluer efficacement la vitesse du trafic routier devient un enjeu très important.

Toutefois, avant d'en arriver là, quelques notions doivent être définies. Dans un premier temps, qu'entend-on par « *vitesse du trafic* » ? La vitesse du trafic routier décrit la fluidité du flot de véhicules circulant sur une portion de route [54]. Pour les méthodes d'estimation de la vitesse avec information visuelle, cette portion de la route correspond à un sous-ensemble du champ visuel du capteur (portion de route visible, voie de circulation, ligne de pixels perpendiculaire à la route, etc.). L'information sur la vitesse du trafic sera utilisée pour détecter les situations de ralentissement, comme les embouteillages et les accidents.

Lorsqu'il est question d'estimer cette vitesse, deux types de mesure sont considérés : les mesures globales et les mesures locales. Les mesures globales sont des indices qui considèrent le flot de véhicules comme un tout. Les informations recueillies servent à caractériser la circulation dans son ensemble, sur une portion de route. Les mesures locales servent davantage à quantifier le mouvement et le comportement routier (variation de la vitesse, changements de voies) de chaque véhicule. À partir de mesures locales, on peut obtenir des informations globales, telles que la moyenne et l'écart-type des vitesses individuelles.

Dans le cadre de ce travail, nous proposons d'utiliser l'information visuelle provenant de caméras de surveillance pour fournir une estimation de la vitesse du trafic routier. Celui-ci s'inscrit dans le cadre d'un projet plus large d'analyse du comportement routier. L'objectif global du projet est de mettre au point une méthode automatisée basée sur l'information visuelle permettant d'évaluer la fluidité du trafic routier afin de détecter, d'identifier et d'anticiper les problèmes. À terme, cette méthode devra être indépendante de la position et de l'orientation de la caméra, des conditions climatiques et de la lumi-

nosité dans la scène. À l'intérieur de ce vaste projet, nous nous attardons dans le présent travail à proposer une technique d'évaluation de la fluidité du trafic routier afin de déceler les ralentissements à partir de l'information visuelle, sans toutefois se préoccuper des conditions d'acquisition. La méthode proposée est non intrusive, puisque les données vidéo sont déjà disponibles là où des caméras de surveillance sont installées, et demande des ressources technologiques réduites, puisqu'un ordinateur personnel suffit à la tâche d'analyse des séquences vidéo. Les investissements requis ne sont donc pas importants.

Dans un premier temps, nous présentons un ensemble de techniques existantes pour mesurer la vitesse du trafic routier. Nous nous attardons d'abord aux approches ne faisant pas appel à l'information visuelle, puis nous enchaînons avec une étude de plusieurs méthodes basées sur l'analyse des séquences vidéo provenant de systèmes de surveillance du réseau routier. Cela nous amène ensuite à proposer une méthode reposant sur l'analyse multirésolution de l'information visuelle pour estimer la fluidité du trafic. Après avoir établi les fondements théoriques de la méthode proposée, nous validons celle-ci sur un ensemble de séquences vidéo présentant des scènes de trafic routier.

# CHAPITRE 1

## Mesure de la vitesse du trafic routier

L'intérêt pour la mesure automatisée de la vitesse et de la densité du trafic routier remonte aux années 1920, alors que le contrôle des véhicules aux intersections vivait une révolution [31]. Des signaux contrôlés manuellement, on passait peu à peu à une gestion automatique et chronométrée des droits de passage. Rapidement, le besoin d'avoir un système adaptatif qui serait non seulement basé sur le temps écoulé, mais aussi sur la présence de véhicules, s'est fait sentir.

Charles Adler, Jr., un ingénieur américain, a été un des premiers chercheurs à s'intéresser à cette problématique en proposant des systèmes adaptatifs basés sur le son [10]. Dans la première version de son système, utilisée dans la ville de Baltimore à partir de 1928, le conducteur d'un véhicule devait s'arrêter à une intersection, puis demander le droit de passage en appuyant sur son klaxon. Des microphones installés à chacune des entrées d'une intersection captaient le son du klaxon et le droit de passage était accordé selon l'ordre d'arrivée. Une évolution apparue en 1931 proposait plutôt d'utiliser le bruit associé à l'arrivée des véhicules à l'intersection, éliminant ainsi la nécessité d'utiliser le klaxon [31].

Environ au même moment, un autre américain, Henry A. Haugh, développait ce qui servirait de base pour la grande majorité des systèmes de mesure du trafic routier pour les trente années suivantes. Son idée était de concevoir un capteur composé de deux plaques de métal insérées dans le pavé qui, lorsqu'en contact, permettraient à un courant électrique de circuler [31]. C'est la pression associée au passage des véhicules qui amènerait les deux plaques à se toucher. Un manque de fiabilité mécanique du système, associé à un coût élevé d'installation et une incapacité à produire des résultats fiables lorsque les conditions climatiques ne sont pas optimales (pluie, neige, etc.) a toutefois rendu impossible la démocratisation de ce produit.

Aujourd'hui, l'estimation de la vitesse du trafic routier peut se faire de plusieurs façons en se basant sur une grande variété d'indices de mesure. Les plus répandus sont :

- le son (capteurs acoustiques) ;
- l'opacité (capteurs optiques et infrarouges) ;
- les mesures de temps de vol (par micro-ondes ou lasers) ;
- l'induction électromagnétique (boucles à induction) ;
- la vibration (capteurs sismiques).

Dans le cadre de ce travail, nous regroupons ces approches en deux grandes familles : les méthodes sans information visuelle et les méthodes avec information visuelle. Les méthodes sans information visuelle sont principalement décrites dans des rapports commandés par des organismes publics de gestion du réseau routier (Gordon et Tighe [19], Klein *et al.* [31], Xie *et al.* [54]). Kastrinaki *et al.* [30] répertorient quant à eux un éventail d'applications de l'information visuelle à l'analyse du trafic routier.

Dans ce chapitre, on s'intéresse dans un premier temps, aux techniques permettant d'évaluer le déplacement des véhicules sans faire appel à l'information visuelle. Il sera question de deux des systèmes les plus répandus sur le réseau routier : les boucles à

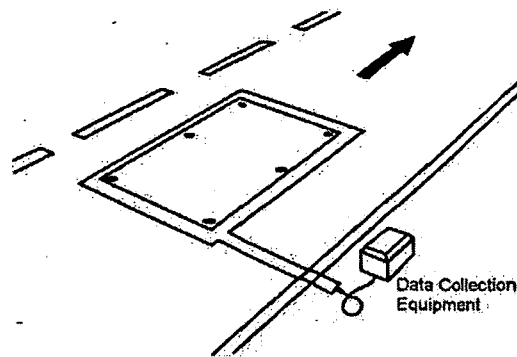


Figure 1.1 – Système de boucle à induction typique, tiré de [31].

induction et les mesures de temps de vol. Par la suite, nous présentons, à travers une série de techniques, comment l'information visuelle peut être utilisée pour estimer la vitesse du trafic.

## 1.1 Mesure de la vitesse sans information visuelle

### 1.1.1 Boucles à induction

Au cours des dernières décennies, les boucles à induction magnétique placées sous les routes sont devenues les détecteurs les plus répandus pour l'acquisition de données sur le trafic routier. Cela est dû au fait qu'elles constituent les capteurs les plus fiables, tant à cause de leur grande précision pour évaluer le nombre de véhicules que pour leur invariance aux conditions météorologiques et d'éclairage [19].

Un système de boucle à induction simple typique est composé de deux éléments : un contrôleur et une boucle à induction magnétique placée sous le pavé. Les deux éléments sont reliés par un câble électrique (figure 1.1). Lorsqu'un véhicule passe sur la boucle à induction, le flot électrique parcourant celle-ci est perturbé et une impulsion est envoyée

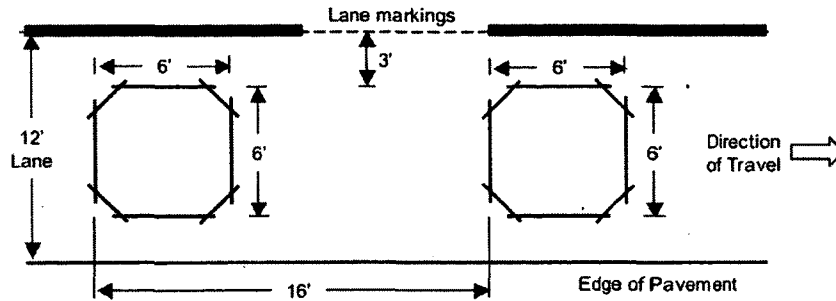


Figure 1.2 – Mesure de la vitesse d’un véhicule avec deux boucles à induction, tiré de [31].

au contrôleur afin de lui indiquer le passage d’un véhicule. Un tel système fournit des données sur le taux d’occupation des routes, puisqu’il se limite à évaluer un nombre de véhicules par unité de temps. Les informations sur la vitesse ne sont donc pas disponibles directement [54].

Plusieurs estimateurs de la vitesse du flot de trafic à partir des données fournies par les boucles à induction existent, le premier remontant à 1965 [54]. Celui-ci fait intervenir le délai entre deux mesures, le nombre de véhicules durant cet intervalle et la longueur moyenne des véhicules. Encore aujourd’hui, les techniques d’estimation de la vitesse à partir de ces données sont raffinées [26, 35], puisqu’il s’agit toujours des détecteurs les plus répandus sur le réseau routier.

Une autre méthode consiste à installer deux boucles à induction à une distance de quelques mètres l’une de l’autre (figure 1.2). Une telle installation porte le nom de *speed trap*. Lorsque la première boucle détecte le passage d’un véhicule, le contrôleur note sa signature (la signature d’un véhicule dépend essentiellement de sa taille et de son poids [40]) et chronomètre le temps nécessaire pour qu’il soit détecté par la deuxième boucle. La vitesse de chaque véhicule est donc obtenue directement.

Les principales faiblesses des boucles à induction concernent leur installation et leur

entretien. En plus de devoir fermer des voies, la mise en place de ces systèmes ne peut se faire sans couper la surface du pavé, ce qui dégrade la qualité de celui-ci, en plus d'engendrer des coûts importants [19]. De plus, les boucles à induction ne donnent pas des résultats concluants dans une situation de trafic lent ou arrêté [30].

### 1.1.2 Mesures de temps de vol

Comme la mise en place de boucles à induction est coûteuse et complexe, les autorités préfèrent parfois se tourner vers des solutions non intrusives. Une approche qui a fait ses preuves comme outil pour mesurer la vitesse est la mesure de temps de vol [22]. On utilise principalement deux types de rayonnements : micro-ondes et laser.

#### Micro-ondes

Deux types de radars à micro-ondes sont utilisés dans les applications routières [31] : ceux émettant des ondes constantes et ceux émettant des ondes de fréquences variables. Dans les deux cas, les émetteurs doivent être positionnés de manière à émettre les ondes dans la direction opposée à la circulation (figure 1.3). Les micro-ondes sont typiquement insensibles aux conditions climatiques, il est toutefois possible que la présence d'importantes structures métalliques influence les résultats [19].

L'estimation de la vitesse à l'aide des micro-ondes à fréquence constante est basé sur l'effet Doppler. Lorsqu'un véhicule passe sous le flux d'énergie émis, une portion de cette énergie est réfléchié dans la direction inverse, revenant ainsi à la source. Cette énergie réfléchié est captée par un récepteur, et en comparant la fréquence reçue par rapport à la fréquence émise, il est possible de déduire la vitesse des véhicules [31]. Ainsi, contrairement aux boucles à induction, la vitesse du trafic est ici un paramètre obtenu directement. En contrepartie, ce type de capteur sera très peu sensible aux mouvements



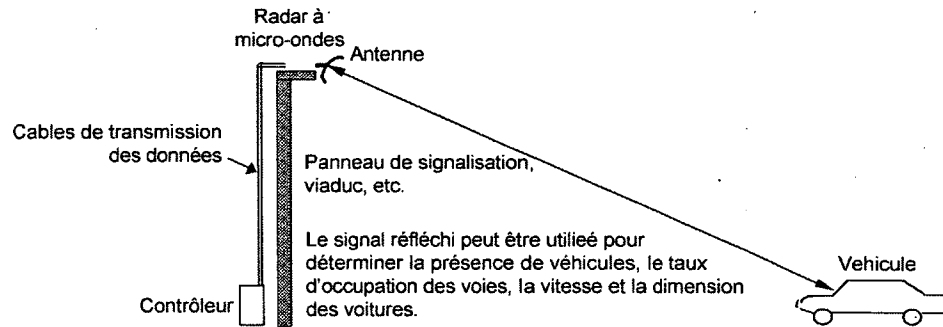


Figure 1.3 – Principe de fonctionnement d’un radar à micro-ondes, tiré de [31].

lents des véhicules et sera tout simplement incapable de détecter les véhicules arrêtés [19].

Le second type de micro-ondes utilisé dans l’analyse routière transmet une onde d’amplitude constante, mais dont la fréquence varie avec le temps. Tout au long de son parcours, l’onde conservera les mêmes propriétés fréquentielles. C’est ce qui permettra au système de distinguer les ondes reçues et, en connaissant le temps écoulé depuis leur émission, d’évaluer la vitesse des véhicules de même que de permettre la détection des véhicules arrêtés [21].

### Laser

Le fonctionnement de la mesure de vitesse par laser reprend en quelque sorte le principe des *speed traps* mentionné à la section 1.1.1. Un rayon laser couvrant la largeur de la route (ou d’une voie) est séparé en deux traverses, tel qu’illustré à la figure 1.4. Un émetteur-récepteur envoie de l’énergie infrarouge cohérente en direction de la chaussée puis évalue le temps nécessaire pour que le signal revienne par réflexion sur les véhicules. Il enregistre les temps de passage des véhicules aux deux traverses, ce qui permet d’évaluer leur vitesse [31].

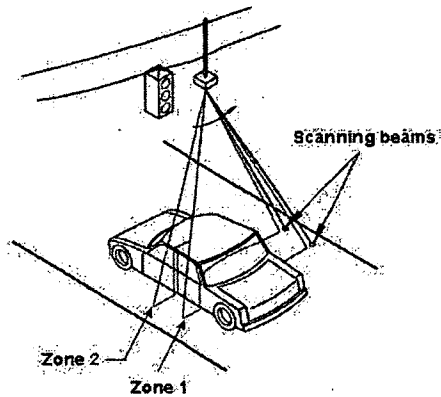


Figure 1.4 – Mesure de vitesse avec un laser, tiré de [31].

Idéalement, les laser doivent être positionnés de 6 à 8 mètres au-dessus de la chaussée. De plus, en raison de l'utilisation d'un capteur optique situé en hauteur, cette technique de mesure est sensible aux conditions météorologiques nuisant à la visibilité [39]. L'utilisation des lasers rend aussi possible l'obtention d'un rendu à trois dimensions du flot de véhicules, permettant ainsi de classifier ceux-ci selon leur format [31].

## 1.2 Mesure de la vitesse avec information visuelle

Les méthodes abordées à la section précédente présentent plusieurs problèmes d'importance : les techniques non intrusives (mesure de temps de vol, analyse sonore) souffrent d'un manque de robustesse dû à une trop grande sensibilité aux facteurs externes (intempéries, trafic dense), tandis que les techniques plus complexes (boucles à induction) offrent la robustesse demandée, mais leur installation et leur entretien est coûteux et complexe [19]. Dans cette optique, l'utilisation de données visuelles est pertinente, à condition de savoir bien gérer les variations dans l'illumination et les perturbations dues aux intempéries. En effet, des réseaux de surveillance par caméras sont bien souvent déjà

en place sur les routes. De plus, l'information contenue dans ces vidéos de surveillance devrait suffire à estimer le paramètre qui nous intéresse, soit la mesure de la vitesse du trafic. Finalement, comme les données visuelles sont très riches en contenu, il sera possible d'adapter les méthodes de traitement des données à mesure que celles-ci évolueront sans changer les infrastructures d'acquisition.

À la base, les caméras vidéo ont été introduites dans les systèmes de gestion de trafic routier afin de transmettre des images de la route à une centrale où une personne pourrait les analyser en direct ou *a posteriori*, en cas d'accident, par exemple [31]. Avec la mise au point de techniques avancées pour l'analyse des images et la montée en puissance du matériel informatique, les systèmes actuels deviennent capables d'analyser automatiquement et en temps réel les images captées afin d'estimer diverses informations sur le flot de voitures. Qui plus est, un inconvénient majeur de la surveillance humaine est qu'il est pratiquement impossible, ou alors très coûteux, de maintenir une attention constante sur l'ensemble des sources d'informations visuelles dans un réseau routier. Il importe donc d'utiliser ces technologies d'automatisation de l'analyse afin d'exploiter au maximum le contenu de l'information visuelle.

Un tel système devrait d'abord pouvoir détecter les véhicules en se basant sur l'intensité des pixels, pour ensuite en évaluer leur mouvement sur la route. Les deux étapes majeures d'un tel processus sont donc la détection des objets en mouvement (véhicules, motocyclettes, camions, etc.), puis la quantification de leur déplacement afin d'en déterminer la vitesse. Le déplacement des véhicules est d'abord évalué en espace image, c'est-à-dire qu'il est exprimé en pixels par unité de temps. Pour arriver à une telle mesure, il est nécessaire de connaître la cadence de prise de vue lors de l'acquisition des séquences vidéo. On sait ainsi combien de temps s'écoule entre deux images consécutives de la séquence, et donc entre les positions d'un même objet dans ces deux images. La conversion en une mesure de vitesse en distance par unité de temps se fait ensuite en

évaluant une correspondance entre les distances en espace image et les distances réelles dans l'espace monde. Cette relation est déterminée par une calibration de la caméra.

### 1.2.1 Calibration de caméra pour l'estimation de la vitesse réelle

L'estimation du mouvement des véhicules dans une séquence vidéo permet facilement d'obtenir une vitesse dont l'unité de déplacement est le pixel. Cette vitesse en espace image  $v_{\text{pix}}$  exprimée en pixels par seconde est obtenue en évaluant l'amplitude du vecteur de déplacement en pixels  $\mathbf{d}_v$  d'un véhicule entre deux images consécutives de la séquence. Cette donnée est ensuite multipliée par la cadence d'acquisition  $C_{\text{acq}}$  exprimée en IPS (images par seconde) :

$$v_{\text{pix}} = |\mathbf{d}_v| C_{\text{acq}}.$$

Pour fournir une estimation de la vitesse réelle du trafic, on doit convertir le mouvement en coordonnées pixel du repère image en une distance réelle dans le repère monde. Cette conversion implique une étape de calibration de la caméra. Pour y arriver, Dailey *et al.* [12] suggèrent de définir, en plus du plan image  $uv$ , deux repères 3D, tels qu'illustrés à la figure 1.5. Le premier système de coordonnées, noté  $x_s y_s z_s$  est centré à la position de la caméra dans l'espace et l'axe  $x_s$  donne la direction d'observation de celle-ci, tandis que le plan  $y_s z_s$  est parallèle au plan image. Le second repère,  $x_e y_e z_e$ , a son origine à la surface du sol. Le plan  $x_e y_e$  définit la surface supposée plane où se déplacent les véhicules. On suppose que l'axe  $x_s$  forme un angle  $\phi$  avec le plan  $x_e y_e$ .

Le passage du plan image au repère caméra est défini par une simple relation de projection perspective, tandis que le changement de repère entre les coordonnées caméra et monde se fait à l'aide d'une matrice de rotation et un vecteur de translation. Afin d'estimer les paramètres de ces transformations pour une caméra de surveillance sur le réseau routier, plusieurs auteurs recommandent l'utilisation des points de fuite comme

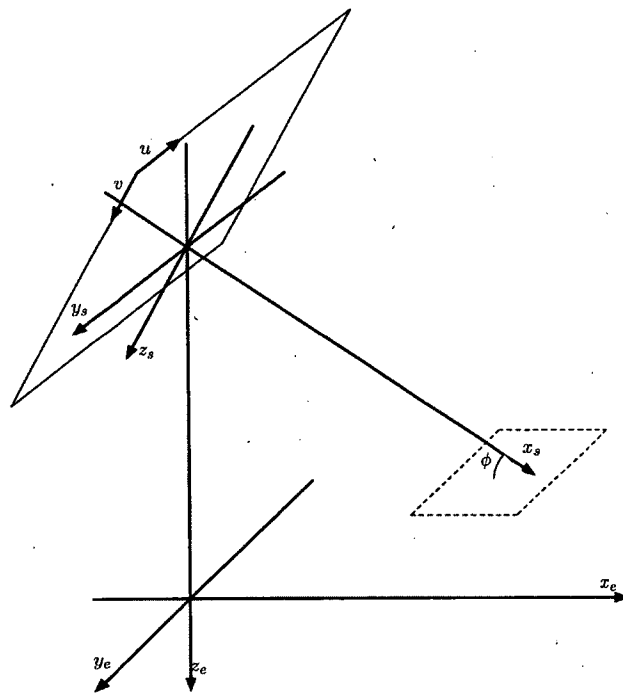


Figure 1.5 – Position relative des repères caméra et monde, inspiré de [12].

source d'information géométrique [44].

Les approches les plus communes permettant d'estimer la vitesse des véhicules à partir de l'information visuelle se séparent en deux classes : celles proposant une analyse globale du flot de véhicule et celles reposant sur le suivi des objets en mouvement. La première catégorie permet d'estimer certains paramètres, tels que la vitesse moyenne et la densité, sur le flot de voitures pris dans son ensemble, tandis que la seconde tire son information sur la vitesse du trafic à partir des données associées à chacun des véhicules, indépendamment.

### **1.2.2 Mesure de la vitesse par l'analyse globale du trafic routier**

Une méthode reposant sur l'élaboration d'un profil d'intensité, qui correspond à l'intensité maximale moyenne le long de la direction du trafic (figure 1.6) a été proposée par Cho et Rice [9]. En partant du principe que la route et les véhicules ont des intensités différentes, on a que chaque pic du profil correspond à un véhicule. En comparant ceux-ci d'une image à l'autre, il est possible d'estimer le mouvement de l'ensemble des véhicules. De plus, comme il n'y a pas d'identification des véhicules pour le suivi, le traitement des données est grandement simplifié. Évidemment, une telle méthode suppose que l'écart entre les véhicules demeure constant. Cette limitation n'est toutefois pas trop contraignante. L'analyse se fera sur des images consécutives (ou assez rapprochées) dans la séquence, de manière à ne pas laisser le temps à l'écart entre les véhicules de varier significativement. Finalement, comme il s'agit d'une analyse unidimensionnelle, les temps de calcul sont assez réduits. Comme cette méthode repose sur un découpage manuel des voies de circulation, elle n'est fonctionnelle que sur des routes qui sont droites. Pour la même raison, les résultats sont aussi sensibles aux vibrations de la caméra dues, par exemple, au vent. Dans ce cas, d'une image à l'autre dans la séquence, les voies ne

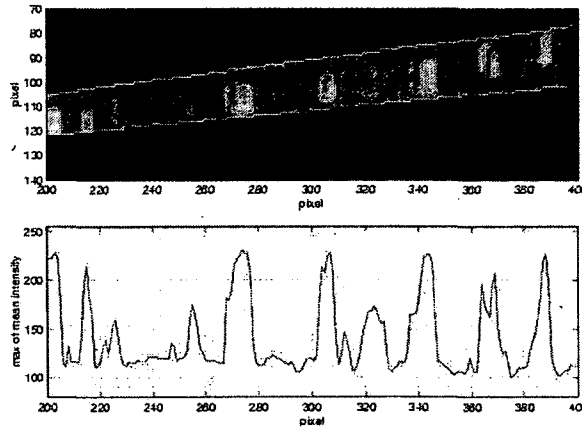


Figure 1.6 – Profil d'intensité d'une voie, tiré de [9].

correspondent pas exactement aux mêmes pixels, ce qui diminue la qualité des résultats.

Yu *et al.* [55] proposent quant à eux de faire appel aux vecteur de mouvement qui sont à la base de la compression vidéo MPEG en les utilisant comme une approximation du flux optique. Un avantage non négligeable de cette méthode est de ne nécessiter aucun traitement supplémentaire (dans la mesure où l'on dispose de vidéos en format MPEG) pour estimer le mouvement. Les auteurs ont toutefois souligné l'instabilité de leur méthode lors de l'évaluation de la vitesse moyenne sur de trop courts laps de temps, en raison d'une trop grande sensibilité au bruit. Bien entendu, si les vidéos ne sont pas, dès le départ, au format MPEG, cette méthode perd de son intérêt en raison du traitement nécessaire pour effectuer une nouvelle compression.

Le problème de l'évaluation du mouvement des véhicules peut aussi être abordé du point de vue fréquentiel. Mohamadi *et al.* [38] utilisent la transformée de Radon pour estimer les paramètres du flou dû au mouvement (« *motion blur* »). Celui-ci est perçu lorsque l'obturateur d'une caméra photo reste ouvert pendant un certain temps et qu'un mouvement relatif entre la caméra et les objets survient. Partant du principe que plus la vitesse d'un véhicule est grande, plus le flou dû au mouvement sera important, la

quantification de ce dernier permet d'estimer la vitesse des véhicules.

### **1.2.3 Mesure de la vitesse par suivi de véhicules**

Il a été mentionné précédemment que c'est entre autres la richesse et la polyvalence des données visuelles qui nous motivaient à les utiliser pour évaluer certains paramètres du trafic routier. Ce faisant, il est logique de souhaiter aller plus loin que la mesure de la vitesse en cherchant à obtenir davantage d'informations sur le trafic. À titre d'exemple, on peut vouloir classer les véhicules selon leur type, évaluer des statistiques de changement de voies, connaître la longueur des embouteillages, etc [16]. Ces nouvelles estimations plus complexes font appel à des techniques de suivi d'objets en mouvement.

Un système basé sur le suivi des véhicules doit être en mesure d'identifier chaque voiture apparaissant dans le champ de vision de la caméra, puis de suivre son déplacement avec précision dans la séquence vidéo et ce, jusqu'à ce qu'il quitte la portion de route visible. À partir des informations sur le mouvement de chaque véhicule, on peut évaluer des paramètres globaux sur le trafic (vitesse moyenne, occupation de la route), de même que l'information sur chaque voiture (fréquence de changement de voie, vitesse maximale, etc.).

Les techniques les plus populaires pour effectuer le suivi d'objets en mouvement peuvent être divisées en trois familles : le suivi basé sur les régions semblables, le suivi basé sur les contours d'objets et le suivi basé sur les caractéristiques.

#### **Suivi basé sur les régions semblables**

Avec le suivi basé sur les régions semblables, l'idée est d'associer une région fermée de l'image à chaque véhicule, puis d'en faire le suivi dans le temps en utilisant une mesure de corrélation. L'obtention des régions associées aux objets en mouvement est



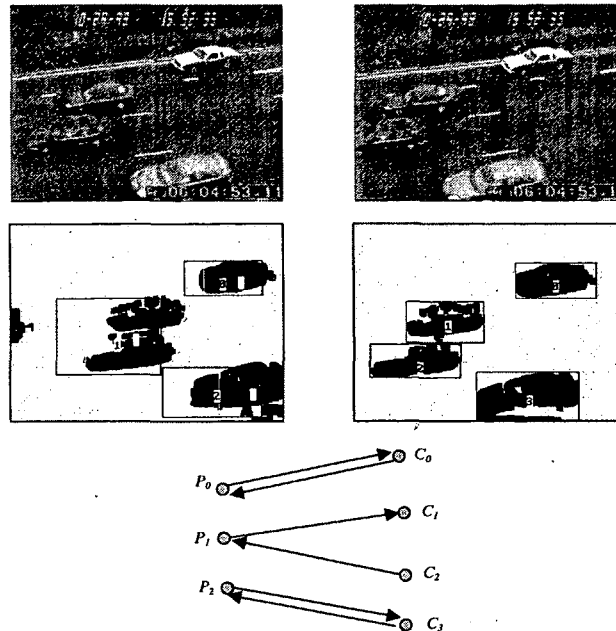


Figure 1.7 – Graphe associé aux véhicules en mouvement dans deux images, tiré de [23].

souvent faite à l'aide d'une technique de soustraction de fond [2]. Ce faisant, comme la qualité du suivi des véhicules dépend de la capacité à obtenir des régions distinctes pour chaque objet, une telle approche risque de moins bien fonctionner dans une situation de trafic dense où les véhicules seront très près les uns des autres et où il y aura de l'occultation [37]. Pour ajouter de la robustesse, Gupte *et al.* [23] suggèrent de modéliser le problème du suivi de véhicule comme un problème d'optimisation de graphe biparti où les sommets correspondent aux régions associées aux véhicules en mouvement dans deux images consécutives de la séquence vidéo, tel qu'illustré à la figure 1.7.

### Suivi basé sur les contours d'objets

Une autre approche pour le suivi de véhicules en mouvement consiste à représenter le contour de l'objet et de le mettre à jour à chaque nouvelle image de la séquence vidéo.

L'intérêt principal d'un suivi basé sur les contours est de réduire la complexité des calculs à effectuer [37]. Les contours peuvent être obtenus à l'aide du gradient temporel [12], des contours actifs [32] ou d'opérateurs morphologiques [15]. Dans une situation de trafic dense, il demeure toutefois possible que plusieurs véhicules soient regroupés en une seule région et ainsi qu'ils soient représentés par un unique contour, ce qui ne permet pas la distinction individuelle des objets [37].

### **Suivi basé sur les caractéristiques**

Finalement, la troisième technique de suivi de véhicules est celle reposant sur le suivi de caractéristiques. Cette façon de faire abandonne l'idée de suivre le déplacement d'objets entiers et se concentre plutôt à faire la détection et le suivi de points caractéristiques appartenant aux objets en mouvement. L'avantage d'une telle approche est de permettre le suivi des véhicules même en situation de trafic dense [37]. En effet, comme un véhicule sera associé à plusieurs points caractéristiques, ils n'est pas nécessaire que ceux-ci soient toujours tous visibles pour le détecter et en évaluer le déplacement. Finalement, pour déterminer si un ensemble de points caractéristiques appartient au même objet, Beymer *et al.* [37] proposent d'utiliser une contrainte d'uniformité du mouvement.

En raison de leur grande fiabilité, les boucles à induction magnétique ont acquis une position dominante sur l'ensemble des autres techniques. Toutefois, cette technique n'est pas parfaite : il s'agit d'abord et avant tout d'un outil pour évaluer la densité du trafic et non des mesures plus complexes (vitesse, types de véhicules, comportement routier, etc.). Des moyens pour contourner ces limitations ont certes été trouvés (utilisation de double-boucles pour le calcul de la vitesse, par exemple), mais la difficulté de modifier les installations existantes afin de pouvoir y arriver rend parfois l'utilisation des boucles à induction plus difficile.

C'est pour compenser ces inconvénients que les recherches de méthodes alternatives

n'ont jamais cessées, et avec le développement de l'analyse d'images sur ordinateur, les méthodes impliquant un traitement automatisé de séquences vidéo se sont développées. L'obtention d'informations sur le flot de véhicules repose avant tout sur notre capacité à détecter les véhicules en mouvement dans une scène. C'est donc là que se situe le principal défi, et c'est aussi dans cette direction que s'effectue la majorité de la recherche. Une avenue prometteuse consiste à combiner la soustraction de fond et le calcul du flux optique tel que suggéré par Ji *et al.* [28], et c'est sur cette idée que se base la méthode que nous proposons et qui sera décrite au chapitre suivant.

## CHAPITRE 2

# Vitesse du trafic routier par information visuelle multirésolution

La surveillance et l'analyse de la circulation automobile sur le réseau routier sont devenues des enjeux primordiaux auprès des autorités concernées [19, 31]. La hausse des performances du matériel informatique permet la mise au point de méthodes automatisées pour accomplir ces tâches en direct. L'avenue que nous proposons est d'utiliser les données captées par des réseaux de caméras de surveillance déjà en place sur les routes. L'information visuelle est donc déjà disponible, mais est présentement peu utilisée pour l'analyse automatique du trafic.

L'objectif est ici d'élaborer une technique automatisée et adaptative basée sur l'information visuelle multirésolution permettant d'évaluer la fluidité de la circulation. Ce travail s'inscrit dans le contexte plus large de la mise au point d'une méthodologie d'évaluation de la vitesse du trafic routier visant à détecter, identifier et anticiper les problèmes de circulation. Partant du principe qu'il n'est pas nécessaire d'évaluer, avec autant de précision l'information sur le déplacement des véhicules si la circulation est fluide, l'utili-

sation d'une approche multirésolution permet d'adapter le niveau de détail des données visuelles à analyser en fonction de la vitesse du trafic. Ce faisant, le temps d'exécution s'en trouve diminué, puisqu'en situation de trafic rapide, les images utilisées seront de résolution inférieure à l'image originale.

Afin de mettre l'accent sur le traitement des données visuelles pour la quantification de la fluidité du trafic, nous formulons certaines hypothèses de travail visant à exclure certaines problématiques associées à la variabilité des conditions d'acquisition. Premièrement, nous prenons pour acquis que la luminosité dans la scène demeure constante tout au long de l'acquisition. Une approche prétendant être insensible aux conditions d'éclairage a été testée (section 2.1.2), mais les résultats obtenus n'étaient pas satisfaisants. Ceux-ci sont présentés à la section 3.1.2. Dans le même ordre d'idées, nous prenons aussi pour acquis que les conditions météorologiques sont stables pour la durée de la séquence analysée. Des variations climatiques pourraient non seulement avoir un impact sur la luminosité dans la scène, mais pourrait aussi nuire à la détection des véhicules, en cas de précipitation.

Il est aussi nécessaire de poser des hypothèses sur l'acquisition des données vidéo en tant que tel. Comme nous le verrons dans la section 2.1.1, la méthode proposée repose en grande partie sur notre capacité à discerner les véhicules en mouvement de l'arrière-plan. En raison de l'approche utilisée pour arriver à cette fin, il est nécessaire de supposer que la caméra servant à l'acquisition soit statique, de manière à avoir un arrière-plan qui occupe les mêmes régions dans l'image tout au long de la séquence. De plus, comme ce travail porte avant tout sur l'analyse des données de déplacement des véhicules, nous laissons de côté les problèmes causés par la déformation géométrique due à la projection en perspective. Tel qu'indiqué à la section 1.2.1, une étape de calibration est nécessaire pour compenser ces déformations. Comme nous omettons cette étape dans le présent travail, l'unité de mesure de déplacement que nous obtenons sera limitée au pixel. La

fréquence d'acquisition de la caméra ne sera pas non plus considérée. Cette dernière entre aussi dans le calcul permettant de passer d'une mesure de déplacement en pixels à une en distance réelle. Le processus de traitement des données présenté dans ce travail est donc valide, mais ne permet pas pour l'instant d'obtenir une mesure de vitesse réelle des véhicules.

Les deux étapes principales de la méthode sont la détection des véhicules et l'estimation du mouvement. En effet, l'évaluation du mouvement étant une opération coûteuse en temps de calcul, on cherche à limiter les régions dans l'image où ces calculs doivent être faits en obtenant dans un premier temps la position des véhicules en mouvement. Par la suite, le déplacement des véhicules est quantifié pour permettre l'estimation de la vitesse.

1. **Détection des véhicules** : la première étape consiste à détecter les régions dans la scène où des déplacements sont observés (les véhicules en mouvement), et c'est uniquement sur ces zones que la vitesse sera estimée. En effet, si aucun déplacement n'est observé, il n'est pas nécessaire de le quantifier. Selon la nature de la scène, il est aussi possible que des objets autres que des véhicules soient détectés (des piétons, par exemple). Il serait important de ne pas considérer ceux-ci dans le calcul de la vitesse du trafic.
2. **Estimation du mouvement** : la quantification du mouvement se fait en évaluant le flux optique entre deux images consécutives de la séquence vidéo. Cette approche de la quantification du mouvement a été choisie en raison de sa polyvalence. Il a l'avantage de permettre à la fois une évaluation de la vitesse instantanée ainsi que la vitesse moyenne. Il rend aussi possible l'estimation de la vitesse moyenne dans la scène autant que celle de la vitesse individuelle de chaque véhicule. Le flux optique illustre le mouvement apparent des objets dans une scène. Comme il n'y a aucun mouvement en dehors des régions d'intérêt obtenues à l'étape précédente, on évalue

le flux optique seulement sur celles-ci.

En tout temps, le déplacement moyen dans la scène sera évalué à partir de données de flux optique afin d'obtenir un indice de fluidité global. On applique d'abord un seuil sur l'amplitude des vecteurs de déplacement, puis on utilise ceux dépassant celui-ci pour estimer le déplacement moyen dans la scène. Tant que la circulation demeure fluide, il n'est pas nécessaire d'estimer le déplacement du trafic avec plus de précision. Toutefois, si cet indice descend sous un certain seuil, l'analyse sera raffinée pour fournir le déplacement moyen par direction, en utilisant une résolution deux fois supérieure à la précédente. Finalement, à partir des images à pleine résolution, il sera possible d'obtenir une mesure du déplacement de chaque véhicule, sans toutefois en faire le suivi, toujours dans le but de localiser avec plus de précision les ralentissements dans le flot de véhicules.

## 2.1 Détection des véhicules par segmentation de l'arrière-plan

L'obtention d'informations par l'analyse du contenu d'une séquence vidéo est souvent une opération coûteuse en temps. En effet, la quantité de données à traiter étant très importante, il est ardu de maintenir de bas temps d'exécution. Le principe de la sous-traction de fond est justement de s'attarder à réduire cette grande quantité de données à traiter en ciblant des zones d'intérêt et en éliminant les régions non essentielles du traitement.

L'objectif général de ce travail est de rendre possible l'utilisation du flux optique dans le calcul automatisé de la vitesse du trafic routier. Comme le flux optique ne doit être calculé que sur les objets en mouvement (les véhicules, dans le cas présent), il serait utile d'effectuer au préalable une détection de ces objets et d'éliminer l'arrière-plan des

calculs reliés au flux optique. Pour arriver à cette fin, deux méthodes seront considérées. La première utilise une approche statistique en modélisant temporellement chaque pixel de la séquence par un mélange de gaussiennes. La seconde fait plutôt appel à une analyse de la texture par histogrammes LBP (*local binary patterns*) afin de séparer les objets en mouvement de l'arrière-plan statique.

### 2.1.1 Détection des véhicules par modélisation statistique

Une des premières publications à mentionner la modélisation statistique comme approche pour la soustraction de fond est celle de Friedman et Russell [17]. Les auteurs y proposent de modéliser temporellement chaque pixel (figure 2.1) par un mélange de trois distributions gaussiennes. Soit  $\{X_1, X_2, \dots, X_{t-1}\}$ , l'historique récent des intensités d'un pixel. La probabilité d'obtenir  $X_t$  comme prochaine mesure d'intensité est alors donnée par

$$P(X_t) = \sum_{i=1}^3 \omega_{i,t} \eta(X_t, \mu_{i,t}, \sigma_{i,t}^2),$$

où  $\eta(X_t, \mu_{i,t}, \sigma_{i,t}^2)$  est la fonction de densité de probabilité d'une distribution gaussienne, c'est-à-dire que

$$\eta(X_t, \mu_{i,t}, \sigma_{i,t}^2) = \frac{1}{\sqrt{2\pi\sigma_{i,t}^2}} \exp\left(-\frac{X_t - \mu_{i,t}}{2\sigma_{i,t}^2}\right).$$

Les paramètres (la proportion  $\omega_{i,t}$ , la moyenne  $\mu_{i,t}$  et la variance  $\sigma_{i,t}^2$ ) de celles-ci sont estimés de manière incrémentale afin qu'il ne soit pas nécessaire de fournir la séquence entière à l'algorithme. Chaque distribution gaussienne sera associée à une des trois classes de pixels : objet en mouvement, arrière-plan ou ombre. Pour chaque nouveau pixel, la probabilité d'appartenance à chacune des trois distributions gaussiennes  $P(i|X_t)$  sera évaluée et le pixel sera assigné à la classe présentant la plus grande probabilité. Suite à cette assignation, les paramètres de la distribution choisie seront mis à jour pour prendre en considération l'intensité du nouveau pixel.



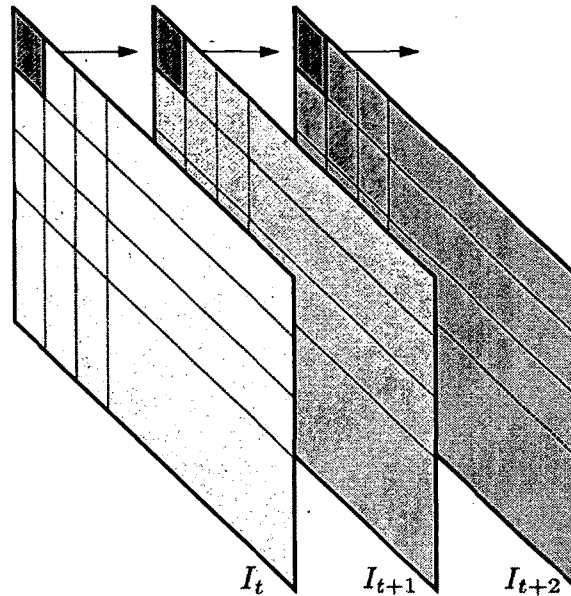


Figure 2.1 – Analyse temporelle de chaque pixel d’une séquence.

Les travaux de Stauffer et Grimson [47, 48] reprennent en grande partie les idées énoncées par Friedman et Russell. Ils considèrent toutefois qu’il est risqué de s’en tenir au même modèle de mélange gaussien tout au long du processus. En effet, considérons le cas d’un changement périodique entre deux états d’illumination dans une scène. Celui-ci aura un impact sur les intensités perçues pour l’arrière-plan et les objets. Un mélange de  $K$  distributions gaussiennes dont les paramètres sont fixes n’est alors plus suffisant : l’arrière-plan doit être représenté par deux distributions gaussiennes correspondant à chacun des états d’illumination. Les modifications proposées par Stauffer et Grimson tentent d’améliorer le fonctionnement de l’algorithme dans l’éventualité où, au fil du temps, un pixel ne correspond plus au modèle à trois classes fixes. Le principe de cette nouvelle approche est de permettre des modifications plus importantes aux distributions composant le mélange.

Soit  $(x_0, y_0)$ , un pixel de la séquence  $I$ . Au temps  $t$ , il est possible d’obtenir l’historique

des intensités de ce pixel :

$$\{X_1, \dots, X_t\} = \{I(x_0, y_0, i) \mid 1 \leq i \leq t\}.$$

Cet historique est modélisé par un mélange de  $K$  distributions gaussiennes. Une nouvelle valeur  $X_{t+1}$  pourra généralement être associée aux composantes actuelles du mélange. Dans un tel cas, les paramètres des  $K$  distributions gaussiennes seront mis à jour afin de considérer cette nouvelle donnée de la façon suivante :

$$\begin{aligned}\omega_{k,t+1} &= (1 - \alpha)\omega_{k,t} + \alpha M_{k,t+1} \\ \mu_{k,t+1} &= (1 - \alpha)\mu_{k,t} + \alpha X_{t+1} \\ \sigma_{k,t+1}^2 &= (1 - \alpha)\sigma_{k,t}^2 + \alpha (X_{t+1} - \mu_{k,t})^2\end{aligned}$$

où  $M_{k,t}$  est la probabilité *a posteriori* de la  $k^e$  Gaussienne, c'est-à-dire que

$$M_{k,t} = \begin{cases} 1 & \text{pour la distribution associée au pixel} \\ 0 & \text{pour les autres distributions} \end{cases},$$

et où  $\alpha$  est le coefficient d'apprentissage.

C'est à cette étape que se situe la principale nouveauté apportée par Stauffer et Grimson. Plutôt que de forcer une assignation à une classe existante, même si l'intensité du pixel ne correspond à aucune des  $K$  composantes du mélange actuel, la distribution la moins importante (celle ayant la plus faible proportion  $\omega$ ) sera remplacée par une nouvelle distribution ayant comme moyenne l'intensité  $X_{t+1}$  et une grande variance. Le mélange de gaussiennes spécifié au départ n'est donc plus statique, puisque les distributions qui le composent peuvent non seulement être modifiées, mais aussi tout simplement remplacées.

### 2.1.2 Détection des véhicules par analyse de la texture

La vaste majorité des techniques de soustraction de fond ne basent leur classification que sur l'intensité des pixels. Souvent en raison de leur plus grande complexité, les ap-

proches prônant une discrimination basée sur des caractéristiques de texture sont laissées de côté. Satpathy *et al.* [43] affirment que l'utilisation des textures comme moyen de discriminer les objets en mouvement de l'arrière-plan statique rend la méthode beaucoup moins sensible aux changements d'illumination dans la scène. Comme nous souhaitons que notre méthode soit le moins sensible possible aux conditions d'acquisition, le fait de ne pas dépendre du niveau d'illumination est attrayant. Ainsi, contrairement à la plupart des méthodes de soustraction de fond, l'information permettant de modéliser l'arrière-plan provient d'une région dans l'image et non pas seulement d'un pixel. Heikkilä et Pietikäinen [24, 25], puis Satpathy *et al.* [43] proposent toutefois d'utiliser les motifs binaires locaux (*local binary patterns*, ou LBP) comme moyen de quantifier les textures.

La valeur du LBP associé à un pixel  $(x, y)$  de l'image  $I_t$  est donnée par

$$\text{LBP}_{P,R}(x, y) = \sum_{p=0}^{P-1} 2^p \cdot s(g_p - I_t(x, y)),$$

où

$$s(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

et  $g_p$  correspond à l'intensité de chacun des  $P$  pixels répartis uniformément sur un cercle de rayon  $R$  autour de  $(x, y)$  (Figure 2.2). Une valeur LBP sera donc évaluée pour tous les pixels de l'image. Par la suite, pour chaque pixel  $(x, y)$  de l'image  $I_t$ , les LBP associés aux sites formant le voisinage de  $(x, y)$  seront additionnés afin d'obtenir un histogramme-LBP pour chaque pixel de l'image. Ce sont ces histogrammes qui serviront à quantifier les textures.

L'historique  $\{X_1, \dots, X_t\}$  d'un pixel sera modélisé par un ensemble de  $K$  histogrammes-LBP adaptatifs  $m_k$ . À chacun de ces  $K$  histogrammes est associé un poids  $w_k \in [0, 1]$  tel que

$$\sum_k w_k = 1.$$

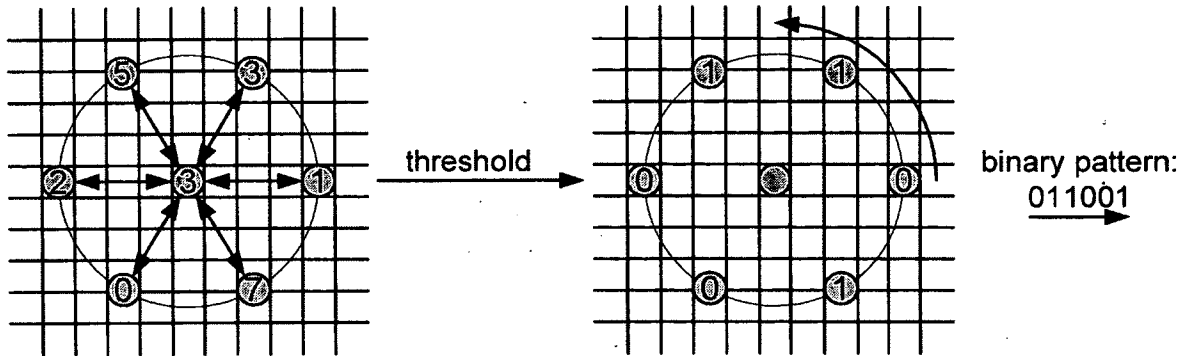


Figure 2.2 – Construction du code LBP d’un pixel, tiré de [24].

Soit  $X_{t+1}$ , l’intensité d’un nouveau pixel à considérer. Son histogramme-LBP  $h$  sera comparé (avec une mesure de similarité) aux  $K$  histogrammes modélisant l’historique du pixel. L’histogramme  $m_i$  situé le plus près de  $h$  sera alors mis à jour :

$$m_i = \alpha_b \cdot h + (1 - \alpha_b) \cdot m_i,$$

où  $\alpha_b$  est une constante fixée par l’utilisateur. L’ensemble des poids  $w_k$  du modèle seront aussi mis à jour :

$$w_k = \alpha_w \cdot M_k + (1 - \alpha_w) \cdot w_k,$$

où

$$M_k = \begin{cases} 1 & \text{si } k = i \\ 0 & \text{sinon} \end{cases}$$

et où  $\alpha_w$  est une constante fixée par l’utilisateur.

La classification avant-plan/arrière-plan se fera en testant la similarité entre  $h$  et chacun des  $K$  histogrammes du modèle. Si, pour au moins un des  $K$  histogrammes, la mesure de similarité est supérieure à un certain seuil  $T_p$  fixé par l’utilisateur, le pixel sera associé à la classe du fond. Sinon, il sera considéré comme un objet en mouvement.

### 2.1.3 Évaluation des régions d'intérêt

Pour faire la détection des véhicules en mouvement, on utilisera une variation proposée par Lee [33] de la méthode de modélisation statistique de Stauffer et Grimson [47] présentée en 2.1.1. On applique ensuite un filtre médian sur l'image binaire résultante afin d'éliminer les pixels isolés. Avec cette opération, l'objectif est d'éliminer la nécessité de faire un calcul de flux optique sur l'image entière et de plutôt se concentrer sur les régions où se trouvent des véhicules en mouvement. La méthode utilisée reprend les principes généraux introduits en 2.1.1, mais diffère dans la définition du coefficient d'apprentissage. Alors que Stauffer et Grimson le définissent comme étant un paramètre constant  $\alpha$ , Lee suggère plutôt de réévaluer celui-ci à chaque nouvelle donnée, pour chaque distribution gaussienne composant le mélange. Le coefficient d'apprentissage suggéré est

$$\eta_{k,t} = M_{k,t} \left( \frac{1 - \alpha}{c_k} + \alpha \right),$$

où  $c_k$  est le nombre d'observations associées à la  $k^e$  distribution. Ce nouveau coefficient sera utilisé pour mettre à jour la moyenne  $\mu$  et la variance  $\sigma^2$ , tandis que la constante  $\alpha$  sera utilisée directement pour mettre à jour la proportion  $\omega$  :

$$\begin{aligned} \omega_{k,t+1} &= (1 - \alpha)\omega_{k,t} + \alpha M_{k,t+1}; \\ \mu_{k,t+1} &= (1 - \eta_{k,t})\mu_{k,t} + \eta_{k,t}X_{t+1}; \\ \sigma_{k,t+1}^2 &= (1 - \eta_{k,t})\sigma_{k,t}^2 + \eta_{k,t}(X_{t+1} - \mu_{k,t})^2. \end{aligned}$$

Partant de cette modélisation statistique par un modèle de mélange gaussien, il sera possible d'effectuer une segmentation entre le fond et les objets en mouvement.

Lorsque l'algorithme reçoit une nouvelle image  $I_{t+1}$  pour la séquence  $I$ , une classification binaire (fond ou objet en mouvement) est effectuée pour chacun des pixels, en mesurant la probabilité  $P(B|X_{t+1})$ , où  $X_{t+1}$  est l'intensité du pixel au temps  $t + 1$  et  $B$

représente la classe associée au fond. On évalue la probabilité  $P(B|X_{t+1})$  comme suit :

$$P(B|X_{t+1}) = \frac{\sum_{k=1}^K P(X_{t+1}|G_k) P(G_k) P(B|G_k)}{\sum_{k=1}^K P(X_{t+1}|G_k) P(G_k)},$$

où  $G_k$  correspond à la  $k^e$  Gaussienne et

$$\begin{aligned} P(X_{t+1}|G_k) &= \frac{1}{\sigma_k \sqrt{2\pi}} e^{-\frac{(X_{t+1}-\mu_k)^2}{2\sigma_k^2}}; \\ P(G_k) &= \omega_k; \\ P(B|G_k) &= \frac{1}{1 + e^{-a\frac{\omega_k}{\sigma_k} + b}}. \end{aligned}$$

$P(B|G_k)$  est un estimateur de densité suggéré par Lee où  $a$  et  $b$  sont des constantes laissées aux valeurs suggérées par l'auteur. Stauffer et Grimson suggéraient plutôt  $P(B|G_k) = 1$  ou  $P(B|G_k) = 0$  selon la valeur du ratio  $\frac{\omega_k}{\sigma_k}$  de la distribution gaussienne. L'estimateur  $P(B|G_k)$  proposé par Lee fait donc preuve de plus de souplesse : peu importe la distribution à laquelle un pixel est associé, la probabilité d'appartenance à l'arrière-plan ne sera pas nulle. Avec l'algorithme de Stauffer et Grimson, les distributions gaussiennes étaient simplement associées à l'arrière-plan ( $P(B|G_k) = 1$ ) ou aux objets en mouvement ( $P(B|G_k) = 0$ ). Une fois  $P(B|X_{t+1})$  estimé pour  $I_{t+1}$ , l'avant-plan (les objets en mouvement) sera composé des pixels pour lesquels  $P(B|X_{t+1}) < 0.5$ .

Pour raffiner le résultat de la segmentation, on effectue par la suite un calcul de rectangles englobants sur toutes les régions détectées dans l'image possédant une aire supérieure à 20 pixels, tel qu'illustré à la figure 2.3. Celle présentant une aire inférieure sont éliminées du traitement. Il s'agit d'un seuil arbitraire visant, d'une part à éliminer les faux positifs, c'est-à-dire les pixels égarés qui devraient faire partie de l'arrière-plan, et d'autre part à éviter que l'on considère dans l'estimation du mouvement des véhicules trop éloignés qui n'occupent qu'une petite portion de l'image et qui pourraient fausser les résultats du calcul du flux optique.

L'utilisation des rectangles englobants permet aussi d'optimiser le calcul du flux op-



Figure 2.3 – Évaluation des régions d'intérêt.

tique, puisque celui-ci sera toujours évalué sur des régions rectangulaires plutôt que sur des régions de formes différentes. Comme on considère que le mouvement des objets est de faible amplitude par rapport à la taille de l'image, on utilise la position des rectangles englobants au temps  $t$  pour évaluer le flux optique entre les images des temps  $t$  et  $t + \Delta t$ .

## 2.2 Estimation du mouvement

L'estimation du mouvement se base sur l'étude de la position des objets dans une séquence vidéo. On recherche un ensemble de vecteurs décrivant le déplacement des différents éléments de la scène entre deux ou plusieurs images consécutives. Les algorithmes d'estimation du mouvement sont parmi les plus utilisés en vision par ordinateur. Ce sont par exemple sur ceux-ci que reposent les fonctionnalités de stabilisation d'images que l'on retrouve dans beaucoup de caméras numériques. Ils sont aussi au centre de méthodes répandues de compression vidéo telles que le MPEG et le H.263 [51]. Dans le cadre de l'estimation de la vitesse du trafic routier, l'utilisation des vecteurs de mouvement sera assez directe : ils serviront à évaluer le déplacement des véhicules.

L'estimation du mouvement entre les éléments de deux images d'une séquence vidéo se fera ici à l'aide du flux optique. Celui-ci est représenté par un champ vectoriel encodant l'information sur le mouvement apparent des objets dans l'image. Ces techniques exploitent la contrainte de constance des intensités, c'est-à-dire le fait que l'intensité des

pixels composant un objet en mouvement varie très peu en cours de déplacement [30].

Les méthodes classiques pour le calcul du flux optique peuvent être séparées en deux familles : les techniques locales qui cherchent à optimiser une fonction d'énergie sur de petites régions dans l'image et les approches globales, où la fonction d'énergie à optimiser fait intervenir la totalité des données disponibles dans l'image considérée [7]. Dans la première catégorie, on retrouve entre autres la technique proposée par Lucas et Kanade en 1981 [36], tandis que les modèles suggérés en 1981 par Horn et Schunck [27] et en 1991 par Black et Anandan [6] font partie de la deuxième famille. Les méthodes dites locales présentent une plus grande robustesse au bruit dans les images, c'est-à-dire lorsque les contours des objets de la scène sont moins bien définis. En contrepartie, ces mêmes méthodes ne permettent pas d'obtenir une représentation dense du flux optique. C'est donc dire qu'il n'est pas possible d'avoir une estimation du mouvement à chaque pixel  $(i, j)$  de chaque image de la séquence vidéo. Les méthodes globales, quant à elles, permettent d'obtenir cette représentation dense, mais sont aussi plus sensibles au bruit [7]. Dans cette optique, comme notre méthode repose sur l'analyse du mouvement des objets sur une image entière ou sur une région de l'image, il importe de disposer des données relatives aux déplacements en chaque point. C'est pourquoi nous optons pour une formulation globale du calcul du flux optique, soit celle proposée par Horn et Schunck [27].

### 2.2.1 Calcul du flux optique

Soient deux images d'une séquence vidéo :  $I(i, j, t)$  correspondant au temps  $t$  et  $I(i, j, t + \Delta t)$  correspondant au temps  $t + \Delta t$ . Le calcul du flux optique suppose qu'un certain point  $(i_1, j_1)$  dans l'image  $I(i, j, t)$  peut être mis en correspondance avec un point



$(i_2, j_2)$  dans l'image  $I(i, j, t + \Delta t)$ , c'est-à-dire que

$$\begin{aligned} I(i_1, j_1, t) &\approx I(i_2, j_2, t + \Delta t) \\ &\approx I(i_1 + \Delta i, j_1 + \Delta j, t + \Delta t), \end{aligned}$$

où  $\Delta i = i_2 - i_1$  et  $\Delta j = j_2 - j_1$ .

Le champ vectoriel associé au flux optique entre deux images sera composé des valeurs de mouvement relatifs  $u$  et  $v$ . Pour chaque pixel d'une image, on aura

$$u = \frac{\Delta i}{\Delta t} \quad \text{et} \quad v = \frac{\Delta j}{\Delta t}.$$

Avec cette notation, on a la formulation suivante pour les vecteurs  $(u, v)$  du flux optique :

$$I(i + u\Delta t, j + v\Delta t, t + \Delta t) - I(i, j, t) = 0.$$

De plus, en faisant la supposition que le déplacement des objets est de faible amplitude en comparaison avec la taille de l'image, on peut remplacer les variations discrètes  $\Delta i$ ,  $\Delta j$  et  $\Delta t$  par leurs équivalents infinitésimaux  $di$ ,  $dj$  et  $dt$  [11]. L'équation devient alors

$$I(i + u dt, j + v dt, t + dt) - I(i, j, t) = 0. \quad (2.1)$$

Soit  $f(x_1, \dots, x_n)$  une fonction différentiable. Son développement en série de Taylor d'ordre  $m$  permet d'écrire

$$\begin{aligned} f(x_1 + dx_1, \dots, x_n + dx_n) - f(x_1, \dots, x_n) &= df + \frac{1}{2!} d^2 f + \dots + \frac{1}{(m-1)!} d^{m-1} f + R_m \\ &\approx df, \end{aligned}$$

où  $R_m$  est le reste associé au développement d'ordre  $m$  et  $df$  est la différentielle totale de la fonction  $f$ , c'est-à-dire que

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \dots + \frac{\partial f}{\partial x_n} dx_n.$$

En appliquant le même principe à l'équation (2.1), on obtient la contrainte d'intensité proposée par Horn et Schunck pour le calcul du flux optique [27] :

$$\frac{\partial I}{\partial i} \frac{di}{dt} + \frac{\partial I}{\partial j} \frac{dj}{dt} + \frac{\partial I}{\partial t} = I_i u + I_j v + I_t = 0,$$

où  $I_i$ ,  $I_j$  et  $I_t$  sont les dérivées partielles de l'image  $I$ . Le problème est toutefois que l'équation précédente contient deux inconnues :  $u$  et  $v$ . Il est donc impossible de déterminer de manière unique les valeurs de  $u$  et  $v$ . Pour contourner ce problème, Horn et Schunck [27] ont proposé l'utilisation d'une deuxième équation faisant intervenir une contrainte de lissage stipulant que le vecteur de mouvement associé à un point  $(i_0, j_0)$  devait être stable dans un certain voisinage  $\Omega$  autour de  $(i_0, j_0)$ . Le problème du calcul du flux optique  $(u_0, v_0)$  au point  $(i_0, j_0)$  devient alors un problème de minimisation :

$$(u_0, v_0) = \underset{u, v}{\operatorname{argmin}} E(u, v),$$

avec

$$E(u, v) = \int_{\Omega} ((I_i u + I_j v + I_t)^2 + \alpha(|\nabla u|^2 + |\nabla v|^2)) \, dx \, dy, \quad (2.2)$$

où  $\alpha > 0$  est le terme de lissage pour le flux optique. Une plus grande valeur de  $\alpha$  engendre un flux optique plus constant. La minimisation de cette fonction revient à résoudre ses équations d'Euler-Lagrange, c'est-à-dire

$$\begin{aligned} \Delta u - \frac{1}{\alpha}(I_i^2 u + I_i I_j v + I_i I_t) &= 0, \text{ et} \\ \Delta v - \frac{1}{\alpha}(I_i I_j u + I_j^2 v + I_j I_t) &= 0, \end{aligned}$$

où  $\Delta$  représente l'opérateur laplacien.

Maintenant que le problème de l'évaluation du flux optique entre deux images a été bien défini, il faut s'attarder aux difficultés de sa mise en application. Il est en effet important de distinguer le modèle utilisé (la fonction objectif définissant le problème) de la méthode de résolution (algorithme d'optimisation et détails d'implantation). Ainsi,

avec la technique d'optimisation itérative proposée originalement par Horn et Schunck, les résultats obtenus à l'aide de la fonction objectif de l'équation 2.2 étaient grandement imprécis [5]. Cette façon de faire n'était toutefois qu'une heuristique et ne fournissait donc qu'une estimation de la véritable solution au problème d'optimisation. Cependant, lorsqu'optimisée à l'aide de techniques actuelles [34, 49, 50], la fonction objectif de Horn et Schunck performe bien par rapport à des méthodes plus récentes [4].

Précédemment, nous avons mentionné que les méthodes locales et globales pour l'évaluation du flux optique présentaient des avantages et des inconvénients complémentaires. Les méthodes locales sont plus robustes au bruit dans les images, mais ne permettent pas d'obtenir une représentation dense du flux optique, et l'inverse pour les méthodes globales. Le procédé d'optimisation utilisé dans notre méthode est celui proposé par Sun *et al.* [50]. À la manière de la technique d'optimisation de Bruhn et Weickert [7], l'approche utilisée propose de combiner les méthodes d'évaluation locales et globales afin de profiter des avantages de chacune.

Pour arriver à cette fin, Sun *et al.* utilisent entre autres une approche incrémentale à plusieurs résolutions. Le point de départ se fait à une résolution faible et le flux optique estimé à cette résolution est utilisé comme initialisation pour la minimisation de la fonction objectif, un niveau de résolution plus haut. Ce processus est répété jusqu'au niveau de résolution original de l'image. Avec cette idée, on diminue l'importance de la contrainte de faible amplitude du déplacement des objets qui a été mentionnée précédemment.

Une précision s'impose toutefois : il ne faut pas confondre l'utilisation de plusieurs niveaux de résolutions pour l'estimation du flux optique et la technique d'analyse multi-résolution des données de mouvement qui est détaillée à la section 2.2.2. Même si l'idée de base est la même, soit celle d'utiliser plusieurs résolutions d'une même image, l'objectif visé est différent. D'un côté on souhaite évaluer le flux optique entre deux images, et de l'autre, on veut analyser les données de déplacement pour extraire des informations

quant à la vitesse dans la scène.

Comme on le verra à la section suivante, c'est la simplicité de la fonction à optimiser, de même que son bon niveau de performance qui font de cette technique un choix pertinent pour l'analyse du trafic routier. Il existe des méthodes qui permettent d'obtenir des résultats plus précis, mais ce gain de performance possède un coût important en temps d'exécution [4]. Ainsi, dans le cadre de la présente méthode d'analyse du trafic routier, on fera appel à l'implantation récente de la méthode de Horn et Schunck réalisée par Deqing Sun [49]. Cette dernière tire profit des optimisations de Matlab pour le calcul matriciel et s'intègre bien à notre séquence de traitement.

Pour représenter visuellement les données de flux optique, nous utilisons le code de couleurs en espace colorimétrique HSV proposé par Baker *et al.* [3] et illustré à la figure 2.4. Ce code de couleurs permet de facilement connaître l'orientation et l'amplitude des vecteurs formant le champ vectoriel associé aux déplacements des objets entre deux images consécutives ou rapprochées dans la séquence. En effet, chaque teinte correspond à un angle différent sur le cercle chromatique et donne ainsi l'information sur l'orientation et la direction du vecteur. La saturation représente quant à elle la distance entre le centre du cercle et une couleur, donnant ainsi l'information sur l'amplitude du vecteur.

### 2.2.2 Flux optique et trafic routier

Dans le contexte de l'analyse du trafic routier, les données du flux optique ont principalement été utilisées comme critère de détection des objets en mouvement [30]. Pour faire la segmentation de tels objets dans une séquence obtenue à l'aide d'une caméra fixe, il suffit d'appliquer un seuil sur l'amplitude des vecteurs formant le flux optique entre deux images. On localisera ainsi facilement les régions correspondant aux véhicules en mouvement. Une telle approche est cependant très coûteuse en temps d'exécution [18].

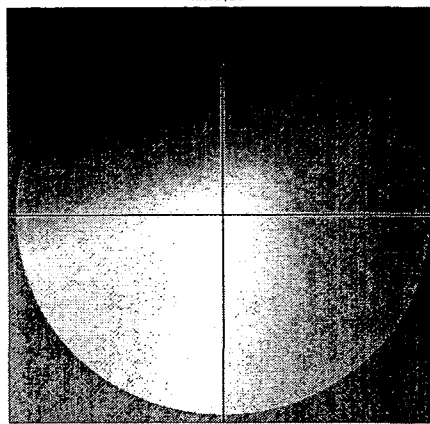


Figure 2.4 – Code de couleurs pour la visualisation du flux optique, proposé par [3].

Le flux optique rend toutefois possible la détection d'objets en mouvement lorsque la caméra faisant l'acquisition est elle aussi mobile, en étant fixée à un autobus [41] ou à un hélicoptère [46], par exemple. Ainsi, lorsqu'un véhicule se déplace sur une route, l'environnement stationnaire qui l'entoure crée un champ vectoriel présentant une structure stable et prévisible [8]. Le flux optique décrivant le mouvement d'une caméra dans un environnement rigide porte le nom d'égo-mouvement. Si un nouvel objet en mouvement devient visible par la caméra en déplacement, le flux optique de l'objet perturbera le flux optique associé à l'égo-mouvement, ce qui rendra possible sa détection [14].

Dans le cadre de ce travail, on suppose toutefois que la position des caméras n'est pas modifiée sur la durée de la séquence vidéo. Ce faisant, la détection des véhicules à l'aide d'une approche statistique, telle que présentée à la section 2.1.1 est plus appropriée. Ainsi, l'idée de localiser des régions d'intérêt avant d'évaluer le flux optique afin de réduire la quantité de données à traiter a déjà été utilisée dans le cadre d'algorithmes de surveillance vidéo [52].

L'estimation de la vitesse des véhicules à l'aide du flux optique n'est pas une idée

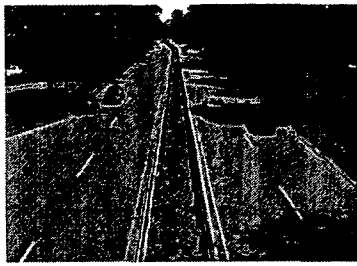
nouvelle [13, 18], mais elle demeure peu exploitée, sans doute en raison de la complexité des calculs à effectuer. S'il n'est pas vraiment envisageable de rendre simple l'évaluation du flux optique, l'approche proposée ici a l'avantage de limiter la quantité de calculs à effectuer de deux façons : d'une part en isolant les régions d'intérêt à l'aide d'une segmentation, et d'autre part en proposant une approche multirésolution de l'analyse du flux optique.

### 2.2.3 Analyse multirésolution du flux optique

Dans la section 2.2.1, on mentionnait comment l'utilisation d'une pyramide d'images pouvait accélérer les calculs du flux optique entre deux images. Le principe est d'utiliser une série d'images de résolutions inférieures comme approximations successives du champ vectoriel associé à l'image de pleine résolution. On limite ainsi la quantité de calcul à effectuer sur les données complètes, et on les remplace par un traitement moins gourmand sur des images de basse résolution. C'est cette idée qui sera adaptée à l'analyse du flux optique associé au trafic routier.

Tel que mentionné en début de travail, l'objectif de la méthode proposée n'est pas d'évaluer avec une grande précision la vitesse de chaque véhicule dans le but, par exemple, de donner des contraventions pour excès de vitesse. On cherche plutôt à obtenir un indice global sur la fluidité du trafic. Partant de ce principe, on peut établir la règle suivante : si le trafic est fluide, aucun traitement supplémentaire n'est nécessaire. Les problèmes surviendront lorsque le trafic perdra son rythme de croisière. Une analyse plus poussée du comportement des véhicules et éventuellement une recherche de la cause du ralentissement devront être effectuées. Ces problématiques dépassent toutefois le cadre de ce travail.

Dans cette optique, la méthode proposée suggère dans un premier temps, de construire



(a) Résolution R1.



(b) Résolution R2.



(c) Résolution R4.

Figure 2.5 – Niveaux de résolution de la pyramide vidéo.

une pyramide vidéo à trois niveaux de résolution pour chacune des images de la séquence, tel qu'illustré à la figure 2.5. Ces trois niveaux seront respectivement appelés R1 (image de dimensions originales), R2 (décimation de facteur 2 en  $x$  et en  $y$  de l'image originale) et R4 (décimation de facteur 4 en  $x$  et en  $y$  de l'image originale). Comme le système doit pouvoir fonctionner en direct, c'est-à-dire en recevant l'information vidéo à mesure qu'elle est acquise, il n'est évidemment pas nécessaire de disposer de la séquence en entier avant le début du traitement. Les trois niveaux de résolution sont plutôt calculés à la volée, au moment où une image donnée est traitée. Le traitement d'une portion de la séquence vidéo est alors effectué selon le niveau de précision souhaité dans les résultats de l'analyse du flux optique.

Trois niveaux de précision sont proposés, chacun de ceux-ci étant associé à un niveau de la pyramide vidéo. L'information obtenue consistera, au choix, en

- une estimation de la vitesse moyenne dans la scène (résolution R4) ;
- une estimation de la vitesse moyenne par direction principale (résolution R2) ;
- une estimation de la vitesse instantanée pour chaque véhicule (résolution R1).

## Estimation de la vitesse moyenne dans une scène

Comme une moins grande précision est nécessaire pour estimer une mesure de vitesse grossière, on utilise moins d'information, ce qui permet d'accélérer le traitement. Ainsi, à partir d'une séquence vidéo ayant une résolution spatiale correspondant au quart de la résolution originale, on évalue une vitesse moyenne  $v_s$  sur l'ensemble de la scène. Ce déplacement moyen permet de définir un indice de fluidité global.

En plus de fournir une estimation de la vitesse moyenne dans la scène qui mène à un indice de fluidité, on peut utiliser les données du flux optique pour évaluer le nombre de directions principales  $n_d$ , c'est-à-dire le nombre de directions où du mouvement est perçu sur la route. Cette information sera utile au niveau de précision supérieur afin d'évaluer la vitesse moyenne par direction. Pour y arriver, on construit un histogramme de mouvement 2D à partir du flux optique, tel que proposé par Zhang *et al.* [56]. Pour obtenir l'information souhaitée, on n'utilise que les vecteurs de flux optique ayant servi au calcul de la moyenne. Un tel histogramme informe sur la répartition des vecteurs du flux optique en fonction de leur direction et de leur amplitude. La figure 2.8 présente un exemple d'un histogramme de mouvement. Le nombre de directions principales  $n_d$  est ensuite obtenu en appliquant un seuil sur l'histogramme et en évaluant le nombre de régions connexes dépassant ce seuil, tel qu'illustré à la section 3.2.1. La figure 2.6 présente les étapes de cette partie de la méthode proposée.

La vitesse moyenne  $v_s$  est évaluée en prenant en considération seulement les pixels  $(x_i, y_j)$  de l'image qui présentent un vecteur de flux optique  $\mathbf{v}_{ij}$  dont l'amplitude est supérieure à un seuil  $S_A > 0$ . De cette façon, on évite que  $v_s$  soit artificiellement tiré vers le bas par des pixels correspondant à des régions ne présentant aucun mouvement. Ainsi, on a que

$$v_s = \frac{1}{n_A} \sum_{|\mathbf{v}_{ij}| > S_A} |\mathbf{v}_{ij}|, \quad (2.3)$$



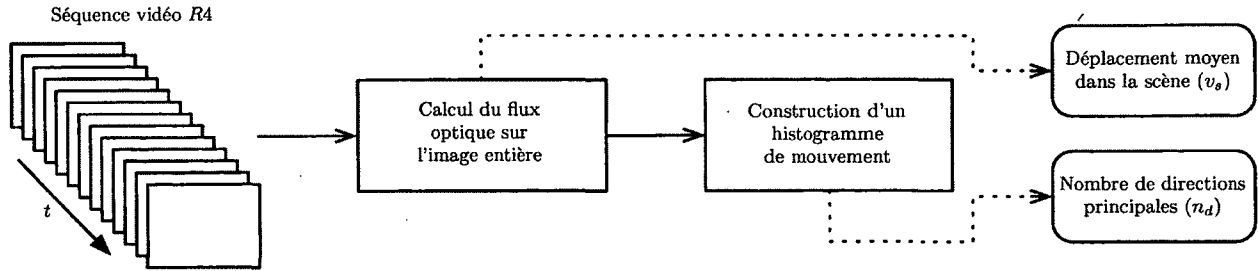


Figure 2.6 – Processus d’estimation de la vitesse moyenne dans une scène.

où  $n_A$  est le nombre de pixels auxquels est associé un vecteur de flux optique dont l’amplitude est supérieure à  $S_A$ . On atteint cet objectif en appliquant sur l’histogramme un seuil de manière à isoler les intervalles contenant au moins 5% à 10% des pixels. En évaluant le nombre de regroupements connexes d’intervalles dépassant ce seuil (figure 2.7); on peut évaluer automatiquement le nombre de directions principales  $n_d$ . Dans le cas illustré, nous trouvons  $n_d = 2$ .

La décision d’évaluer ici le flux optique sur l’image en entier plutôt que sur des régions d’intérêt est motivée par la mauvaise performance observée des algorithmes de détection d’objets en mouvement sur des images de très faible résolution.

### Estimation de la vitesse moyenne par direction

Si l’indice de fluidité global obtenu à partir de la vitesse moyenne dans la scène chute sous un certain seuil  $S_1$ , la précision des calculs sera augmentée pour permettre de mieux localiser les problèmes. On s’intéresse alors à la vitesse moyenne pour chacune des directions principales dans la scène.

Pour obtenir celles-ci, on utilise la résolution R2 correspondant à la moitié de la résolution originale. À partir de ces images, on localise les véhicules en mouvement par modélisation statistique de l’arrière-plan, ce qui permet d’obtenir des régions d’intérêt

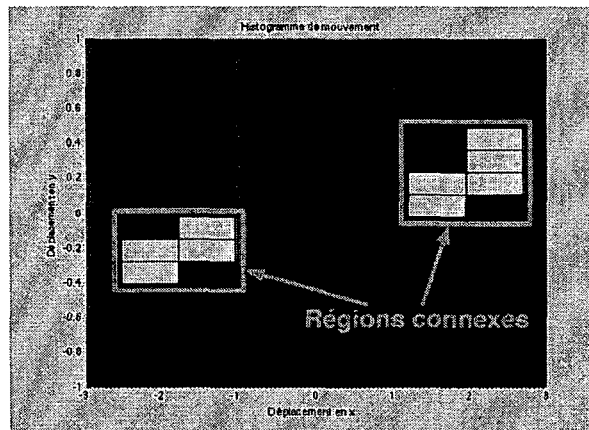
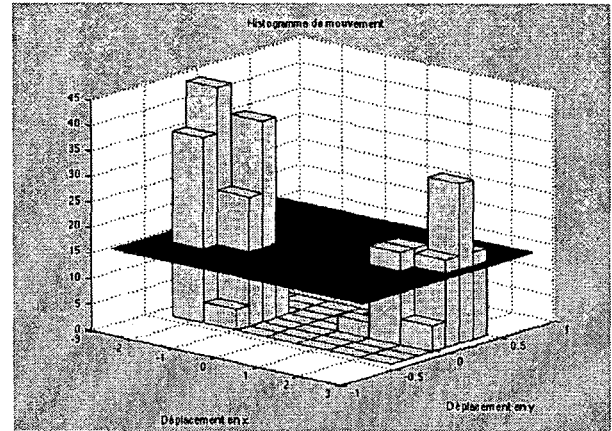
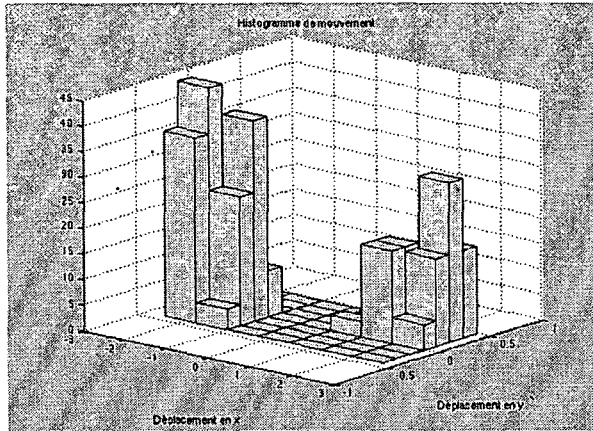
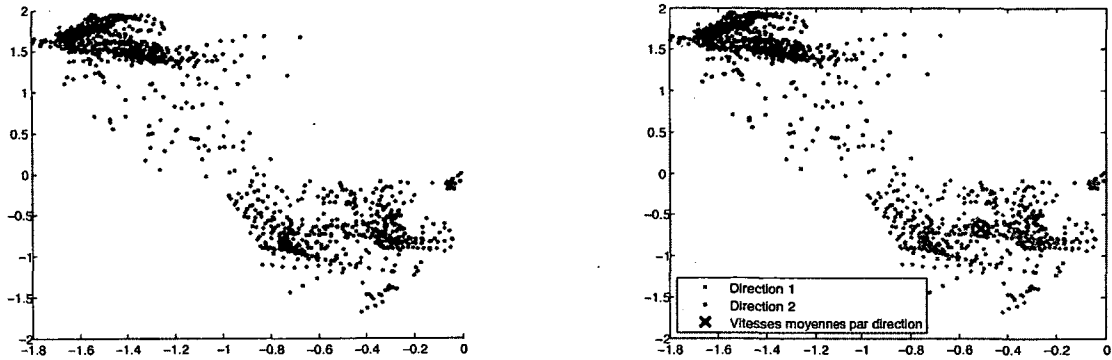


Figure 2.7 – Évaluation du nombre de directions de trafic.



Figure 2.8 – Histogramme de mouvement, tiré de [56].



(a) Flux optique sur les régions d'intérêt. (b) Vitesses moyennes des directions principales.

Figure 2.9 – Estimation de la vitesse moyenne par direction

dans l'image, puis on évalue le déplacement uniquement sur celles-ci. Dans l'implantation, chaque région d'intérêt est considérée comme une petite image indépendante des autres régions pour le calcul du flux optique. En limitant le calcul du flux optique aux régions où nous savons qu'il y a du mouvement, nous diminuons de nouveau la quantité de données à traiter.

On peut représenter les vecteurs du mouvement du flux optique par des points sur le plan correspondant au déplacement en  $x$  et  $y$ , tel qu'illustré à la figure 2.9a. Connaissant le nombre de directions principales  $n_d$ , on peut segmenter les données en  $n_d$  classes à l'aide de l'algorithme des  $k$ -moyennes, et évaluer les centroïdes de chacune de celles-ci. Ces centroïdes correspondent aux vitesses moyennes de chacune des directions; tel que présenté à la figure 2.9b. L'algorithme des  $k$ -moyennes est une méthode dont le but est de diviser des observations en  $k$  partitions dans lesquelles chaque observation appartient à la partition avec la moyenne la plus proche.

La figure 2.10 illustre les étapes de traitement associées à la résolution R2. De plus, à partir des données du flux optique, on peut continuer à évaluer la vitesse moyenne  $v_s$  de même que le nombre de directions principales  $n_d$  que l'on évaluait précédemment à

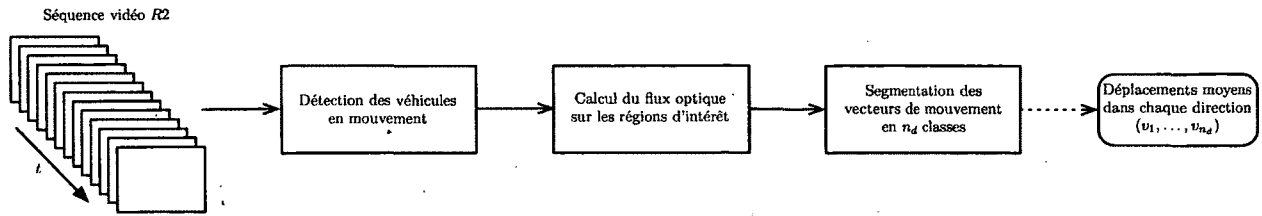


Figure 2.10 – Processus d’estimation de la vitesse moyenne par direction.

partir de la résolution R4. Deux options sont alors envisageables : soit la vitesse moyenne  $v_s$  augmente et l’indice de fluidité global repasse au-dessus du seuil  $S_1$  et le système considère que la situation est revenue à la normale, soit la vitesse  $v_s$  demeure trop basse et le système décide de raffiner de nouveau le traitement des données. Dans le premier cas, cela entraîne un retour à la résolution R4, tandis que dans l’autre on passe à la résolution  $R_1$  pour obtenir encore plus de précision sur l’état du trafic.

### Estimation de la vitesse par véhicule

Dans l’éventualité où l’analyse des données du flux optique doivent encore être raffinées, on passe à la résolution R1, soit la résolution originale des données vidéo. Avec les données à pleine résolution, on met à profit les deux méthodes d’analyse des vecteurs de flux optique utilisées pour les résolutions R2 et R4 afin d’estimer une vitesse instantanée pour chaque véhicule de la scène. Comme à la résolution R2, le traitement débute par une soustraction de fond afin de mettre en évidence les objets en mouvement dans la scène. Ensuite, pour chacune des régions ainsi obtenues, nous évaluons le flux optique, toujours en suivant le même principe qu’à la résolution R2. Toutefois, pour la résolution R1, à la différence du traitement aux résolutions R2 et R4, l’analyse des données devient locale, c’est-à-dire que le flux optique est évalué sur chaque région correspondant à du mouvement et est analysé indépendamment des autres régions. En reprenant le procédé utilisé à la résolution R4, nous trouvons le nombre de véhicules dans chaque région, puis à

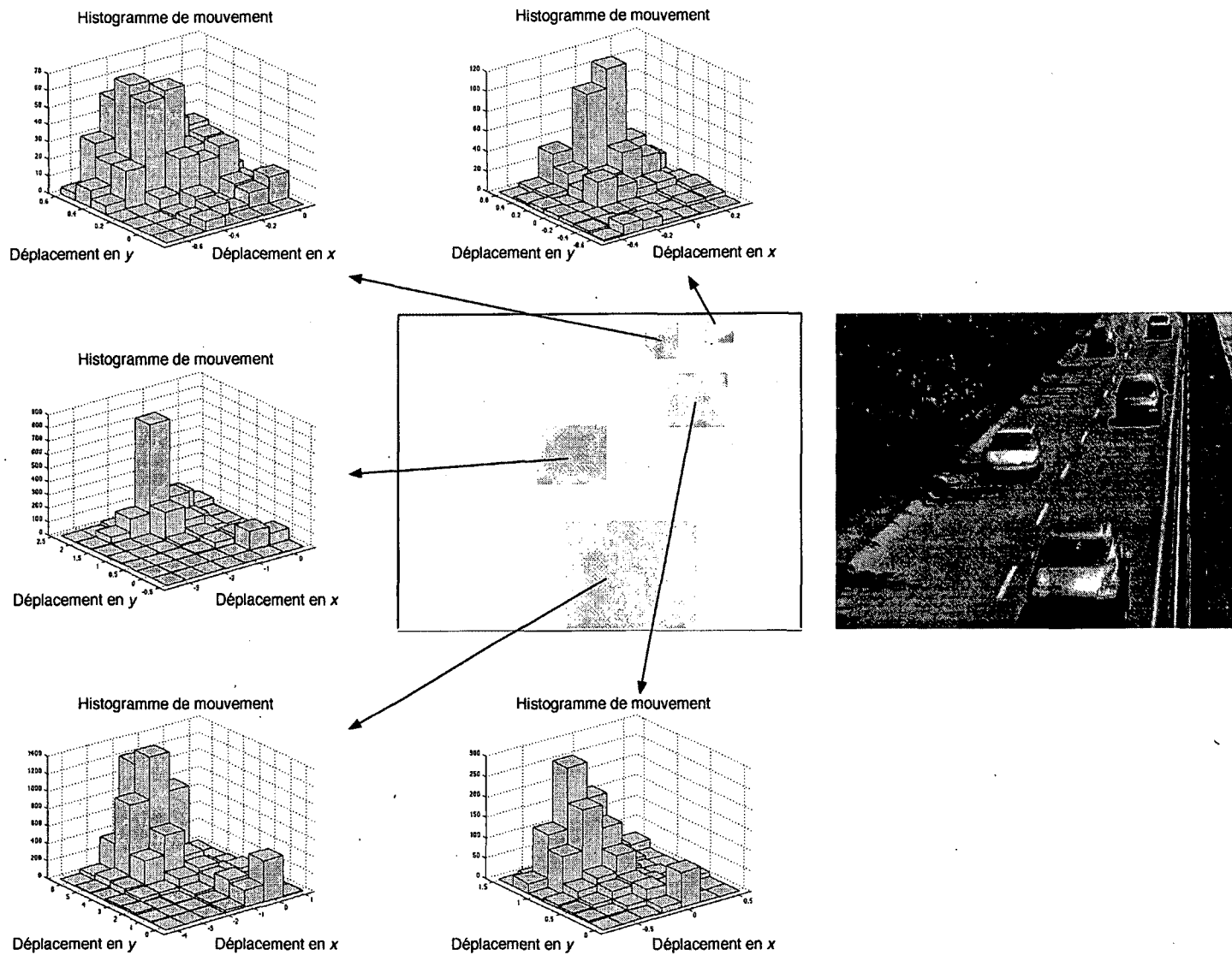
l'aide de l'algorithme des  $k$ -moyennes comme à la résolution R2, nous évaluons la vitesse instantanée de ceux-ci.

Pour chaque région correspondant à un objet en mouvement, on évalue le flux optique. Avec ces données, nous construisons pour chaque région un histogramme de mouvement (figure 2.11). En reprenant le procédé utilisé avec la résolution R4, nous évaluons le nombre de régions connexes suite à l'application d'un seuil sur l'histogramme (figure 2.12). Un seuil correspondant à 5% du nombre de pixels dans la région a été utilisé dans cet exemple. Pour toutes les régions, nous trouvons une seule région connexe et donc un seul véhicule par région. L'estimation du déplacement pour chaque véhicule est alors simple : il suffit d'évaluer le vecteur moyen parmi l'ensemble des vecteurs de flux optique de chaque région.

Dans le cas où une même région contient plus d'un véhicules, comme c'est le cas à la figure 2.13, l'analyse des régions connexes de l'histogramme donne un nombre de véhicules  $n_v > 1$  à condition que les deux voitures se déplacent à des vitesses ou dans des directions différentes. Dans ce cas, on sépare l'ensemble des vecteurs de flux optique de la région en  $n_v$  classes à l'aide de l'algorithme des  $k$ -moyennes. Chaque classe contient alors les vecteurs de flux optique qui sont associés à un seul véhicule. Les centroïdes de chaque classe permettent d'obtenir le déplacement instantané des véhicules, et à partir de celui-ci, nous pouvons évaluer la vitesse des voitures. Il devient alors possible d'identifier automatiquement les véhicules présentant un comportement hors normes. La figure 2.14 illustre les étapes de traitement associées à la résolution R1.

Finalement, comme aux résolutions inférieures, on peut continuer à utiliser l'ensemble des données de flux optique des objets en mouvement pour évaluer la vitesse moyenne dans la scène, ce qui permet de mettre à jour l'indice de fluidité global du trafic. De cette façon, le système peut, en cas de retour à la normale, revenir aux résolutions inférieures. La méthode proposée étant adaptative, ce ne sont pas toutes les portions d'un réseau

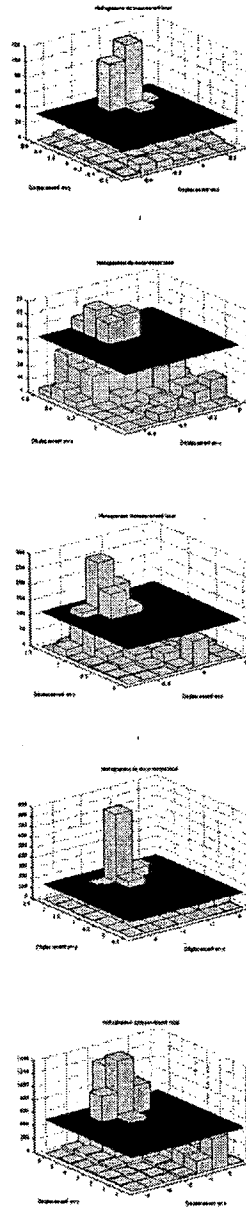
Figure 2.11 – Construction d'un histogramme pour chaque région en mouvement.



Flux optique local



Histogramme de mouvement



Régions connexes

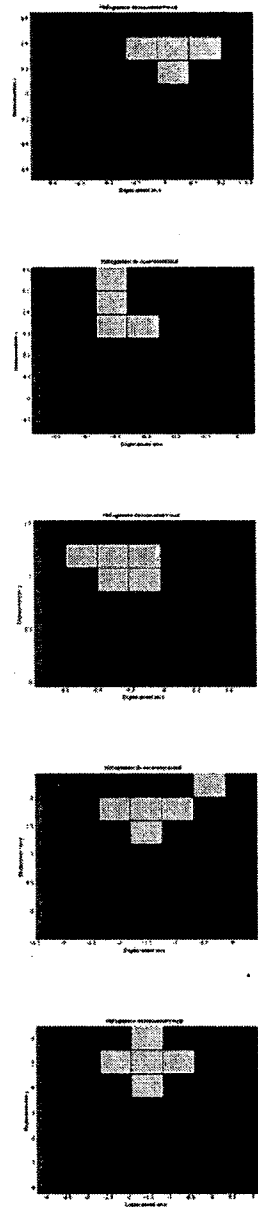
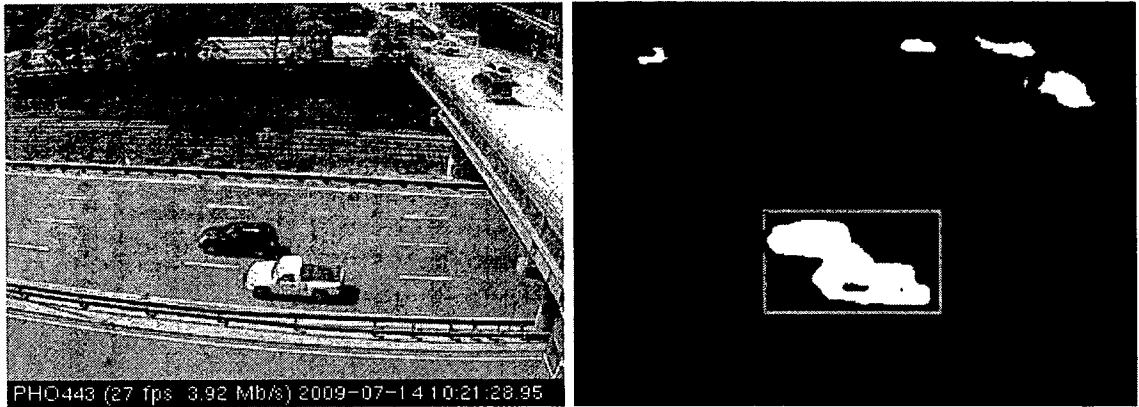


Figure 2.12 – Évaluation du nombre de véhicules par région sur la séquence *highway* ( $t = 1070$ ) de la série *Base*.





(a) Image originale R1.

(b) Détection des objets en mouvement.

Figure 2.13 – Séquence *streetLight* ( $t = 925$ ) de la série *Mouvement intermittent*.

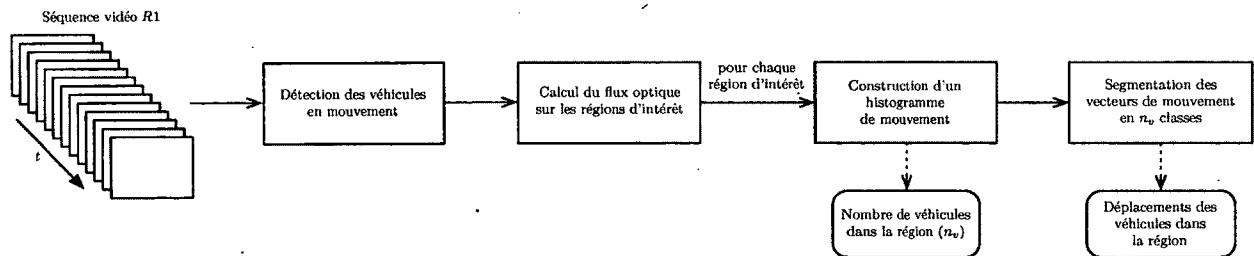


Figure 2.14 – Processus d'estimation de la vitesse par véhicule.

routier qui devront être analysées en résolutions R2 et R1 simultanément en cas de problème. L'analyse des différents points de vue se fera majoritairement en résolution R4, ce qui augmente le potentiel de traitement d'un parc informatique.

En résumé, nous avons présenté une méthode d'estimation de la vitesse du trafic automobile sur le réseau routier à partir de l'information visuelle multirésolution dont la figure 2.15 présente les étapes principales. Le choix de travailler à plusieurs niveaux de précision diminue le temps de traitement requis pour une séquence. Présentement, les déplacements de véhicules sont mesurés en pixels. Cela est dû au fait que, tel que mentionné en début de chapitre, nous n'effectuons pas d'étape de calibration de la caméra pour faire la rectification des images. Ce faisant, les images analysées peuvent présenter une déformation due à la perspective, ce qui empêche l'évaluation de déplacements réels. Une autre hypothèse qui a été faite est celle de la stabilité de la luminosité dans la scène. Un changement brusque des conditions d'illumination rendrait nécessaire le calcul d'un nouveau modèle statistique pour l'arrière plan. La méthode proposée demeure fonctionnelle dans la mesure où les intensités correspondant aux véhicules peuvent être distinguées de celle associées à l'arrière-plan. Ce faisant, la méthode n'est pas fonctionnelle la nuit. Maintenant que nous avons vu les fondements théoriques de la méthode proposée, attardons-nous à la validation expérimentale.

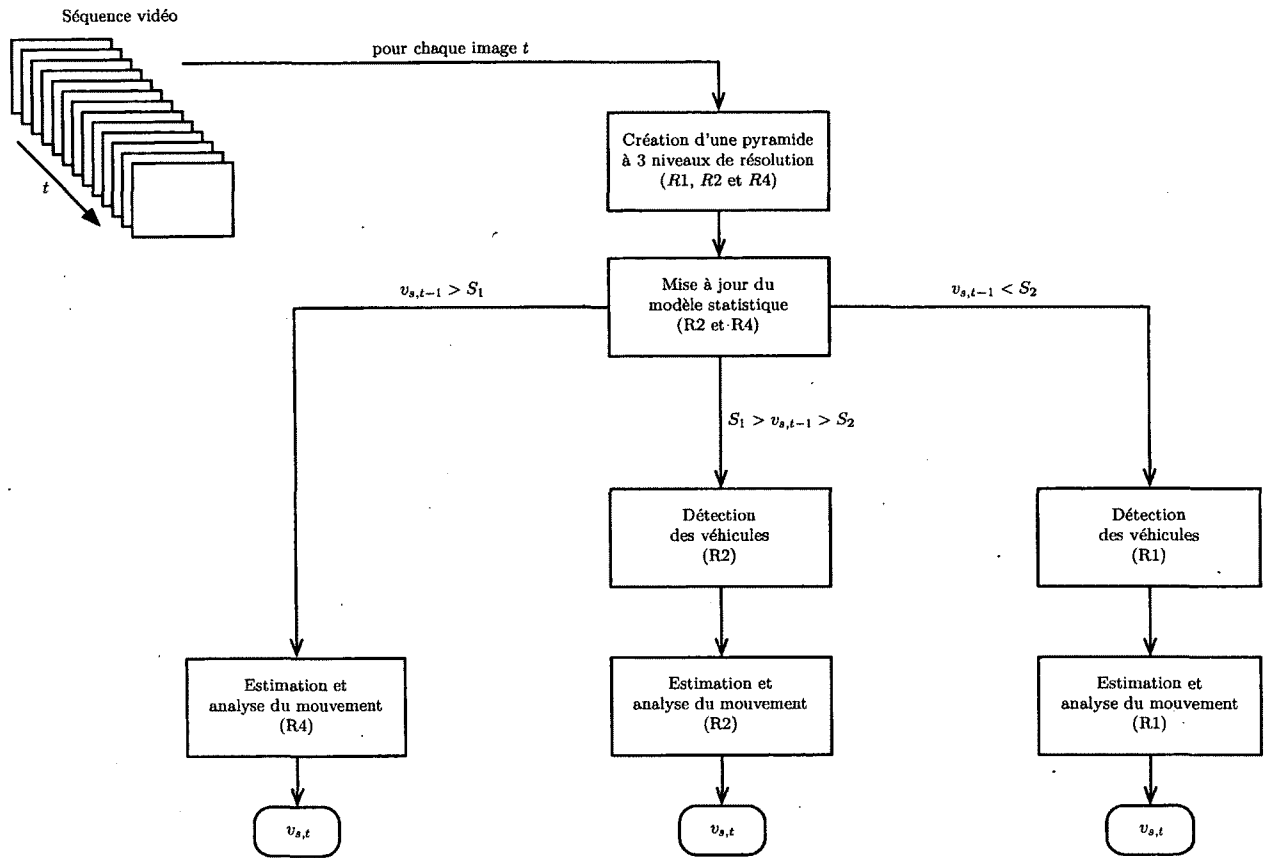


Figure 2.15 – Organigramme de la méthode d'estimation de la vitesse du trafic automobile.

# CHAPITRE 3

## Résultats expérimentaux

Dans ce chapitre, nous nous intéressons à deux problématiques principales. Dans un premier temps, nous cherchons à valider expérimentalement les résultats générés par la méthode de détection des véhicules en mouvement préalable à l'estimation de la vitesse du trafic routier. Par la suite, nous nous attardons à l'approche d'évaluation des déplacements par le flux optique combinée à la soustraction de fond. Nous discutons finalement de l'impact de l'utilisation d'une approche multirésolution, à savoir les conséquences sur la précision des résultats.

### 3.1 Détection des véhicules en mouvement

#### 3.1.1 Méthodologie d'évaluation

Afin de valider les résultats obtenus par la méthode de détection des véhicules, nous utilisons des séquences vidéo de la banque *changedetection.net* [20]. Leur ensemble de test comprend 31 séquences vidéo réelles (aucune séquence synthétique) séparées en six

Tableau 3.1 – Données d'évaluation de la performance d'un classificateur

	Vérité terrain	
	Mouvement	Statique
Résultat de la segmentation		
<div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <span style="display: inline-block; width: 10px; height: 10px; background-color: #cccccc; border: 1px solid black; margin-bottom: 2px;"></span> Mouvement         </div> <div style="text-align: center;"> <span style="display: inline-block; width: 10px; height: 10px; background-color: #808080; border: 1px solid black; margin-bottom: 2px;"></span> Statique         </div> </div>	<div style="background-color: #cccccc; padding: 5px;">           positif (<math>T_P</math>)  <small>(comme mouvement)</small> </div>	<div style="background-color: #808080; padding: 5px;">           Faux positif (<math>F_P</math>)  <small>(statique détecté comme mouvement)</small> </div>
	<div style="background-color: #808080; padding: 5px;">           négatif (<math>F_N</math>)  <small>(é comme statique)</small> </div>	<div style="background-color: #808080; padding: 5px;">           Vrai négatif (<math>T_N</math>)  <small>(statique détecté comme statique)</small> </div>

catégories élaborées pour proposer différents défis aux algorithmes de détection d'objets en mouvement. Les séquences se veulent représentatives des données visuelles captées par des systèmes de surveillance intérieurs et extérieurs. La banque de données inclut aussi, pour chaque séquence, une vérité terrain élaborée à la main. Cette dernière permet une comparaison quantitative précise des méthodes testées. Même si le sujet de ce travail porte sur l'évaluation du mouvement des véhicules, des séquences illustrant autre chose que des scènes de trafic routier ont été utilisées pour les tests. Les défis associés à la détection des objets en mouvement ne dépendent pas vraiment du contenu sémantique des séquences. En plus d'évaluer la qualité de la détection, les tests portent sur la gestion d'un arrière-plan dynamique, sur les changements brusques dans la position et l'orientation de la caméra et sur les mouvements intermittents des objets.

Dans un contexte de segmentation en deux classes comme nous avons présentement (statique et mouvement), les notions de *vrai positif*, *vrai négatif*, *faux positif* et *faux négatif* permettent d'évaluer la performance d'un algorithme de classification par rapport à une vérité terrain. Les termes *positif* et *négatif* réfèrent au résultat du classificateur : *positif* correspond au mouvement et *négatif* correspond au statique. Les termes *vrai* et *faux* indiquent quant à eux si le résultat de la segmentation correspond à la réalité. Le tableau 3.1 résume ces notions. À partir de ces définitions, on établit une série de mesures

pour évaluer et comparer le rendement des méthodes de segmentation. La procédure de test de *changedetection.net* [20] indique à la fois sur quelle portion de l'image porte le test (région d'intérêt dans l'image) et donne aussi un intervalle temporel indiquant sur quelle portion de chaque séquence les mesures doivent être calculées. Ces mesures font intervenir le nombre total de vrais positifs, de faux positifs, et faux négatifs et de vrais négatifs pour les régions d'intérêt sur l'ensemble des images considérées. Ces valeurs sont respectivement notées  $T_P$ ,  $F_P$ ,  $F_N$  et  $T_N$  dans les formules des mesures qui suivent :

- Le rappel ( $Re$ ) : il donne la probabilité qu'un pixel réellement en mouvement soit bien classé.

$$Re = \frac{T_P}{T_P + F_N}$$

Sa valeur est comprise entre 0 et 1, où 1 est la valeur idéale, puisqu'un pixel réellement en mouvement est alors toujours bien classé.

- La précision ( $Pr$ ) : elle donne la probabilité qu'un pixel classé comme étant en mouvement le soit réellement.

$$Pr = \frac{T_P}{T_P + F_P}$$

S'agissant d'une probabilité, sa valeur est aussi comprise entre 0 et 1: Une valeur de 1 indique que le classificateur ne se trompe jamais lorsqu'il établit qu'un pixel est en mouvement.

- La spécificité ( $Sp$ ) : elle donne la probabilité qu'un pixel réellement statique soit bien classé.

$$Sp = \frac{T_N}{T_N + F_P}$$

La valeur de spécificité est aussi située entre 0 et 1. Cette mesure est analogue au rappel, mais cette fois-ci pour le cas des pixels statiques.

- Le score  $F_1$  : il correspond à la moyenne harmonique de la précision et du rappel.

$$F_1 = \frac{2 \cdot Pr \cdot Re}{Pr + Re}$$

Cette mesure a aussi 1 comme meilleure valeur et 0 comme pire. Comme il s'agit d'une moyenne pondérée de la précision et du rappel, le score  $F_1$  résume en une seule mesure les performances du classificateur pour ces deux statistiques.

- Le taux de faux positifs ( $FPR$ ) : il donne la proportion de mauvaise classification pour les pixels réellement statiques.

$$FPR = \frac{F_P}{F_P + T_N}$$

Cette donnée mesure le taux d'erreurs du classificateur pour les pixels statiques selon la vérité terrain. Un petit taux d'erreur étant souhaitable, sa valeur idéale est 0 et la pire valeur est 1.

- Le taux de faux négatifs ( $FNR$ ) : il donne la proportion de mauvaise classification pour les pixels réellement en mouvement.

$$FNR = \frac{F_N}{T_P + F_N}$$

Cette mesure est analogue au  $FPR$ , mais pour les pixels en mouvement. Les meilleures et pires valeurs sont encore 0 et 1, respectivement.

- Le pourcentage de mauvaise classification : il donne le pourcentage d'erreur global (faux positifs et faux négatifs) du classificateur.

$$PWC = \frac{100(F_N + F_P)}{T_P + F_N + F_P + T_N}$$

En une seule mesure comprise entre 0 et 100, nous résumons les erreurs faites par le classificateur (faux négatifs et faux positifs). La meilleure valeur est 0 et la pire est 100, qui indiquerait que l'algorithme de segmentation donne systématiquement la mauvaise classe.

Ces mesures sont celles suggérées par *changedetection.net* [20]. Leur utilisation rend possible la comparaison de la méthode proposée à d'autres algorithmes de segmentation.

Comme mentionné précédemment, les séquences vidéo utilisées à des fins d'évaluation des méthodes sont regroupées en quatre catégories dans la banque de test de *changedetection.net* [20] :

- Base (*Baseline*), qui regroupe quatre séquences vidéo simples dont la longueur varie entre 1099 et 2050 images.
- Arrière-plan dynamique (*Dynamic Background*), qui comprend six séquences vidéo dont la longueur se situe entre 1184 et 7999 images et où l'arrière-plan présente un certain mouvement.
- Vibration de caméra (*Camera Jitter*), où quatre séquences vidéo possédant entre 1150 et 2500 images présentent d'importantes vibrations de caméra.
- Mouvement intermittent (*Intermittent Object Motion*), où six séquences vidéo de 2500 à 4500 images présentent des objets dont le mouvement est interrompu puis repris en cours de séquence.

Pour chaque séquence, on dispose d'une vérité terrain qui permet l'évaluation des statistiques détaillées précédemment. Ces statistiques sont mesurées, pour chaque séquence, sur plusieurs centaines d'images de manière à avoir un portrait général des performances et non pas un instantané pris à un moment précis. On s'assure ainsi d'évaluer le comportement des algorithmes de soustraction de fond non seulement dans une situation idéale (série *Base*), mais aussi dans des contextes problématiques pouvant survenir en pratique (séries *Arrière-plan dynamique*, *Vibration de caméra* et *Mouvement intermittent*).

Les tableaux de résultats associés à chaque catégorie de séquences vidéo présentent un compte rendu des performances des deux méthodes considérées. La première, notée LBP-Satpathy a été détaillée à la section 2.1.2 et repose sur une segmentation basée sur la texture. La seconde, notée GMM-Lee et détaillée à la section 2.1.3, fait appel à un mélange de gaussiennes pour distinguer les objets en mouvement de l'arrière-plan. En plus de ces deux méthodes, on inclut les données associées à quatre autres techniques



de soustraction de fond reposant sur un mélange de gaussiennes, afin de mieux situer les performances de la méthode choisie. Comme les fondements théoriques des approches basées sur les mélanges de gaussiennes sont les mêmes, les temps d'exécution sont similaires. Les méthodes proposées (LBP-Satpathy et GMM-Lee) sont donc comparées à des techniques équivalentes. Les résultats de ces quatre méthodes supplémentaires proviennent de *changedetection.net*. Parmi les quatre techniques supplémentaires, il y a celle proposée par Stauffer et Grimson [47] qui a été introduite à la section 2.1.1 en tant que source principale de la méthode proposée.

Finalement, en plus de la qualité des résultats aux tests de segmentation, il sera important que la méthode proposée présente d'assez bas temps de calcul afin de pouvoir maintenir une exécution en temps réel. La notion de temps réel est ici fixée à 15 images par seconde.

### 3.1.2 Résultats et analyse

#### Détection des objets basée sur la texture

Dans un premier temps, attardons-nous aux résultats de la méthode de segmentation LBP-Satpathy, basée sur la texture. Rappelons que cette méthode avait d'abord été considérée puisque les auteurs affirmaient qu'elle possédait une moins grande sensibilité aux changements d'illumination. Le nombre de paramètres à fixer pour cette technique est, aux dires même des auteurs, très élevé [24]. Ceci rend l'obtention d'une combinaison performante extrêmement difficile, puisque le choix des paramètres dépend grandement du contexte des séquences vidéo à traiter. Parmi les éléments qui influencent le choix des paramètres, notons principalement les variations dans l'illumination, la gestion des ombres et la structure multimodale de l'arrière-plan, c'est-à-dire lorsque plusieurs intensités doivent être associées au fond pour une même position de pixel.

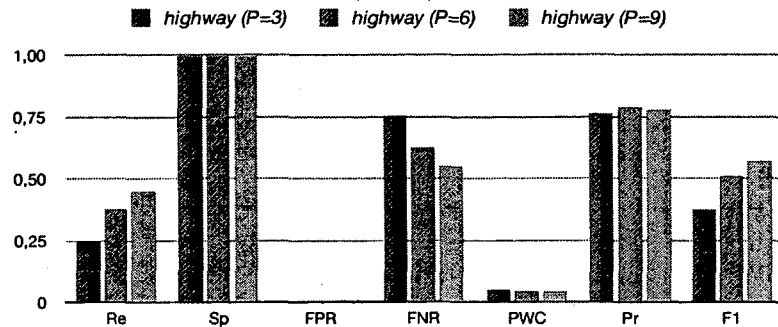


Figure 3.1 – Influence du choix de  $P$  sur les mesures d'évaluation de LBP-Satpathy

On doit d'abord déterminer le rayon  $R$  et le nombre de pixels  $P$  servant à calculer le motif binaire local d'un pixel donné, tel qu'illustré à la figure 2.2. Comme la corrélation entre les intensités de deux pixels diminue à mesure que la distance entre ceux-ci augmente, il est préférable d'utiliser une valeur de  $R$  correspondant à 2 ou 3 pixels. En ce qui concerne le choix de  $P$ , il faut faire un compromis entre la qualité des résultats et le temps d'exécution de l'algorithme. Une plus grande valeur de  $P$  améliorera le score de la méthode pour les différentes mesures d'évaluation du classificateur, mais augmentera aussi la quantité de calculs à effectuer. Heikkila *et al.* [24] recommandent d'utiliser  $P = 6$ , et nos expérimentations confirment ce choix. La figure 3.1 présente les résultats de la méthode LBP-Satpathy pour différentes valeurs de  $P$  sur la séquence *highway* de la série *Base*. L'utilisation de  $P = 9$  par rapport aux valeurs inférieures améliore le score de la méthode pour la plupart des mesures d'évaluation (rappel, ratio de faux négatif, score  $F_1$ ). Toutefois, cette augmentation de la qualité des résultats présente un intérêt limité en raison de son coût en termes de temps de calcul qui est alors doublé par rapport à  $P = 3$ . La vitesse de traitement passe en effet de 6 images par seconde pour  $P = 3$  à moins de 3 images par seconde pour  $P = 9$ , pour une séquence de résolution  $320 \times 240$ .

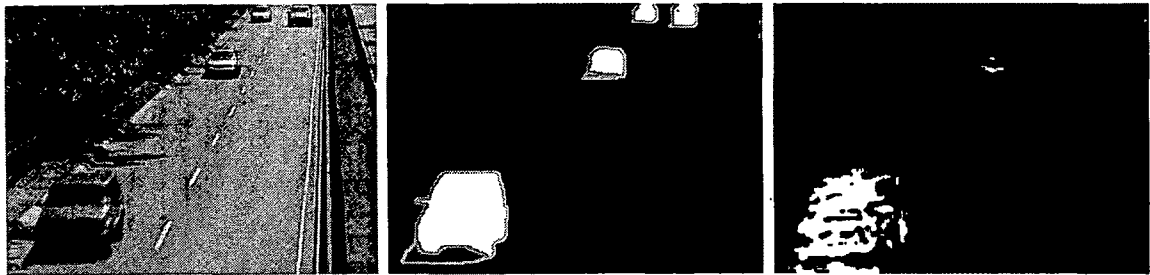
Ensuite, il est nécessaire de fixer le nombre d'histogrammes  $K$  à utiliser pour modéliser temporellement chaque pixel d'une séquence vidéo. La valeur de  $K$  à utiliser dépend de

la nature de l'arrière-plan de la scène. Pour un arrière-plan complètement statique, une petite valeur de  $K$  fera l'affaire. Toutefois, dès qu'on peut observer des variations dans les intensités associées à l'arrière-plan, il faut opter pour une plus grande valeur de  $K$ , ce qui diminue la vitesse de traitement de l'algorithme. On dira alors que la séquence possède un arrière-plan multimodal. Ces variations peuvent être causées, par exemple, par de brusques changements d'illumination dans la scène, par de l'ombre ou par un mouvement de la caméra dû au vent. Les auteurs suggèrent une valeur de  $K$  entre 2 et 5, selon la nature de l'arrière-plan.

On doit aussi trouver une valeur satisfaisante pour  $T_B \in [0, 1]$ , qui est le seuil déterminant quels histogrammes du modèle parmi les  $K$  sont associés à l'arrière-plan, pour chaque pixel. Le choix de  $T_B$  est intimement lié à celui de  $K$ . Une petite valeur de  $T_B$  est appropriée pour les séquences présentant un fond statique, tandis qu'une plus grande valeur doit être utilisée en présence d'un arrière-plan multimodal. Pour s'assurer de bien gérer la nature dynamique des arrière-plans, on utilise le seuil  $T_B = 0.8$ . On constate ainsi une plus faible sensibilité aux changements d'illumination soudains dans une séquence. Ce résultat est observable en comparant les figures 3.2 et 3.3 avec les figures 3.9 et 3.10 qui présentent le résultat d'une classification suivant de près un changement brusque d'illumination dans la séquence.

Finalement, les paramètres  $\alpha_b \in [0, 1]$  et  $\alpha_w \in [0, 1]$  sont des coefficients d'apprentissage qui déterminent la vitesse d'adaptation de l'algorithme aux observations récentes. En prenant des valeurs supérieures à 0.1 pour  $\alpha_b$  et  $\alpha_w$ , l'algorithme devient moins sensible aux déplacements lents des objets dans la scène. Ceux-ci sont alors considérés comme faisant partie de l'arrière-plan. Les auteurs suggèrent de prendre  $\alpha_b = \alpha_w = 0.01$  [24]. Nos expérimentations avec plusieurs valeurs allant de 0.001 à 0.2 confirment ce choix, puisque c'est avec  $\alpha_b = \alpha_w = 0.01$  que nous avons observé les meilleurs résultats.

Toutes les séquences vidéo de la banque de test de *changedetection.net* ne présentent



(a) Image originale

(b) Vérité terrain

(c) LBP-Satpathy

Figure 3.2 – Faible sensibilité aux changements d’illumination avec LBP-Satpathy (séquence *highway*,  $t = 930$ )



(a) Image originale

(b) Vérité terrain

(c) LBP-Satpathy

Figure 3.3 – Faible sensibilité aux changements d’illumination avec LBP-Satpathy (séquence *office*,  $t = 662$ )

pas des scènes de trafic routier. Considérant l'objectif premier de ce travail, qui est d'évaluer la vitesse de véhicules en mouvement à partir d'information visuelle, les paramètres des algorithmes ont été optimisés de manière à fournir les meilleures performances dans les séquences présentant des véhicules en mouvement. Notons aussi que les paramètres n'ont pas été modifiés indépendamment pour chacune des séquences de test. Pour une utilisation concrète de la méthode, il sera possible d'ajuster ceux-ci afin de maximiser la qualité des résultats dans un environnement donné.

Le tableau 3.2 présente les résultats moyens des différentes méthodes de soustraction de fond sur les quatre séquences de la série *Base*. Dans ce tableau et dans les suivants du même type, les meilleurs résultats pour chaque mesure de performance sont indiqués en gras, et les pires sont en italique. Comme les séquences d'une même catégorie présentent les mêmes défis pour l'algorithme de segmentation, l'utilisation de la moyenne des statistiques sur l'ensemble des séquences d'une même catégorie donne une bonne mesure de la qualité des résultats de la méthode, sans être trop sensible aux données aberrantes. Les meilleurs résultats sont identifiés en gras pour chaque mesure et, réciproquement, l'italique dénote les moins bonnes performances. D'entrée de jeu, on remarque que sur sept mesures, la méthode de soustraction de fond LBP-Satpathy présente à cinq reprises le moins bon résultat. Les deux statistiques où LBP-Satpathy n'arrive pas en dernière position sont la spécificité ( $Sp$ ) et le taux de faux positif ( $FPR$ ). La spécificité donne la probabilité qu'un pixel réellement statique soit bien classé, tandis que le taux de faux positif donne la proportion de mauvaise classification pour les pixels réellement statiques. Ainsi, un classificateur qui donnerait comme résultat que la totalité des pixels est statique aurait un score parfait sur ces deux mesures, mais serait potentiellement très bas pour les autres. C'est précisément ce qui semble se passer avec l'algorithme LBP-Satpathy. En mettant en relation les données statistiques du tableau avec les résultats de classification évalués avec les paramètres choisis ( $P = 6$ ,  $R = 3$  et  $K = 5$ ) pour la série *Base* présentés

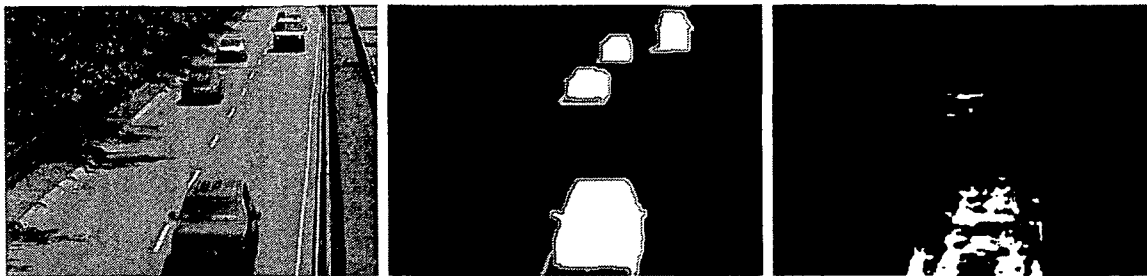
Tableau 3.2 – Performances moyennes pour la série *Base* [20]

	<i>Re</i>	<i>Pr</i>	<i>F<sub>1</sub></i>	<i>Sp</i>	<i>FPR</i>	<i>FNR</i>	<i>PWC</i>
<b>LBP-Satpathy</b>	0.2206	0.6041	0.2984	0.9968	0.0084	0.7793	3.3574
<b>GMM-Lee</b>	<b>0.8982</b>	0.6705	0.7628	0.9764	0.0295	<b>0.1017</b>	2.7851
<b>GMM-Stauffer [47]</b>	0.8180	0.8461	0.8245	0.9948	0.0052	0.1820	1.5325
<b>GMM-KaewTraKulPong [29]</b>	0.5863	<b>0.9532</b>	0.7119	<b>0.9987</b>	<b>0.0013</b>	0.4137	1.9381
<b>GMM-RECTGAUSS-TeX [42]</b>	0.6669	0.9175	0.7500	0.9979	0.0021	0.3331	1.5342
<b>GMM-Zivkovic [57]</b>	0.8085	0.8993	<b>0.8382</b>	0.9972	0.0028	0.1915	<b>1.3298</b>

(meilleur résultat, pire résultat)

aux figures 3.4 à 3.7, on peut tirer comme conclusion que la méthode LBP-Satpathy présente un biais important vers la classe de l'arrière-plan. Bien peu de pixels réellement statiques sont mal classifiés, mais un grand nombre de pixels appartenant à des objets en mouvement aboutissent dans la classe de l'arrière-plan. Ce résultat est particulièrement visible à la figure 3.5 où pratiquement tous les pixels ont été classifiés comme faisant partie de l'arrière-plan. On peut penser qu'un mouvement trop lent est à l'origine de ce mauvais résultat. Le même problème se produit à la figure 3.7 où le sac le long du mur demeure immobile pendant trop longtemps et est ainsi considéré comme faisant partie de l'arrière-plan par l'algorithme. Ce comportement n'est pas souhaitable dans un contexte d'analyse du trafic routier. En situation d'embouteillage, par exemple, on ne veut pas que les véhicules soient interprétés comme s'ils faisaient partie de l'arrière-plan. La figure 3.4 présente un problème d'un tout autre ordre : ce sont les petits objets qui ne sont pas détectés. La sensibilité de l'algorithme aux objets de différentes tailles est gérée par le paramètre  $R$  introduit plus tôt. Celui-ci devant être fixe, il est préférable que les objets devant être détectés occupent une surface similaire dans l'image. Or, la qualité de la segmentation ne devrait pas dépendre de la taille des véhicules.

En plus de produire des résultats de soustraction de fond peu concluants, l'algorithme LBP-Satpathy ne se démarque pas par sa rapidité d'exécution. Le domaine d'application visé impose d'être en mesure d'effectuer le traitement des séquences vidéo en temps réel

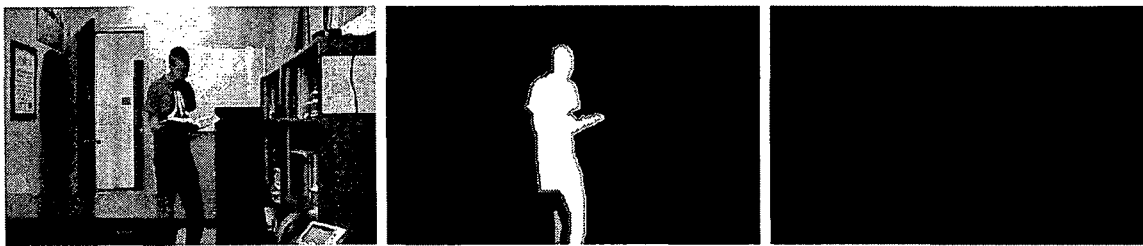


(a) Image originale

(b) Vérité terrain

(c) LBP-Satpathy

Figure 3.4 – Application de LBP-Satpathy ( $P = 6$ ,  $R = 3$  et  $K = 5$ ) sur la séquence *highway* de la série *Base* ( $t = 1419$ )

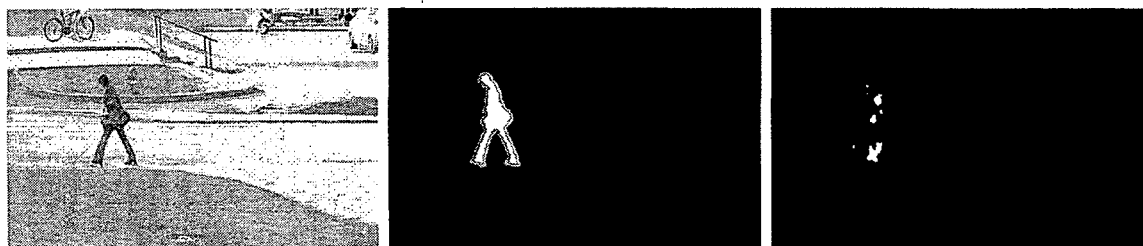


(a) Image originale

(b) Vérité terrain

(c) LBP-Satpathy

Figure 3.5 – Application de LBP-Satpathy ( $P = 6$ ,  $R = 3$  et  $K = 5$ ) sur la séquence *office* de la série *Base* ( $t = 1400$ )



(a) Image originale

(b) Vérité terrain

(c) LBP-Satpathy

Figure 3.6 – Application de LBP-Satpathy ( $P = 6$ ,  $R = 3$  et  $K = 5$ ) sur la séquence *pedestrians* de la série *Base* ( $t = 1000$ )

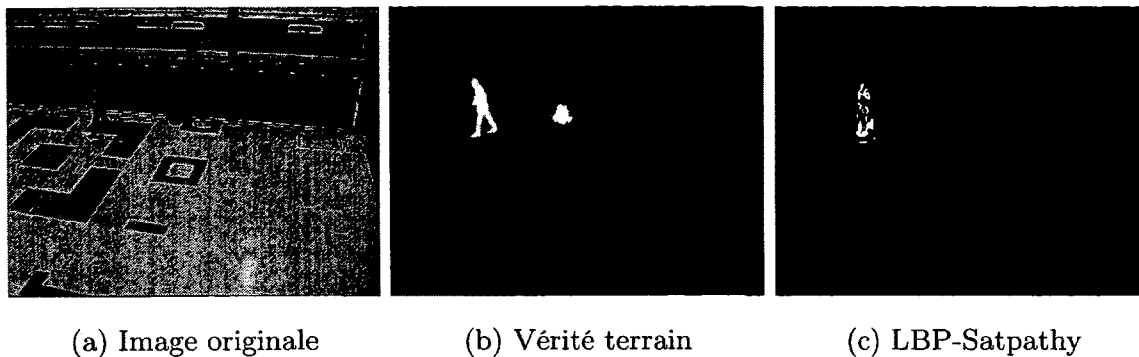


Figure 3.7 – Application de LBP-Satpathy ( $P = 6$ ,  $R = 3$  et  $K = 5$ ) sur la séquence *PETS2006* de la série *Base* ( $t = 1150$ )

(15 images par seconde, environ). Or, la méthode LBP-Satpathy présente une vitesse d'exécution avoisinant les 5 images par seconde pour une séquence de résolution  $320 \times 240$  avec une implantation en C++ sur un Core i7. Ainsi, malgré des caractéristiques intéressantes (utilisation de l'information de texture, invariance aux conditions d'éclairage), cette technique de soustraction de fond ne présente pas des performances suffisantes, tant en ce qui concerne la qualité des résultats que la vitesse d'exécution, pour être intégrée à notre méthode d'estimation de la vitesse du trafic routier à partir de l'information visuelle. Au lieu de cela, on se tourne plutôt vers une approche plus conventionnelle de la soustraction de fond, soit la modélisation statistique de l'arrière-plan basée sur un mélange de distributions gaussiennes.

### Détection des objets basée sur un mélange de gaussiennes

Une des difficultés intrinsèques à la méthode LBP-Satpathy est de trouver un ensemble de paramètres permettant d'obtenir des résultats de classification satisfaisants. Ces derniers étant très nombreux, il devient difficile de savoir lequel ajuster pour améliorer les performances de classification. Ce problème ne se pose pas avec la technique GMM-Lee de soustraction de fond basée sur la modélisation de l'arrière-plan par un mé-



lange de distributions gaussiennes. Il n'y a, pour cette approche, que deux paramètres à ajuster : le nombre  $K$  de distributions gaussiennes dans le modèle et le terme  $\alpha$  qui est utilisé dans le calcul du coefficient d'apprentissage du modèle statistique.

Comme c'est le cas pour la méthode LBP-Satpathy, le coefficient  $\alpha \in [0, 1]$  détermine à quelle vitesse l'algorithme s'adapte aux observations récentes. Un  $\alpha$  plus faible aura comme conséquence qu'un objet s'étant immobilisé prendra plus de temps à être considéré comme faisant partie de l'arrière-plan (figure 3.8). Ce comportement est souhaitable afin de bien gérer les situations de trafic lent. On ne veut pas, comme c'était le cas avec LBP-Satpathy, que les objets en déplacement soient trop rapidement intégrés à l'arrière-plan (figure 3.7).

À l'opposé, un  $\alpha$  trop petit rendra la détection des objets en mouvement trop sensible, et le modèle statistique de l'arrière-plan sera incapable de s'adapter aux changements permanents dans la scène. On peut observer un tel comportement, avec  $\alpha = 0.001$ , aux figures 3.9 et 3.10, où un changement dans l'illumination de la scène est interprété comme du mouvement. Une plus grande valeur de  $\alpha$  permet au modèle de s'adapter plus rapidement aux nouvelles conditions d'éclairage et ainsi de fournir une meilleure détection des objets. Les figures 3.11 et 3.12 présentent les résultats de classification pour  $\alpha = 0.1$ . Pour l'ensemble des tests, nous avons opté pour un  $\alpha$  de 0.001, soit la valeur suggérée par Stauffer et Grimson[47] et par Lee [33]. Malgré une sensibilité accrue aux changements d'illumination brusques, c'est cette valeur qui permettait d'obtenir les meilleurs résultats dans un contexte d'application sur des séquences vidéo de trafic routier.

Deuxièmement, le paramètre  $K$  détermine le nombre de distributions gaussiennes à inclure dans le modèle statistique des intensités des pixels. Si un pixel donné ne représentait qu'un type de surface sous des conditions d'éclairage constantes, une seule distribution gaussienne serait suffisante pour modéliser ses intensités sur la durée d'une séquence vidéo. Comme la couleur perçue de la surface filmée ne changerait pas, la dis-

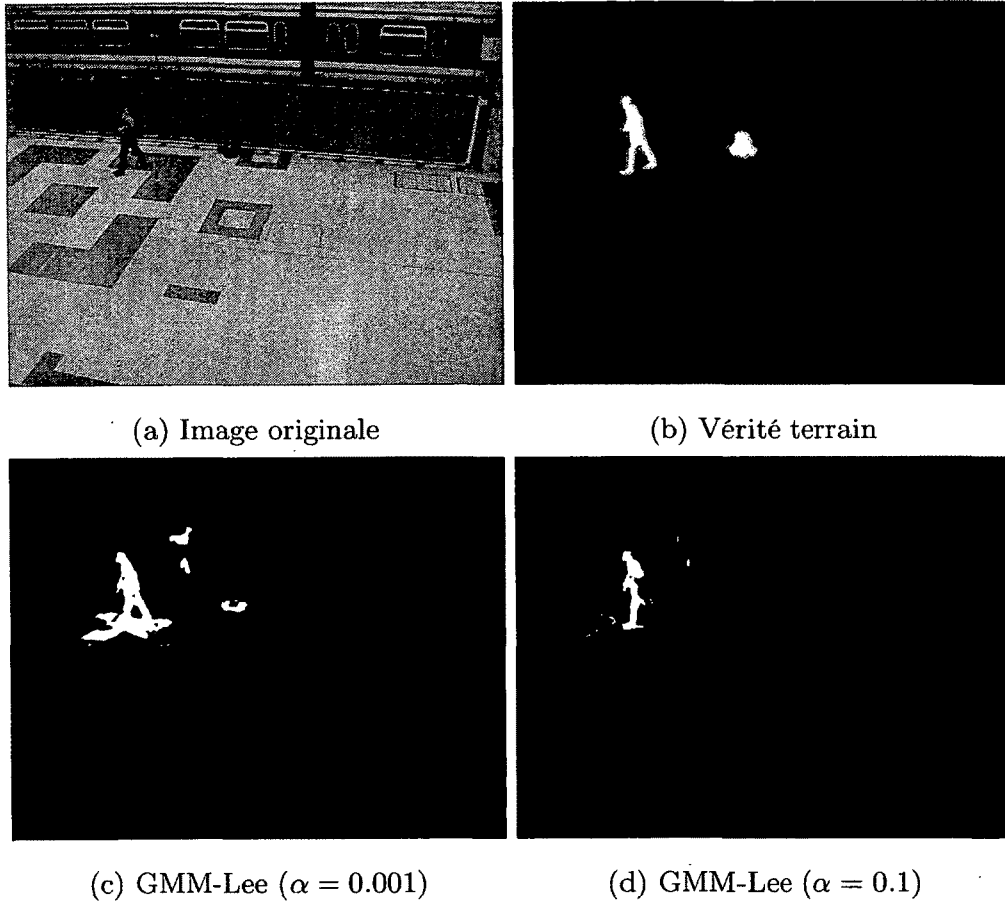
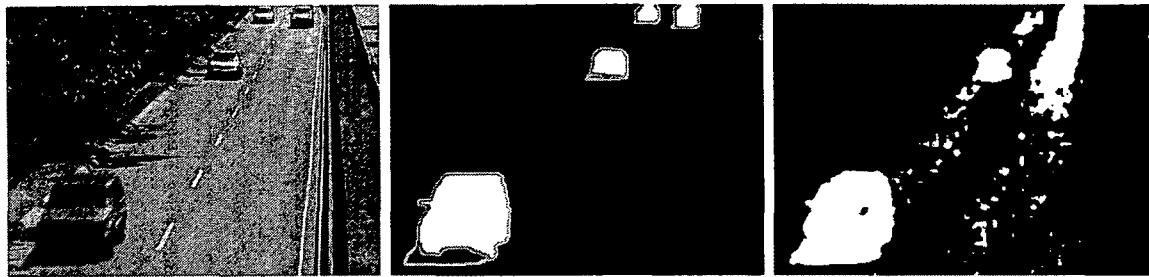


Figure 3.8 – Application de GMM-Lee pour différentes valeurs de  $\alpha$  sur la séquence *PETS2006* de la série *Base* ( $t = 1150$ )



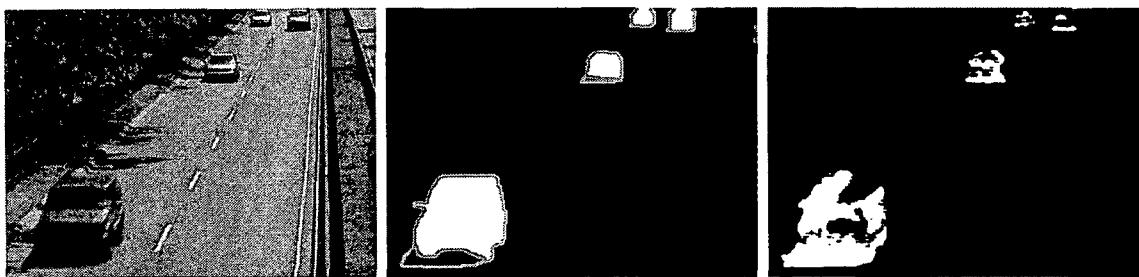
(a) Image originale      (b) Vérité terrain      (c) GMM-Lee ( $\alpha = 0.001$ )

Figure 3.9 – Impact d’un changement brusque d’illumination sur GMM-Lee (séquence *highway*,  $t = 930$ )



(a) Image originale      (b) Vérité terrain      (c) GMM-Lee ( $\alpha = 0.001$ )

Figure 3.10 – Impact d’un changement brusque d’illumination sur GMM-Lee (séquence *office*,  $t = 662$ )



(a) Image originale      (b) Vérité terrain      (c) GMM-Lee ( $\alpha = 0.1$ )

Figure 3.11 – Impact d’un changement brusque d’illumination sur GMM-Lee (séquence *highway*,  $t = 930$ )

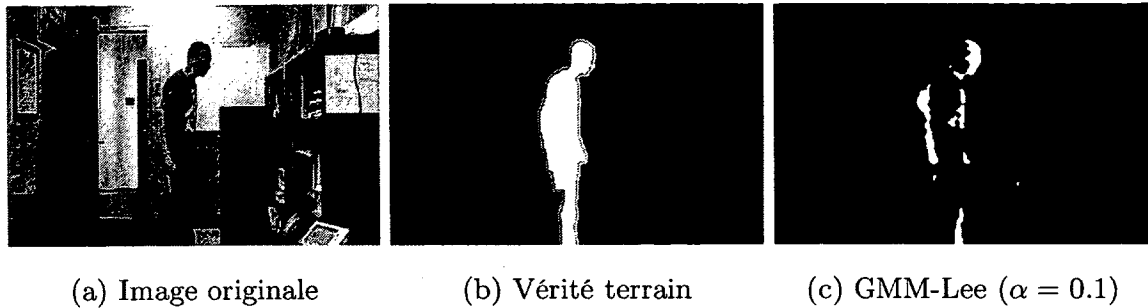


Figure 3.12 – Impact d’un changement brusque d’illumination sur GMM-Lee (séquence *office*,  $t = 662$ )

tribution gaussienne permettrait de modéliser le bruit inhérent à l’acquisition des images par un capteur. De même, si seules les conditions d’illumination de la scène variaient, une unique distribution gaussienne adaptative par pixel serait suffisante pour modéliser les intensités. Or, en pratique, plusieurs surfaces sous des conditions d’éclairage changeantes apparaissent dans la région de la scène correspondant à un pixel donné. Ainsi, plusieurs distributions gaussiennes sont nécessaires pour modéliser adéquatement les intensités d’un pixel sur la durée d’une séquence vidéo [47].

La question est donc de savoir combien de distributions utiliser. Friedman et Russell [17] apportent une réponse à cette question en séparant les intensités que peut prendre un pixel en trois catégories : l’arrière-plan, les objets en mouvement et l’ombre associée à un objet en mouvement. La suggestion est donc de prendre  $K = 3$ , ce qui concorde avec les recommandations de Lee [33] qui conseillent des valeurs de  $K$  entre 3 et 5. Nos

Tableau 3.3 – Influence du nombre de distributions gaussiennes sur les performances de GMM-Lee sur la série *Base* [20]

	Re	Pr	F <sub>1</sub>	Sp	FPR	FNR	PWC
GMM-Lee ( $K = 2$ )	<b>0.8982</b>	<b>0.6705</b>	<b>0.7628</b>	0.9764	0.0235	0.1017	2.7851
GMM-Lee ( $K = 3$ )	0.8195	0.5986	0.6874	<b>0.9783</b>	<b>0.0217</b>	0.1805	3.0128

(meilleur résultat)

expérimentations ont toutefois montré de meilleurs résultats de classification en utilisant  $K = 2$ , tel que l'illustre le tableau 3.3 pour la série *Base*. Non seulement les résultats sont-ils plus près de la vérité terrain, mais le fait d'utiliser une distribution gaussienne de moins accélère de beaucoup le traitement des séquences vidéo.

Les tableaux 3.2 (page 62), 3.4, 3.5 et 3.6 montrent les résultats de la méthode GMM-Lee sur un ensemble de séquences vidéo de la banque d'évaluation *changedetection.net* [20]. En comparant les résultats avec les autres méthodes basées sur les mélanges de gaussiennes, on constate que les performances de l'algorithme GMM-Lee sont du même ordre que celles des autres algorithmes. Cependant, il importe de souligner que, considérant le domaine d'application visé, toutes les mesures de performances n'ont pas la même importance.

En effet, dans un contexte où la soustraction de fond n'est pas une fin en soi, mais plutôt une étape intermédiaire visant à diminuer la quantité de données à traiter, certaines mesures de performance d'un classificateur présentent plus d'intérêt que d'autres. C'est par exemple le cas du rappel (Re) et du taux de faux négatifs (FNR). Rappelons que le rappel donne la probabilité qu'un pixel réellement en mouvement soit bien classé et le taux de faux négatif donne la proportion de mauvaise classification pour les pixels réellement en mouvement. Ces deux mesures sont cruciales pour nous, puisqu'il est beaucoup plus grave de mal classer un pixel réellement en mouvement que le contraire. Une erreur de ce type (un faux négatif, voir le tableau 3.1 à la page 53) entraîne qu'aucune évaluation du déplacement de l'objet ne sera faite en ce point, diminuant du même coup la précision de l'évaluation de la vitesse du trafic.

Il est donc préférable d'avoir une méthode qui présente un fort rappel et un faible taux de faux négatif, quitte à moins bien performer pour les autres mesures. C'est précisément dans cette situation que se retrouve la méthode GMM-Lee avec l'ensemble de paramètres choisis, comme en font foi les résultats sur les quatre catégories de séquences

Tableau 3.4 – Performances moyennes ( $K = 2$ ) pour la série *Arrière-plan dynamique* [20]

	Re	Pr	F <sub>1</sub>	Sp	FPR	FNR	PWC
GMM-Lee	0.7342	0.5957	0.5523	0.9875	0.0125	0.2658	1.6365
GMM-Stauffer [47]	<b>0.8344</b>	0.5989	0.6330	0.9896	0.0104	<b>0.1656</b>	1.2083
GMM-KaewTraKulPong [29]	0.6303	<b>0.7700</b>	<b>0.6697</b>	<b>0.9983</b>	<b>0.0017</b>	0.3697	<b>0.5405</b>
GMM-RECTGAUSS- <i>Tex</i> [42]	0.4776	0.6478	0.4296	0.9838	0.0162	0.5224	1.9735
GMM-Zivkovic [57]	0.8019	0.6213	0.6328	0.9903	0.0097	0.1981	1.1725

(meilleur résultat, pire résultat)

de la banque de test. Le fait de mal classifier un pixel réellement statique a seulement comme conséquence que le flux optique sera évalué pour un pixel qui n'est pas en mouvement, ce qui n'influence en rien la qualité des résultats, mais impacte seulement le temps d'exécution de l'algorithme. En effet, la vitesse des véhicules ne sera estimée qu'à partir des vecteurs présentant une amplitude supérieure à un certain seuil.

Les figures 3.13 à 3.15 présentent les résultats du classificateur GMM-Lee sur des séquences de la banque d'évaluation *changedetection.net* [20]. Les images de la série *Arrière-plan dynamique* (figure 3.13) posent un certain problème à l'algorithme qui, logiquement, détecte les mouvements de l'arrière-plan comme faisant partie des régions d'intérêt, d'où les résultats du tableau 3.4 qui sont en baisse par rapport à ceux du tableau 3.2. Dans cette situation, le fait d'augmenter le nombre de distributions gaussiennes dans le modèle statistique des pixels (paramètre  $K$ ) n'a pas d'impact réel sur les résultats, comme l'illustre les images 3.13c et 3.13d.

Le même problème se produit avec les images de la série *Vibration de la caméra* (figure 3.14), dont les résultats sont présentés au tableau 3.5. La différence est qu'ici, le passage de  $K = 2$  à  $K = 3$  permet d'améliorer légèrement la qualité de la classification. La raison est que, à cause des vibrations de la caméra, plusieurs intensités différentes doivent être associées à l'arrière-plan pour un même pixel. Ainsi, l'utilisation d'un nombre plus élevé de distributions gaussiennes permet de mieux modéliser ce cas de figure.

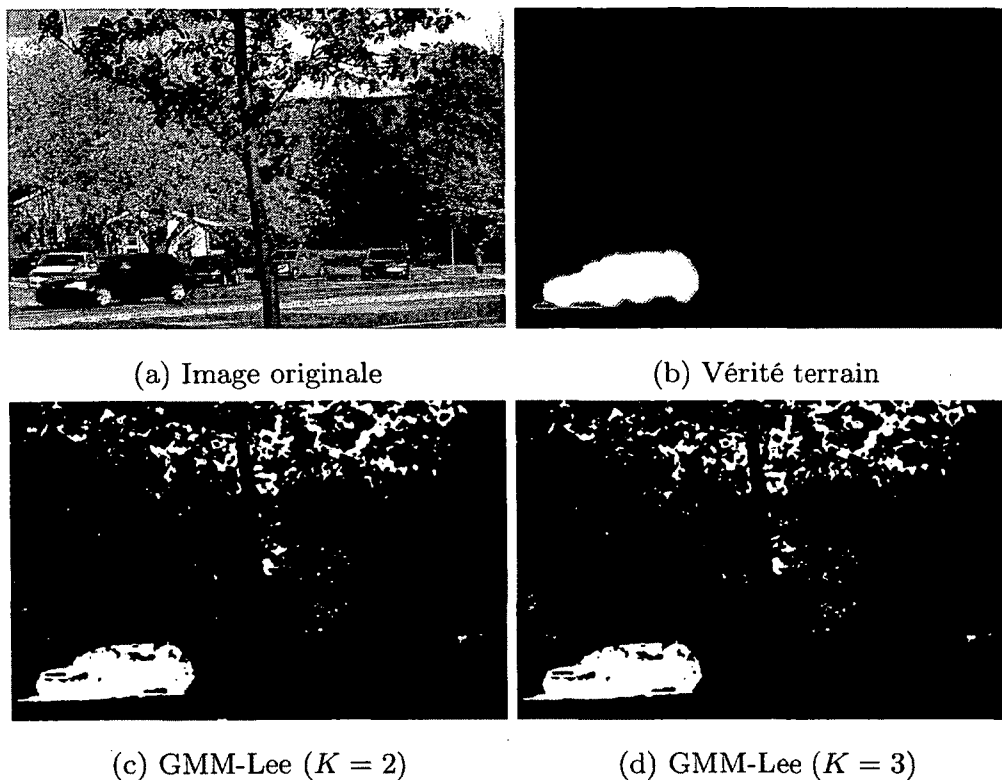


Figure 3.13 – Application de GMM-Lee sur la séquence *fall* de la série *Arrière-plan dynamique* ( $t = 2305$ )

Tableau 3.5 – Performances moyennes ( $K = 2$ ) pour la série *Vibration de caméra* [20]

	Re	Pr	$F_1$	Sp	FPR	FNR	PWC
GMM-Lee	<b>0.7684</b>	0.4811	0.5767	0.9549	0.0451	<b>0.2316</b>	5.2688
GMM-Stauffer [47]	0.7334	0.5126	<b>0.5969</b>	0.9666	0.0334	0.2666	4.2269
GMM-KaewTraKulPong [29]	0.5074	<b>0.6897</b>	0.5761	<b>0.9888</b>	<b>0.0112</b>	0.4926	<b>3.0233</b>
GMM-RECTGAUSS-TeX [42]	0.7649	0.4179	0.5370	0.9497	0.0503	0.2351	5.6663
GMM-Zivkovic [57]	0.6900	0.4872	0.5670	0.9665	0.0335	0.3100	4.4057

(meilleur résultat, pire résultat)

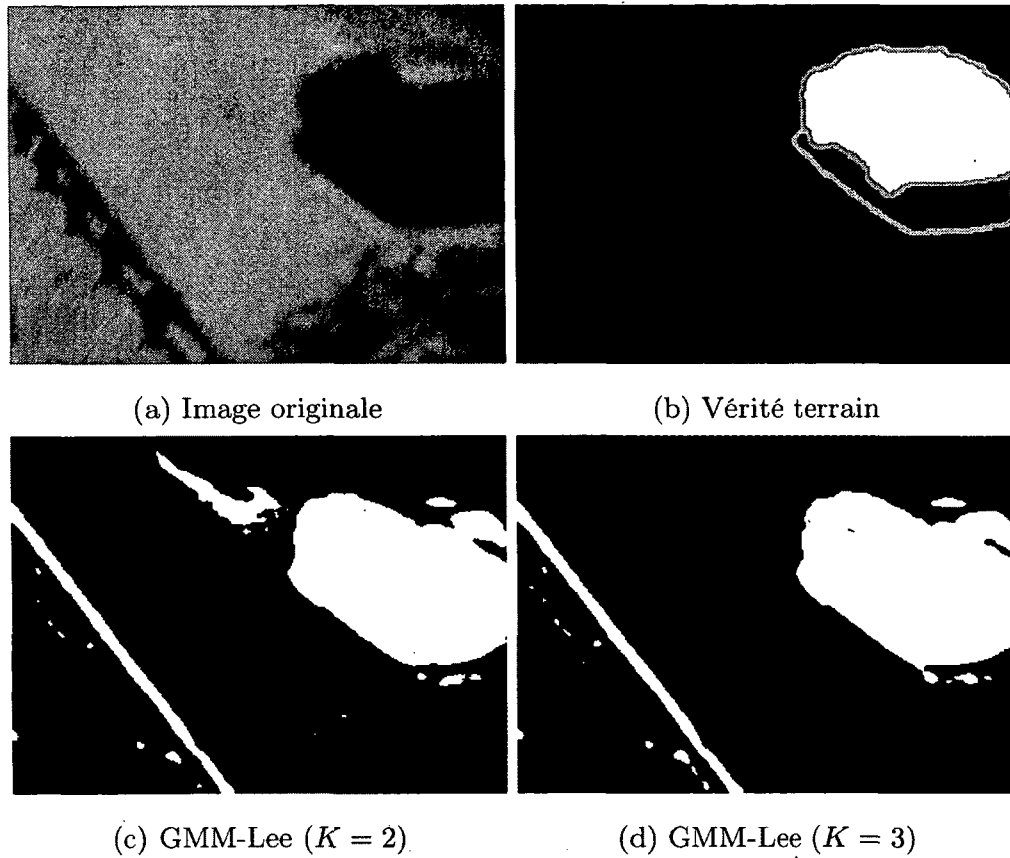


Figure 3.14 – Application de GMM-Lee sur la séquence *traffic* de la série *Vibration de la caméra* ( $t = 1451$ )



Tableau 3.6 – Performances moyennes ( $K = 2$ ) pour la série *Mouvement intermittent* [20]

	Re	Pr	F <sub>1</sub>	Sp	FPR	FNR	PWC
GMM-Lee	<b>0.6180</b>	0.4089	0.4504	0.9030	0.0970	<b>0.3820</b>	11.463
GMM-Stauffer [47]	0.5142	0.6688	0.5207	0.9835	0.0165	0.4858	<b>5.1955</b>
GMM-KaewTraKulPong [29]	0.3476	<b>0.6953</b>	0.3903	0.9892	0.0108	0.6524	5.9854
GMM-RECTGAUSS-TeX [42]	0.2190	0.5850	0.3146	<b>0.9977</b>	<b>0.0023</b>	0.7810	5.2547
GMM-Zivkovic [57]	0.5467	0.6458	<b>0.5325</b>	0.9712	0.0288	0.4533	5.4986

(meilleur résultat, pire résultat)

Enfin, la catégorie *Mouvement intermittent* présente des objets dont le mouvement est interrompu en cours de séquence. On met ainsi en évidence la rapidité d'apprentissage du modèle statistique. Les performances de la méthode GMM-Lee sont détaillées au tableau 3.6. Tel que mentionné précédemment et illustré à la figure 3.8, ce comportement est contrôlé par la variable  $\alpha$ . Avec un coefficient  $\alpha$  fixé à 0.001 et une vitesse de défilement supposée de 15 images par seconde, un laps de temps d'environ 30 secondes est nécessaire pour qu'un objet immobile soit considéré comme faisant partie de l'arrière-plan à cause de son immobilité, tel qu'illustré à la figure 3.15. Un grand laps de temps avant que l'arrière-plan assimile un objet immobile est souhaitable dans une situation d'analyse du trafic routier afin de bien gérer les moments où le flot de voitures est immobile, à un feu de circulation, par exemple.

### 3.2 Estimation du mouvement par une approche multi-résolution du flux optique

Dans cette section, nous nous intéressons à l'évaluation et à l'analyse des données de flux optique. Tel que mentionné à la section 2.2.3, notre approche vise à utiliser l'informations visuelle à trois niveaux de détails différents (trois résolutions : R1, R2 et

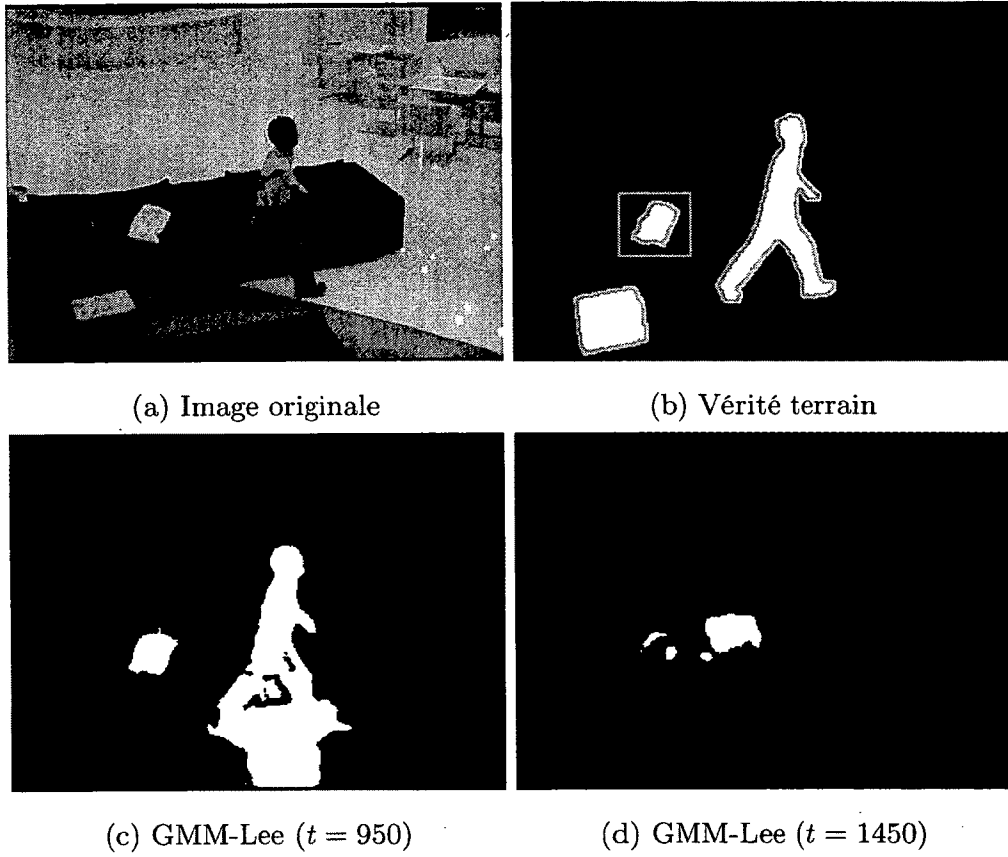


Figure 3.15 – Application de GMM-Lee sur la séquence *sofa* de la série *Mouvement intermittent*

R4, telles que présentées à la section 2.2.3) pour obtenir diverses mesures allant de la vitesse moyenne dans la scène à la vitesse instantanée de chaque véhicule.

Nous présentons ici les résultats associés aux résolutions R1, R2 et R4. Celles-ci correspondent respectivement à la résolution originale et à des décimations d'ordre 2 et 4 de celle-ci. Par exemple, pour une séquence vidéo dont les images ont une résolution de  $320 \times 240$ , la résolution R1 conservera la totalité des données, la résolution R2 sera de  $160 \times 120$  et la résolution R4 sera de  $80 \times 60$ . Nous procédons en ordre croissant du niveau de détail, c'est-à-dire en abordant d'abord les résultats de la résolution R4, puis ceux de la résolution R2 et finalement ceux de la résolution R1.

### 3.2.1 Analyse du flux optique en résolution R4

L'objectif de l'analyse du flux optique en résolution R4 (correspondant au quart de la résolution originale) est double : nous devons d'abord évaluer un indice de fluidité global pour l'ensemble de la scène, puis déterminer automatiquement le nombre de directions principales, c'est-à-dire le nombre de directions dans lesquelles se déplacent les véhicules, et ce, à chaque instant.

Comme il a été mentionné à la section 2.2.3, l'indice de fluidité correspond à la vitesse moyenne  $v_s$  des véhicules. Dans l'équation 2.3, le seul paramètre à fixer est celui du seuil  $S_A$  en deçà duquel un vecteur de flux optique  $\mathbf{v}_{ij}$  associé à un pixel  $(x_i, y_j)$  n'est pas considéré dans le calcul de la vitesse moyenne. Nous avons constaté qu'un  $S_A$  situé entre 20% et 40% de l'amplitude maximale de tous les vecteurs du flux optique permet de bien cerner les régions présentant un réel mouvement, tel qu'illustré aux figures 3.16 et 3.17. Les figures 3.16c et 3.17c illustrent les masques appliqués sur les données correspondant au seuil  $S_A$  choisi. Dans les deux cas, un seuil  $S_A$  de 40% a été utilisé, éliminant ainsi du calcul de vitesse moyenne une bonne partie des régions ne présentant pas de mouvement

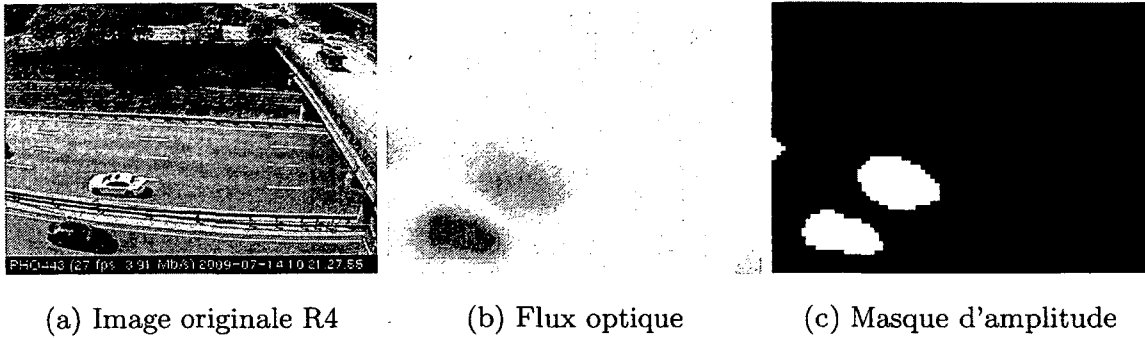


Figure 3.16 – Évaluation de la vitesse moyenne sur la séquence *streetLight* ( $t = 887$ ) de la série *Mouvement intermittent*

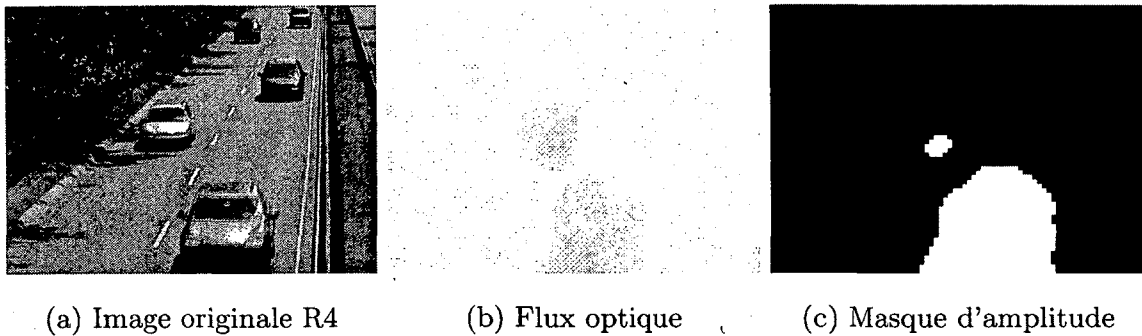


Figure 3.17 – Évaluation de la vitesse moyenne sur la séquence *highway* ( $t = 1070$ ) de la série *Base*

réel, mais aussi celles où l'on retrouve des déplacements de plus faible amplitude, comme dans le coin supérieur droit de la figure 3.16a, ou encore dans toute la portion supérieure de la figure 3.17a. Ce dernier résultat peut toutefois être évité en effectuant d'abord une rectification des images, tel que détaillé à la section 1.2.1.

Lorsque la vitesse moyenne  $v_s$  descend sous un certain seuil  $S_1$  fixé par l'utilisateur et variant selon la situation, l'analyse du déplacement des véhicules doit être raffinée. La résolution R4, correspondant au quart de la résolution originale, ne suffit alors plus pour obtenir des informations plus précises. Il faut plus de données. L'algorithme passe alors

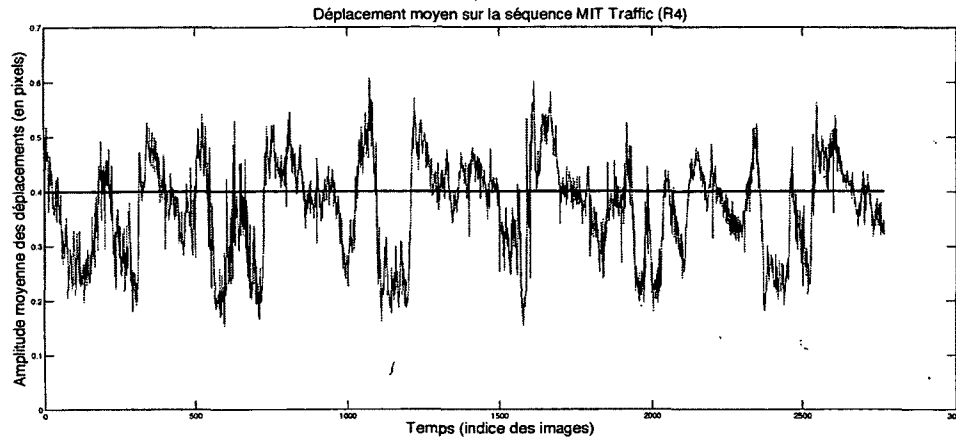


Figure 3.18 – Déplacement moyen pour la séquence *MIT Traffic* (R4) [53]

automatiquement à la résolution R2. La figure 3.18 illustre le déplacement moyen (en pixels) évalué à chaque image de la séquence *MIT Traffic* [53]. Le seuil  $S_1$  y est illustré par une droite horizontale. Ainsi, lorsque le mouvement moyen passe sous  $S_1$ , le système change de résolution afin de raffiner ses calculs. Un principe similaire est appliqué en résolution R2 pour revenir à la résolution R1 lorsque le trafic redevient fluide.

Dans l'état actuel, l'estimation du déplacement se fait en pixel et non en unité de distance réelle. Ce faisant, le choix de  $S_1 = 0.4$  sur la figure 3.18 est arbitraire et n'a pour objectif que d'illustrer à quel moment le système passera en résolution R2. Dans une situation réelle, le choix du seuil  $S_1$  aura d'importants impacts sur le comportement du système. Le choix d'un  $S_1$  trop élevé entraînera des passages fréquents aux résolutions supérieures. La performance du système en sera donc affectée. À l'opposé, un  $S_1$  trop bas rendra le système trop tolérant aux ralentissements du trafic et fournira ainsi moins d'information. Il importe donc de porter une attention particulière au choix de  $S_1$ , et ce pour chaque point de vue ou chaque zone observée par les caméras. Cette observation est, bien entendu, aussi valide pour le choix de  $S_2$ . De plus, comme les déplacements

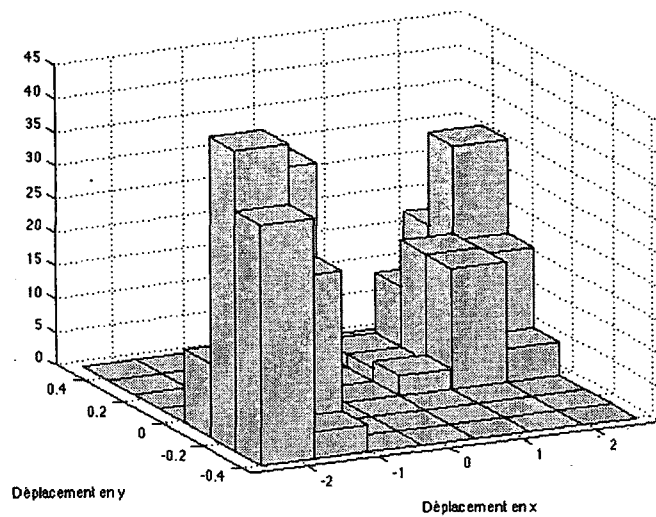


Figure 3.19 – Histogramme de mouvement de la figure 3.16b

obtenus sont en pixels, il est nécessaire d'adapter les choix de  $S_1$  et  $S_2$  en fonction de la résolution où l'on se trouve. En effet, comme les images de résolution R1 correspondent au quart de la résolution originale, il en sera de même pour les déplacements obtenus, d'où la nécessité de définir les seuils de déplacement en fonction de la résolution choisie. Lorsque des vitesses réelles seront évaluées, ce problème ne se posera plus.

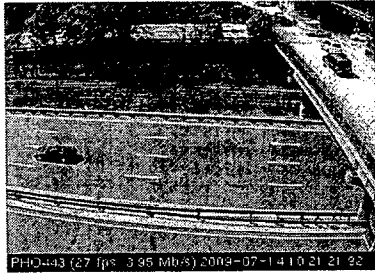
L'autre objectif à atteindre avec l'information visuelle de la résolution R4 est d'évaluer automatiquement et ce, à chaque image de la séquence, le nombre de directions où il y a du trafic routier en mouvement. Cette opération se fait en évaluant un histogramme de mouvement sur les vecteurs de flux optique ayant servi au calcul de la vitesse moyenne. La figure 3.19 présente l'historgramme de mouvement correspondant au flux optique de la figure 3.16b.

La figure 3.20 illustre le procédé de seuillage détaillé à la section 2.2.3 appliqué à d'autres images de la séquence *streetLight* de la série *Mouvement intermittent*. Pour

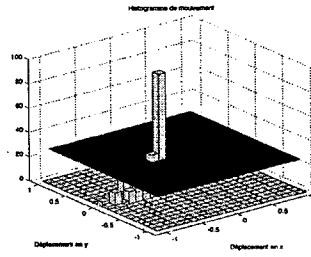
l'image 3.20a, nous obtenons un seul regroupement d'intervalles connexes ( $n_d = 1$ ) pour un seul véhicule en mouvement, ce qui est le résultat attendu. L'image 3.20g présente quant à elle deux véhicules en mouvement, mais qui vont cette fois-ci dans la même direction, comparativement à la scène de la figure 3.16a où les directions étaient opposées. Les deux véhicules ayant la même direction de déplacement sont regroupés dans la même région connexe suite à l'application du seuil sur l'histogramme ( $n_d = 1$ ). Le cas de la figure 3.20d est particulière, puisqu'aucun véhicule en mouvement n'est visible dans la scène. Ce faisant, l'algorithme évalue un histogramme et effectue un seuillage, mais aucune région ne dépasse le seuil. Il retourne alors simplement  $n_d = 0$ .

La figure 3.21 illustre l'évaluation du nombre de directions sur une séquence plus complexe tirée de [53]. On obtient ici  $n_d = 4$ . Attardons-nous maintenant à illustrer la correspondance entre les régions connexes dans l'histogramme et les directions présentes dans l'image. Pour arriver à cela, on isole chaque direction manuellement en appliquant un masque à l'image originale (figures 3.22a, 3.22b et 3.22c). En analysant le flux optique associé à cette portion de l'image en particulier, on trouve la région connexe de l'histogramme associée à chaque direction (figures 3.22d, 3.22e et 3.22f). On constate alors la correspondance entre les directions de trafic et les régions connexes de l'histogramme illustré à la figure 3.21c.

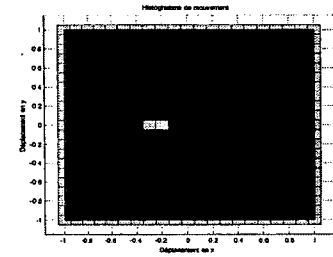
Le même procédé est appliqué à deux autres images de la séquence *MIT Traffic* [53] (figures 3.21, 3.22 et 3.23, 3.24). On constate que malgré la présence de déformations dues à la perspective, nous sommes en mesure de convenablement détecter le nombre de directions.



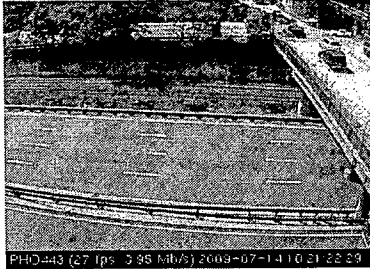
(a) Image originale ( $t = 735$ )



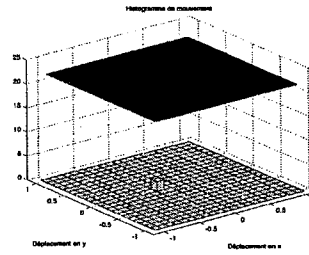
(b) Histogramme



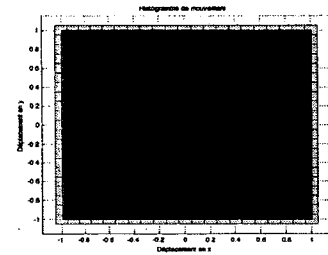
(c) Régions connexes



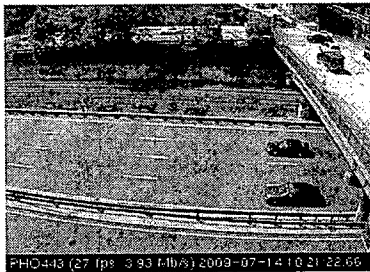
(d) Image originale ( $t = 745$ )



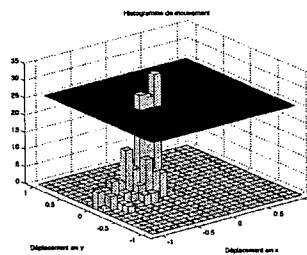
(e) Histogramme



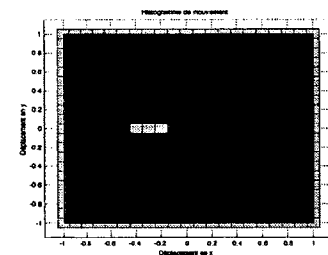
(f) Régions connexes



(g) Image originale ( $t = 755$ )



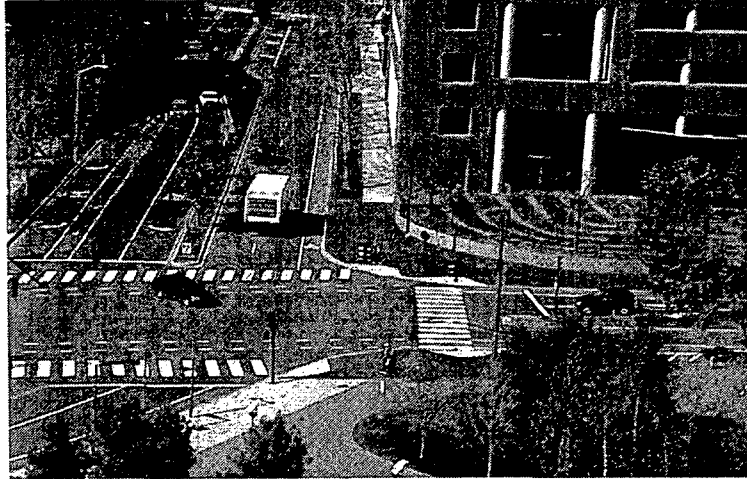
(h) Histogramme



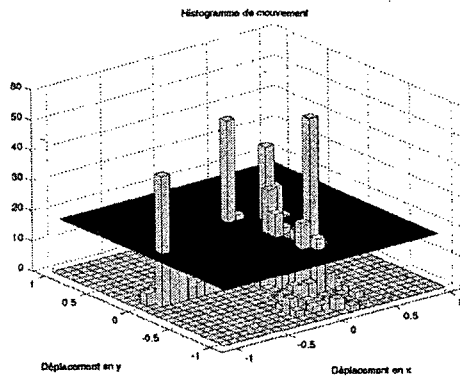
(i) Régions connexes

Figure 3.20 – Évaluation du nombre de directions de trafic sur la séquence *streetLight* de la série *Mouvement intermittent*

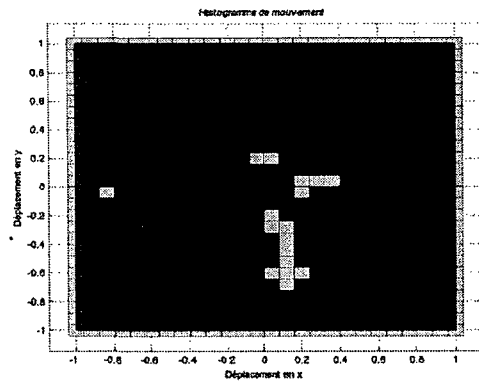




(a) Image originale



(b) Histogramme



(c) Régions connexes

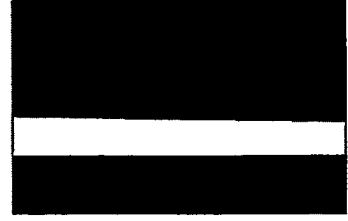
Figure 3.21 – Évaluation du nombre de directions de trafic sur la séquence *MIT Traffic* ( $t = 850$ ) tirée de [53].



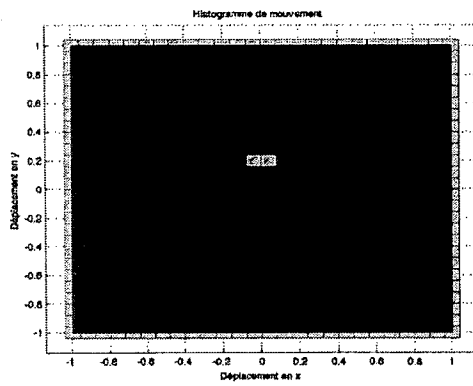
(a) Masque (direction 1)



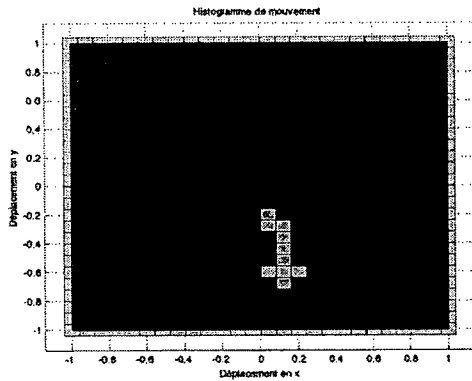
(b) Masque (direction 2)



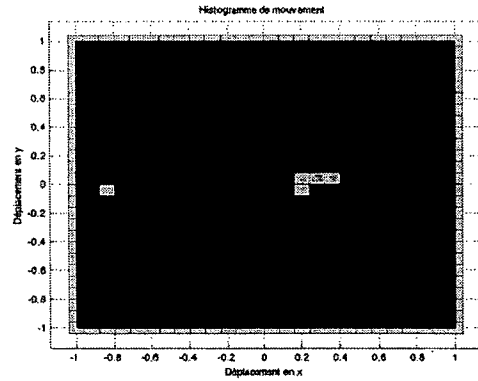
(c) Masque (directions 3 et 4)



(d) Région connexe (direction 1)

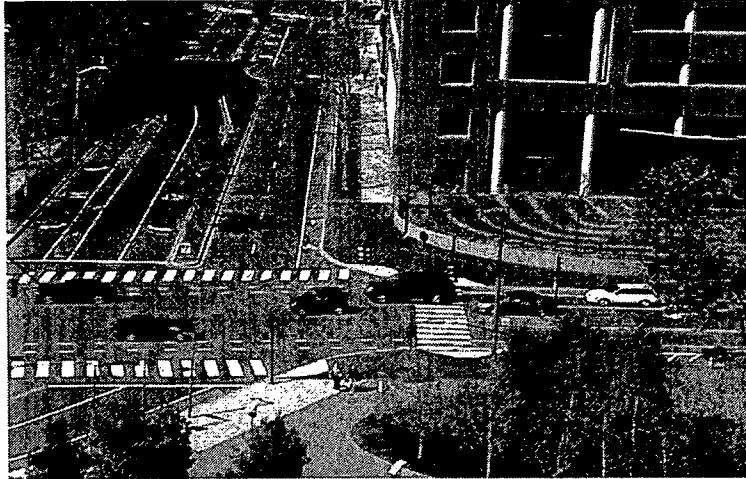


(e) Région connexe (direction 2)

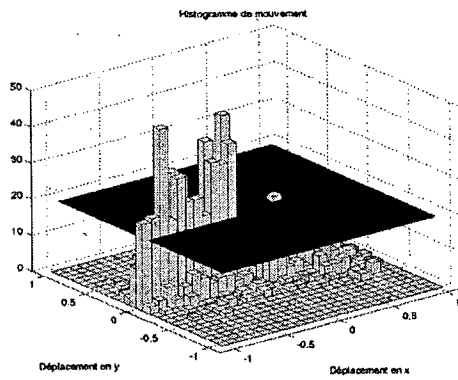


(f) Régions connexes (directions 3 et 4)

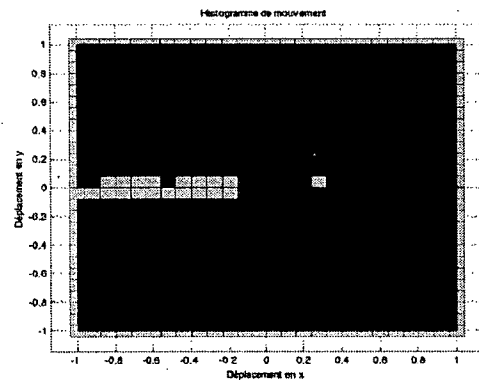
Figure 3.22 – Correspondance entre les directions de trafic et les régions connexes sur la séquence *MIT Traffic* ( $t = 850$ ) tirée de [53].



(a) Image originale



(b) Histogramme



(c) Régions connexes

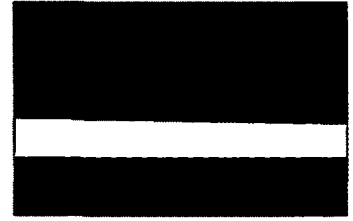
Figure 3.23 – Évaluation du nombre de directions de trafic sur la séquence *MIT Traffic* ( $t = 1245$ ) tirée de [53].



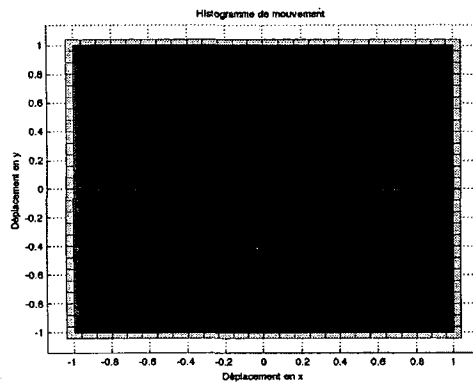
(a) Masque (direction 1)



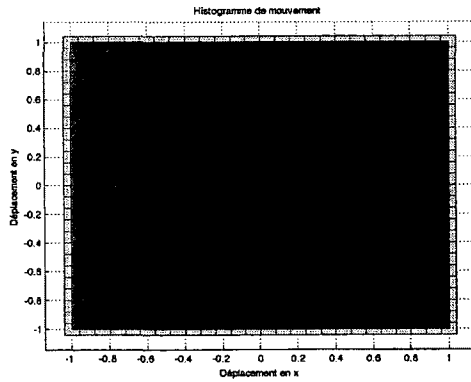
(b) Masque (direction 2)



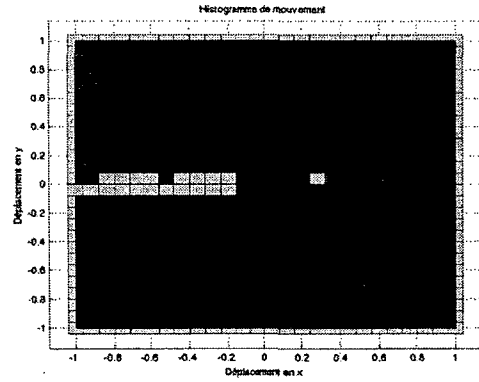
(c) Masque (directions 3 et 4)



(d) Région connexe (direction 1)

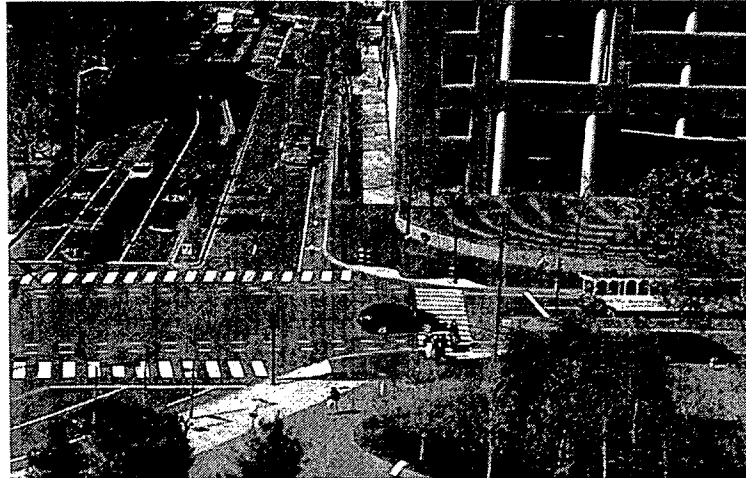


(e) Région connexe (direction 2)

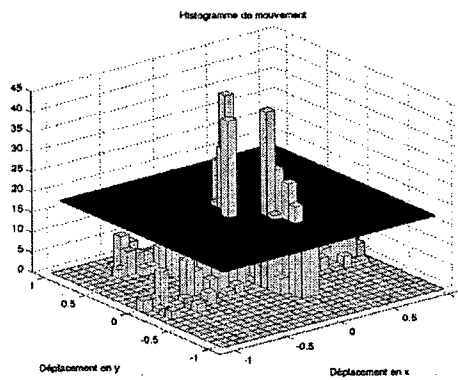


(f) Régions connexes (directions 3 et 4)

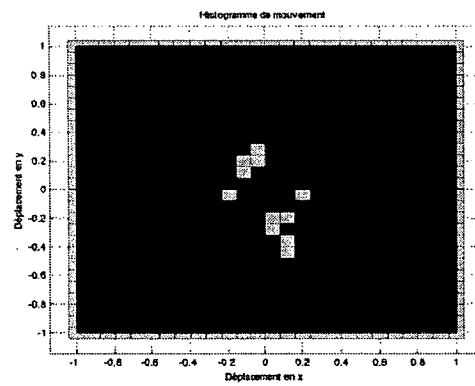
Figure 3.24 – Correspondance entre les directions de trafic et les régions connexes sur la séquence *MIT Traffic* ( $t = 1245$ ) tirée de [53].



(a) Image originale



(b) Histogramme



(c) Régions connexes

Figure 3.25 – Évaluation du nombre de directions de trafic sur la séquence *MIT Traffic* ( $t = 2705$ ) tirée de [53].



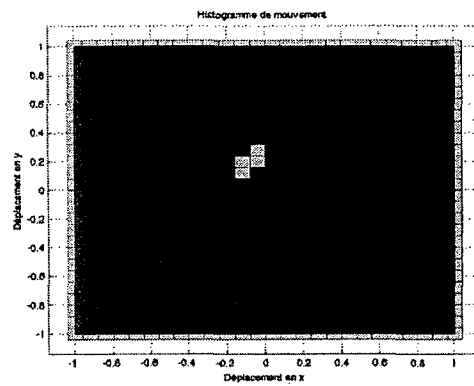
(a) Masque (direction 1)



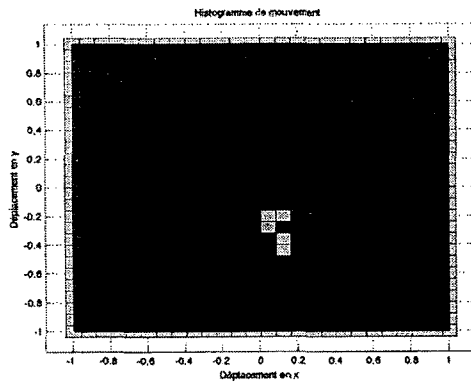
(b) Masque (direction 2)



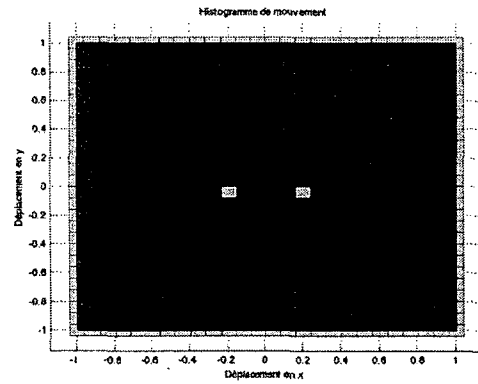
(c) Masque (directions 3 et 4)



(d) Région connexe (direction 1)



(e) Région connexe (direction 2)



(f) Régions connexes (directions 3 et 4)

Figure 3.26 – Correspondance entre les directions de trafic et les régions connexes sur la séquence *MIT Traffic* ( $t = 2705$ ) tirée de [53].

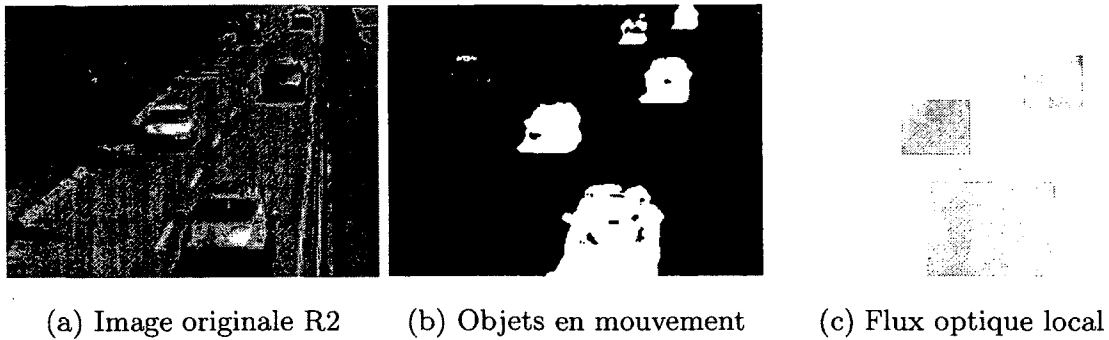


Figure 3.27 – Évaluation locale du flux optique sur la séquence *highway* ( $t = 1070$ ) de la série *Base*

### 3.2.2 Analyse du flux optique en résolution R2

Rappelons que l'objectif de l'analyse de l'information visuelle en résolution R2 est d'évaluer la vitesse moyenne pour chacune des  $n_d$  directions principales dans la scène. Pour y arriver, nous estimons le flux optique uniquement sur les rectangles englobants correspondant aux objets en mouvement (au lieu de la totalité de l'image) et dont l'aire est supérieure à 75 pixels (figure 3.27). L'implantation de l'algorithme utilisé pour le calcul du flux optique ne permet pas l'évaluation du déplacement sur de plus petites régions. Par la suite, on regroupe l'ensemble des vecteurs du flux optique en  $n_d$  classes à l'aide de l'algorithme des  $k$ -moyennes.

Dans un premier temps, nous vérifions que le fait d'évaluer localement (sur les régions en mouvement) le flux optique plutôt que globalement (sur l'image entière) n'affecte pas significativement les résultats. Pour y arriver, nous comparons le flux optique évalué uniquement sur les régions présentant du mouvement (flux optique local) avec celui obtenu sur l'image entière (flux optique global). Les figures 3.28 et 3.29 illustrent bien que le fait d'évaluer le flux optique sur un ensemble de régions de plus petites tailles dans l'image ne modifie pas la direction de celui-ci. Les orientations obtenues dans les figures 3.28b et

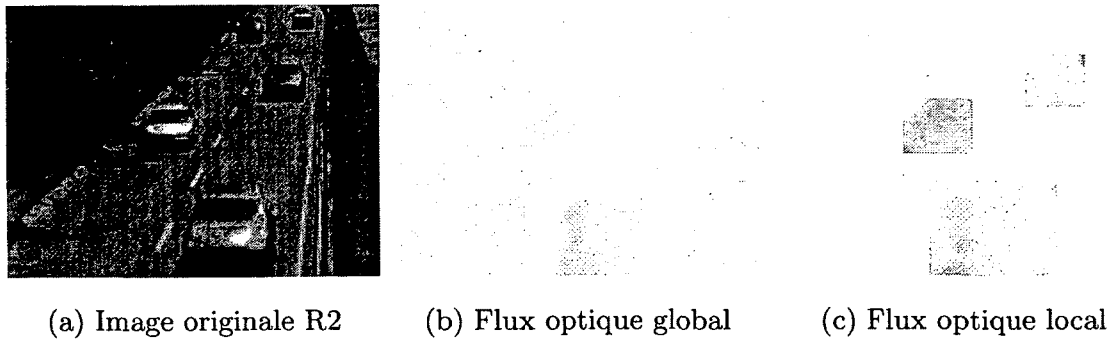


Figure 3.28 – Comparaison des flux optiques sur la séquence *highway* ( $t = 1070$ ) de la série *Base*

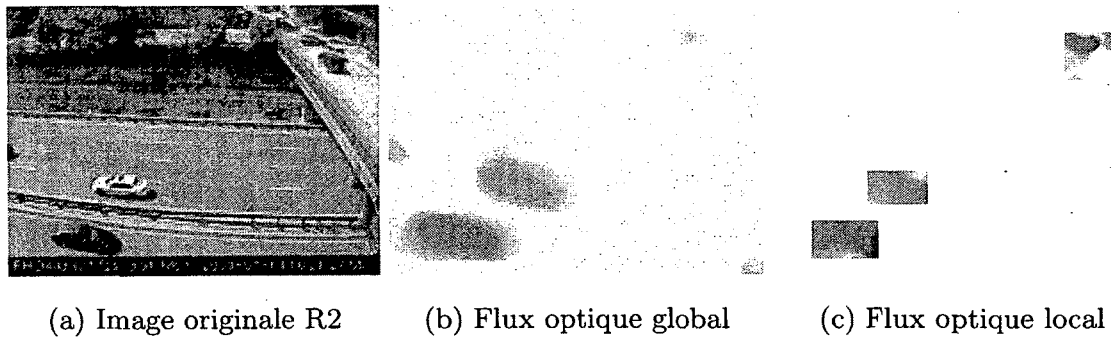


Figure 3.29 – Comparaison des flux optiques sur la séquence *streetLight* ( $t = 887$ ) de la série *Mouvement intermittent*

3.29b se retrouvent aux mêmes positions dans les figures 3.28c et 3.29c, respectivement.

Les graphiques des figures 3.30 et 3.31 illustrent les positions des vecteurs de flux optique dans le plan pour le flux optique global (en bleu) et le flux optique local (en rouge) des figures 3.28 et 3.29, respectivement. Les positions des deux familles de vecteurs (global et local) sont, dans les deux cas, très similaires, ce qui confirme que l'évaluation locale du flux optique n'affecte pas les résultats de manière importante. Dans les deux cas, la principale différence est l'absence des vecteurs de faible amplitude autour de l'origine pour l'évaluation locale. Cela est dû au fait qu'une majorité des pixels ne présentant aucun



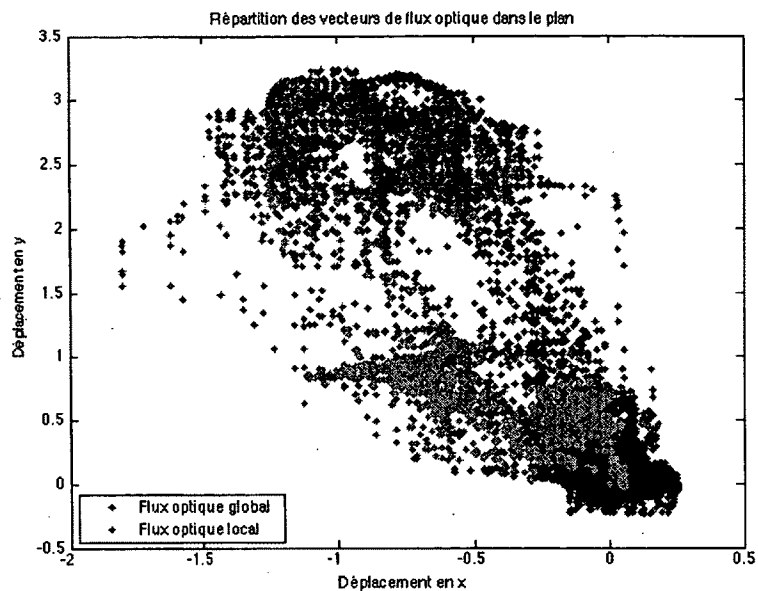


Figure 3.30 – Représentation dans le plan des vecteurs de flux optique de la figure 3.28

mouvement ont été éliminés du traitement par la soustraction de fond. Finalement, la différence dans les intensités des couleurs aux figures 3.28c et 3.29c n'est pas due à une variation de l'amplitude des vecteurs de flux optique, mais plutôt à une opération de mise à l'échelle des vecteurs dans la méthode de conversion en couleurs proposée par Baker *et al.* [3].

Ayant validé que la technique de calcul local du flux optique ne donne pas des résultats anormaux, nous pouvons procéder à l'évaluation des vitesses moyennes par direction. À la section 3.2.1 (figure 2.7), nous avons trouvé  $n_d = 2$  pour les vecteurs de flux optique de l'image 3.29a. Ainsi, en segmentant ces vecteurs en 2 classes, nous sommes en mesure d'obtenir les vecteurs centroïdes correspondant aux déplacements moyens pour chacune des deux directions, tel qu'illustré à la figure 3.32.

La résolution R2 doit aussi elle-même permettre l'évaluation du nombre de directions

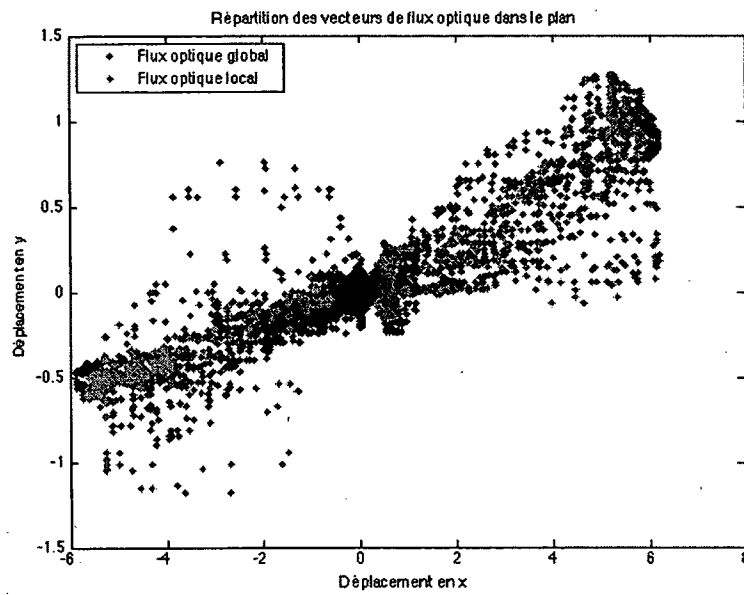


Figure 3.31 – Représentation dans le plan des vecteurs de flux optique de la figure 3.29

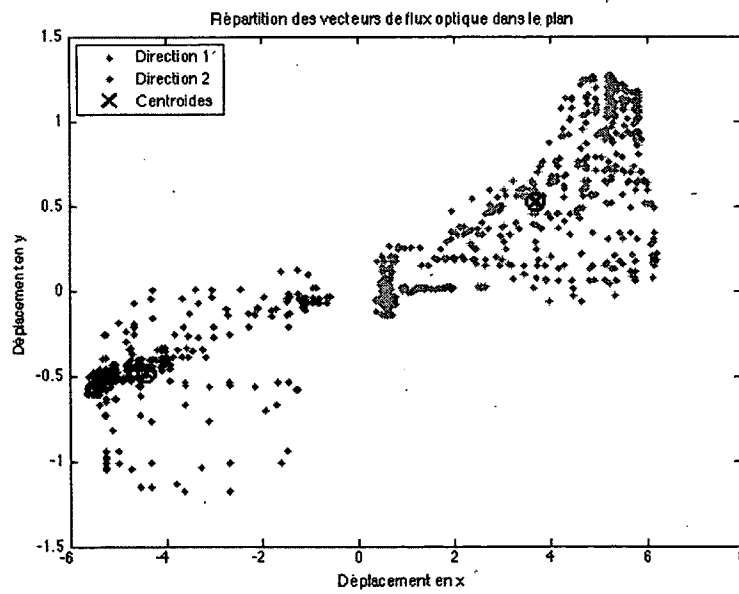


Figure 3.32 – Évaluation du déplacement moyen par direction

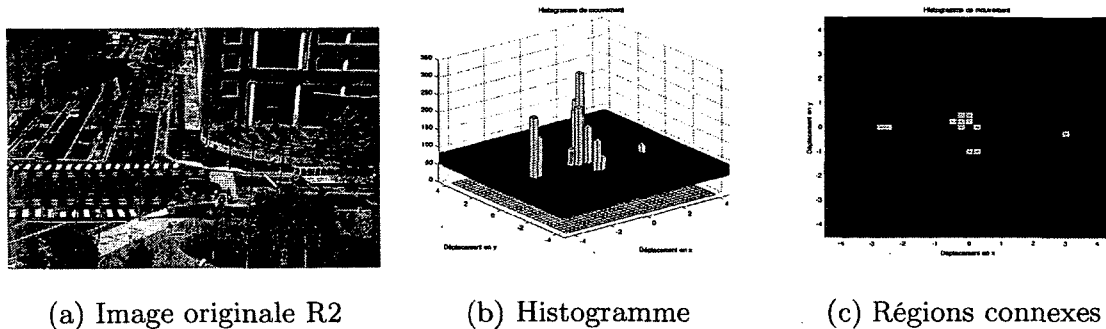


Figure 3.33 – Évaluation du nombre de directions de trafic sur la séquence *MIT Traffic* ( $t = 2705$ ) tirée de [53].

afin de pouvoir poursuivre la segmentation dans le cas probable où l’analyse demeurerait à la résolution R2. Pour y arriver, on procède de manière identique à ce qui a été présenté à la section 2.2.3 et illustré à la section 3.2.1. La construction de l’histogramme de mouvement se fait cette fois-ci à l’aide des vecteurs de flux optique évalués sur les régions en mouvement, et le nombre de régions connexes détermine de nouveau le nombre de directions de trafic. La figure 3.33 illustre ce processus sur une image de la séquence *MIT Traffic*. On obtient alors  $n_d = 4$ , ce qui correspond au résultat obtenu précédemment, à la figure 3.25.

Finalement, il est de nouveau important de mentionner le processus de changement de résolution. À la différence de ce qui a été présenté à la section précédente, le changement de résolution peut ici se faire non pas dans un, mais dans deux sens. En effet, si le trafic ralentit, le système doit passer en résolution R1 et, au contraire, revenir à la résolution R4 si la vitesse moyenne dans la scène augmente. Ces transitions surviennent si la vitesse moyenne évaluée avec les données de la résolution R2 descend sous un seuil  $S_2$  ou remonte au-dessus d’un seuil  $S_1$ , respectivement. Rappelons que, tel que mentionné à la section précédente, il est nécessaire d’adapter les choix de  $S_1$  et  $S_2$  en fonction de la résolution où l’on se trouve puisque l’amplitude des déplacements dépend de la résolution où l’on se

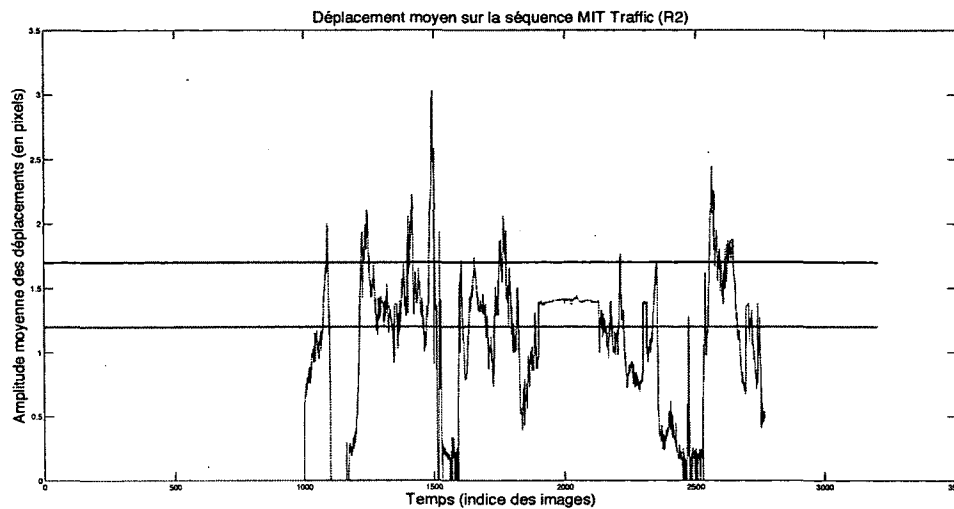


Figure 3.34 – Déplacement moyen pour la séquence *MIT Traffic* (R2) [53]

trouve : un pixel de résolution R2 représente une plus courte distance réelle qu'un pixel de résolution R4, mais une plus longue distance réelle qu'un pixel de résolution R1.

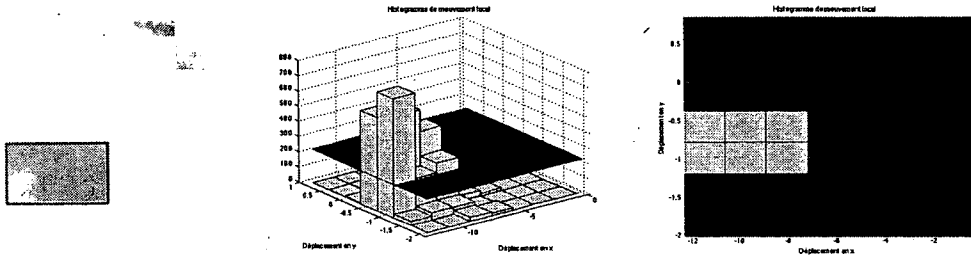
La figure 3.34 illustre le déplacement moyen pour chaque instant  $t > 1000$  de la séquence *MIT Traffic*, de même que les seuils  $S_1$  (en noir) et  $S_2$  (en rouge). De nouveau, nous devons mentionner que les seuils sont ici fixés de manière arbitraire et n'ont donc pas de signification réelle. Cela est dû au fait que le système évalue présentement les déplacements en pixels au lieu de fournir une mesure en distance réelle. Encore une fois, les seuils  $S_1$  et  $S_2$  sont fixés par l'utilisateur et dépendent du contexte d'utilisation. Enfin, si aucune mesure de déplacement n'est donnée pour  $t \in [0, 1000]$ , c'est en raison de l'utilisation en résolution R2 de la soustraction de fond. En effet, le modèle statistique de l'arrière-plan nécessite une certaine quantité de données pour se stabiliser. Les premiers résultats qu'il fournit ne sont donc pas assez fiables pour être utilisés. Les sections dans le graphique où le déplacement est nul sont aussi dues à l'étape de la soustraction de fond. Cela signifie alors qu'aucune région en mouvement n'a été détectée et donc qu'aucun flux

optique n'a été évalué.

### 3.2.3 Analyse du flux optique en résolution R1

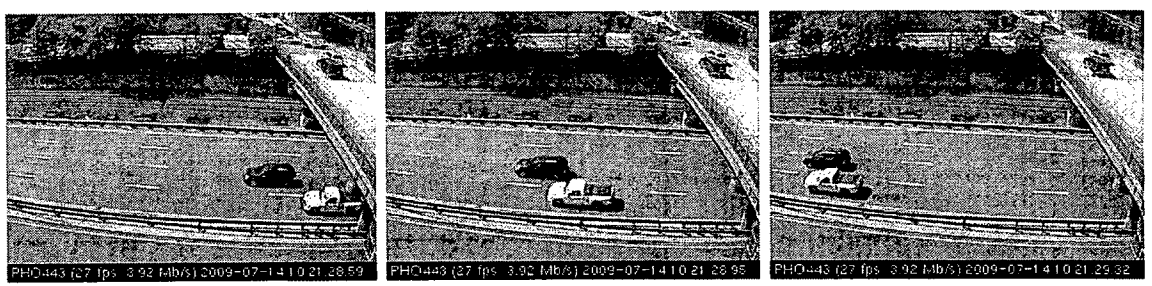
L'utilisation de la séquence à sa pleine résolution permet d'obtenir de l'information sur la vitesse de chaque véhicule. En combinant les deux techniques précédentes, la méthode proposée doit pouvoir déterminer le nombre de véhicules dans chaque région d'intérêt, de même qu'une estimation de la vitesse de ceux-ci. Or, comme l'illustre la figure 3.35a, les vecteurs de flux optique associés aux deux véhicules peuvent posséder des orientations et des amplitudes très similaires, et ce même si les deux voitures ne vont pas à la même vitesse (figure 3.36). Cette trop grande similarité ne permet pas une discrimination des deux véhicules à l'aide de l'histogramme (figure 3.35b). Le flux optique étant calculé entre deux images consécutives de la séquence vidéo, il n'offre pas la possibilité de discerner les deux véhicules sur la seule base du déplacement instantané. Il n'est pas non plus envisageable d'évaluer le flux optique sur des images plus temporellement espacées dans la séquence. En effet, tel que présenté à la section 2.2.1, la théorie sur laquelle repose l'évaluation du flux optique stipule que les déplacements doivent être minimes à l'échelle de la taille de l'image. Ainsi, en choisissant des images plus éloignées, le mouvement apparent entre les objets augmente, ce qui va à l'encontre de la condition précédente. De plus, l'utilisation de la soustraction de fond proscrit l'utilisation d'images non consécutives (ou très rapprochées), puisque les régions détectées comme étant en mouvement sont alors très différentes sur l'une et l'autre des images.

Toutefois, l'algorithme utilisé pour l'évaluation du flux optique (Horn et Schunck, [27]) ayant été choisi avant tout pour sa simplicité et sa rapidité d'exécution, il est envisageable qu'une technique plus moderne permette d'évaluer avec plus de précision le déplacement des véhicules. Il est alors permis de penser que cette augmentation de la précision des



(a) Flux optique local                      (b) Histogramme                      (c) Régions connexes

Figure 3.35 – Analyse du flux optique sur la séquence *streetLight* ( $t = 925$ ) de la série *Mouvement intermittent*



(a)  $t = 915$                       (b)  $t = 925$                       (c)  $t = 935$

Figure 3.36 – Extrait de la séquence *streetLight* de la série *Mouvement intermittent*

vecteurs composant le flux optique rende possible une meilleure discrimination de ceux-ci avec la technique de l'histogramme. En ce sens, le site web *vision.middlebury.edu/flow* [4] recense un grand nombre de techniques d'évaluation du flux optique et les trie selon plusieurs mesures de performance.

Rendu à ce point de la méthode proposée, si la vitesse du trafic ne s'améliore pas, l'algorithme poursuit son analyse à la résolution R1. Si, au contraire, la circulation reprend un rythme convenable et repasse au dessus du seuil  $S_2$ , le système repasse à la résolution R2. Ainsi, afin d'être prêt à tout moment à repasser en résolution R2, le système évalue le nombre de directions  $n_d$  pour chaque image de la séquence traitée en résolution R1. Le calcul du nombre de directions se fait de la même façon qu'en résolution R2, c'est-à-dire à l'aide d'un histogramme construit à partir des vecteurs de flux optique calculés sur les régions en mouvement.

Ainsi, à défaut de fournir des résultats fiables pour détecter chaque véhicule individuellement, l'utilisation des données à pleine résolution permet d'évaluer avec davantage de précision l'ensemble des mesures qui sont obtenues aux résolutions inférieures R2 et R4. Ainsi, même si la mesure de vitesse instantanée de chaque véhicule ne donne pas, dans l'état actuel des choses, les résultats escomptés, l'utilisation d'une résolution supérieure permet malgré tout un raffinement des mesures globales de vitesse du trafic routier (vitesse moyenne dans la scène et vitesse moyenne par direction).

Nous nous intéressons finalement à l'impact de la soustraction de fond sur le calcul du flux optique. À la section 3.2.2, nous avons vérifié que le calcul du flux optique sur des régions donnait des résultats analogues au calcul du flux optique sur l'image entière pour ce qui est de l'orientation des vecteurs obtenus. Nous nous intéressons maintenant à l'effet de l'incorporation de la soustraction de fond sur le temps d'exécution du calcul du flux optique. Nous abordons ce sujet sous deux angles : premièrement, en évaluant la proportion des pixels qui sont éliminés du traitement par la soustraction de fond ;

deuxièmement, en comparant les temps d'exécution du calcul du flux optique avec et sans la soustraction de fond.

Après vérification sur cinq séquences vidéo de trafic routier, nous constatons que la méthode de soustraction de fond présentée à la section 2.1.3 permet d'éliminer plus de 75% des pixels du processus de calcul du flux optique. Le graphique 3.37 illustre la proportion des pixels que la soustraction de fond a permis d'éliminer du traitement pour chacune des cinq séquences de test. Ainsi, c'est sans réelle surprise que nous observons une importante diminution des temps d'exécution lorsque la soustraction de fond est présente dans le traitement. Le temps requis pour traiter une séquence en résolution R1 est environ 15 fois plus rapide lorsqu'on utilise la soustraction de fond. Le graphique 3.38 met de l'avant la différence entre les temps d'exécution en secondes pour les cinq séquences de test, ce qui illustre bien l'importance de l'étape de soustraction de fond dans le traitement.



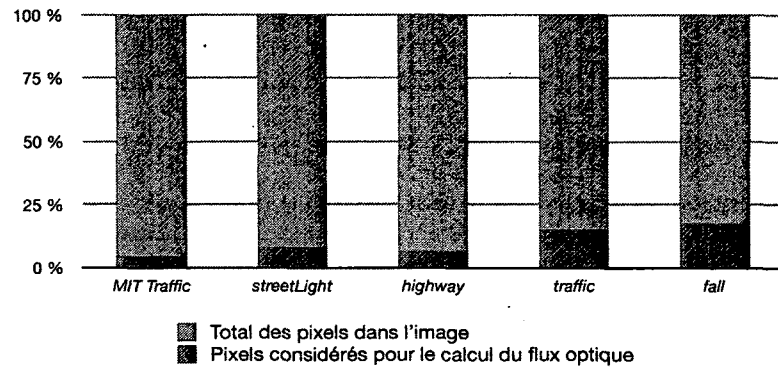


Figure 3.37 – Proportion des pixels éliminés par la soustraction de fond en résolution R1

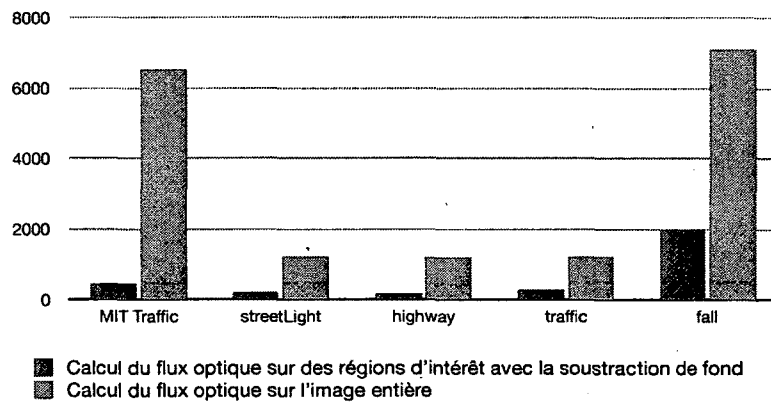


Figure 3.38 – Temps d'exécution en sec. pour le calcul du flux optique en résolution R1

# CONCLUSION ET PERSPECTIVES

Nous avons présenté une méthode simple d'estimation du mouvement du trafic routier à plusieurs niveaux de précision à partir de l'information visuelle multirésolution. De plus, la simplicité des opérations et l'utilisation de la soustraction de fond pour réduire la quantité de données à traiter permet le maintien de temps d'exécution assez bas. Il est ainsi envisageable d'inclure cette approche dans un système de traitement de l'information visuelle en temps réel.

Toutefois, pour être viable, les performances de la méthode devront être indépendantes des conditions climatiques et de luminosité, ce qui n'est pas le cas présentement. Cette robustesse représente un défi de taille, à la fois pour la détection des objets en mouvement et pour l'évaluation du déplacement des véhicules à l'aide du flux optique. En ce sens, il pourrait être pertinent d'explorer l'avenue des mélanges de gaussiennes généralisées pour la soustraction de fond, tel que suggéré par Allili *et al.* [1].

Il sera aussi nécessaire d'inclure une étape de correction de la déformation due à la perspective dans les images. Présentement, l'amplitude des déplacements estimés dépend de la distance entre les véhicules et la caméra : un véhicule plus éloigné aura un mouvement apparent inférieur à celui d'un véhicule à proximité de la caméra. La correction de l'effet de perspective, tel que proposé par Dailey *et al.* [12], règlera ce problème. Finalement, il pourrait être intéressant d'explorer des avenues permettant l'évaluation

automatique des seuils de changement de résolution  $S_1$  et  $S_2$ . Ceux-ci pourraient être estimés empiriquement pour chaque zone du réseau routier analysée.

Dans cette optique, il est important de rappeler que ce travail ne représente qu'une étape, dans un projet plus large visant l'élaboration d'un système complet pour l'analyse automatisée du trafic routier. L'accent était ici mis sur l'élaboration d'une stratégie de traitement multirésolution utilisant les données de flux optique comme indice d'estimation du déplacement des véhicules. Le fait d'avoir mis de côté d'importantes contraintes d'acquisition rendait impossible l'évaluation de la vitesse réelle des véhicules en mouvement, mais nous avons malgré tout illustré le potentiel des données de flux optique comme source d'information pour l'analyse de la circulation routière.

Il ne s'agit bien entendu que d'un aperçu de ce que l'information visuelle permettra d'obtenir comme information dans le futur. Il pourrait par exemple être possible d'intégrer les données routières obtenues avec de l'information visuelle dans des systèmes de guidage automatisé des véhicules de même que dans des outils d'aide à la navigation. Il serait aussi intéressant de permettre une analyse à plus grande échelle, sur un réseau de caméras de surveillance afin, par exemple, de rendre possible la recherche de la source des ralentissements dans le flot de trafic. La gestion et la surveillance automatisées du trafic routier à l'aide de l'information visuelle n'en sont qu'à leurs balbutiements.

# Bibliographie

- [1] M. S. ALLILI, N. BOUGUILA et D. ZIOU : Finite General Gaussian Mixture Modeling and Application to Image and Video Foreground Segmentation. *Journal of Electronic Imaging*, 17(1), 2008.
- [2] J. BADENAS, M. BOBER et F. PLA : Segmenting Traffic Scenes from Grey Level and Motion Information. *Pattern Analysis & Applications*, 4(1):28–38, 2001.
- [3] S. BAKER, D. SCARSTEIN, J.P. LEWIS, S. ROTH, M.J. BLACK et R. SZELISKI : A Database and Evaluation Methodology for Optical Flow. *International Journal of Computer Vision*, 92(1):1–8, 2011.
- [4] S. BAKER, D. SCHARSTEIN, J.P. LEWIS, S. ROTH, M. BLACK et R. SZELISKI : [vision.middlebury.edu/flow](http://vision.middlebury.edu/flow/), 2012. URL <http://vision.middlebury.edu/flow/>.
- [5] J. L. BARRON, D. J. FLEET, S. S. BEAUCHEMIN et T. A. BURKITT : Performance of Optical Flow Techniques. *International Journal of Computer Vision*, pages 236–242, 1994.
- [6] M.J. BLACK et P. ANANDAN : Robust Dynamic Motion Estimation Over Time. Dans *Computer Vision and Pattern Recognition*, 1991, pages 296–302, 1991.
- [7] Andrés BRUHN, Joachim WEICKERT et Christoph SCHNÖRR : Lucas/Kanade Meets

- Horn/Schunck : Combining Local and Global Optic Flow Methods. *International Journal of Computer Vision*, 61:211–231, 2005.
- [8] W. BURGER et B. BHANU : Estimating 3D Egomotion from Perspective Image Sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12 (11):1040–1058, 1990.
- [9] Y. CHO et J. RICE : Estimating Velocity Fields on a Freeway from Low-Resolution Videos. *IEEE Transactions on Intelligent Transportation Systems*, 7(4):463–469, 2006.
- [10] C. COHEN : Prime Mover. *Baltimore City Paper*, 2000.
- [11] B. CYGANEK et J.P. SIEBERT : *An Introduction to 3D Computer Vision Techniques and Algorithms*. Wiley, 2009.
- [12] D.J. DAILEY., F.W. CATHEY et S. PUMRIN : An Algorithm to Estimate Mean Traffic Speed using Uncalibrated Cameras. *IEEE Transactions on Intelligent Transportation Systems*, 1(2):98–107, 2000.
- [13] S. DOĞAN, M.S. TEMİZ et S. KÜLÜR : Real Time Speed Estimation of Moving Vehicles from Side View Images from an Uncalibrated Video Camera. *Sensors*, 10 (5), 2010.
- [14] W. ENKELMANN : Interpretation of Traffic Scenes by Evaluation of Optical Flow Fields from Image Sequences. *International Federation of Automatic Control : Symposium of Control, Computer, Communications in Transportation*, pages 43–50, 1989.
- [15] M. FATHY et M.Y. SIYAL : A Window-Based Image Processing Technique for Quantitative and Qualitative Analysis of Road Traffic Parameters. *IEEE Transactions on Vehicular Technology*, 47(4):1342–1349, 1998.

- [16] A. FERNÁNDEZ-CABALLERO, F. GÓMEZ et J. LÓPEZ-LÓPEZ : Road-Traffic Monitoring by Knowledge-Driven Static and Dynamic Image Analysis. *Expert Systems with Applications : An International Journal*, 35(3):701–719, 2008.
- [17] N. FRIEDMAN et S. RUSSELL : Image Segmentation in Video Sequences : A Probabilistic Approach. *Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence*, pages 175–181, 1997.
- [18] A. GIACHETTI, M. CAMPANI et V. TORRE : The Use of Optical Flow for Road Navigation. *IEEE Transactions on Robotics and Automation*, 14(1):34–48, 1998.
- [19] R. GORDON et W. TIGHE : Traffic Control Systems Handbook. Rapport technique, Federal Highway Administration, 2005.
- [20] N. GOYETTE, P.M. JODOIN, F. PORIKLI, J. KONRAD et P. ISHWAR : changedetection.net : A New Change Detection Benchmark Dataset. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2012.
- [21] H. GRIFFITHS : New Ideas in FM Radar. *Electronics & Communication Engineering Journal*, 2(5), 1990.
- [22] D. GRIMES et T. JONES : Automotive Radar : A Brief Review. *Proceedings of the IEEE*, 62(6), 1974.
- [23] S. GUPTA, O. MASOUD, R.F.K. MARTIN et N.P. PAPANIKOLOPOULOS : Detection and Classification of Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 3(1):37–47, 2002.
- [24] M. HEIKKILA et M. PIETIKAINEN : A Texture-Based Method for Modeling the Background and Detecting Moving Objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):657–662, 2006.

- [25] M. HEIKKILA, M. PIETIKAINEN et J. HEIKKILA : A texture-Based Method for Detecting Moving Objects. *British Machine Vision Conference*, pages 187–196, 2004.
- [26] B. HELLINGA : Improving Freeway Speed Estimates from Single-Loop Detectors. *Journal of Transportation Engineering*, 128(1):58–67, 2002.
- [27] B.K.P. HORN et B.G. SCHUNCK : Determining Optical Flow. *Artificial Intelligence*, 17:185–203, 1981.
- [28] X. JI, Z. WEI et Y. FENG : Effective Vehicle Detection Technique for Traffic Surveillance Systems. *Journal of Visual Communication and Image Representation*, 17(3):647–658, 2006.
- [29] P. KAEWTRAKULPONG et R. BOWDEN : An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection. *European Workshop on Advanced Video Based Surveillance Systems*, 2001.
- [30] V. KASTRINAKI, M. ZERVAKIS et K. KALAITZAKIS : A Survey of Video Processing Techniques for Traffic Applications. *Image and Vision Computing*, 21(4):359–381, 2003.
- [31] L. KLEIN, M. MILLS et D. GIBSON : Traffic Detector Handbook. Rapport technique, Federal Highway Administration, 2006.
- [32] D. KOLLER, J. WEBER, T. HUANG, J. MALIK, G. OGASAWARA, B. RAO et S. RUSSELL : Towards Robust Automatic Traffic Scene Analysis in Real-time. *International Conference on Pattern Recognition*, 1:3776–3781, 1994.
- [33] D.S. LEE : Effective Gaussian Mixture Learning for Video Background Subtraction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):827–832, 2005.

- [34] Victor LEMPITSKY, Stefan ROTH et Carsten ROTHER : FusionFlow : Discrete-Continuous Optimization for Optical Flow Estimation. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [35] W.H. LIN, J. DAHLGREN et H. HUO : An Enhancement to Speed Estimation using Single Loop Detectors. *Proceedings of 2003 Intelligent Transportation Systems*, 1: 417–422, 2003.
- [36] B.D. LUCAS et T. KANADE : An Iterative Image Registration Technique with an Application to Stereo Vision. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, 2:674–679, 1981.
- [37] P. McLAUHLAN, D. BEYMER, B. COIFMAN et J. MALI : A Real-time Computer Vision System for Measuring Traffic Parameters. *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition*, pages 495–501, 1997.
- [38] J. MOHAMMADI, R. AKBARI et M. KESHAVARZ BA HAGHIGHAT : Vehicle Speed Estimation Based on the Image Motion Blur using RADON Transform. *International Conference on Signal Processing Systems*, 1:243–247, 2010.
- [39] L. NEUDORFF, J. RANDALL, R. REISS et R. GORDON : Freeway Management and Operations Handbook. Rapport technique, Federal Highway Administration (U.S. Department of Transportation), 2003.
- [40] T. POTTER : The Evolution of Inductive Loop Detector Technology. Rapport technique, Reno A&E, 2005.
- [41] T. RABIE, B. ABDULHAI, A. SHALABY et A. EL-RABBANY : Mobile Active-Vision Traffic Surveillance System for Urban Networks. *Computer-Aided Civil and Infrastructure Engineering*, 20(4):231–241, 2005.



- [42] D. RIAHI, P.L. ST-ONGE et G.A. BILODEAU : RECTGAUSS-*Tex* : Block-based Background Subtraction. Rapport technique, École Polytechnique de Montréal, 2012.
- [43] A. SATPATHY, H.L. ENG et X. JIANG : Difference of Gaussian Edge-Texture Based Background Modeling for Dynamic Traffic Conditions. *Proceedings of the 4th International Symposium on Advances in Visual Computing*, pages 406–417, 2008.
- [44] T.N. SCHOEPFLIN et D.J. DAILEY : Dynamic Camera Calibration of Roadside Traffic Management Cameras for Vehicle Speed Estimation. *IEEE Transactions on Intelligent Transportation Systems*, 4(2):90–98, 2003.
- [45] D. SCHRANK, B. EISELE et T. LOMAX : Tti’s 2012 urban mobility report. Rapport technique, Texas A&M Transportation Institute, 2012.
- [46] A.C. SHASTRY et R.A. SCHOWENGERDT : Airborne Video Registration and Traffic-flow Parameter Estimation. *IEEE Transactions on Intelligent Transportation Systems*, 6(4):391–405, 2005.
- [47] C. STAUFFER et W.E.L. GRIMSON : Adaptive Background Mixture Models for Real-time Tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2, 1999.
- [48] C. STAUFFER et W.E.L. GRIMSON : Learning Patterns of Activity Using Real-time Tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):747–757, 2000.
- [49] D. SUN, S. ROTH et M.J. BLACK : Secrets of Optical Flow Estimation and Their Principles. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2432–2439, 2010.

- [50] Deqing SUN, J. P. LEWIS et Michaelj. BLACK : Learning optical flow. Dans *In Proceedings of European Conference on Computer Vision*, pages 83–97, 2008.
- [51] Richard SZELISKI : *Computer Vision : Algorithms and Applications*. Springer, 2010.
- [52] S.A. VELASTIN : Image Processing for On-line Analysis of Crowds in Public Areas. *Transportation systems : theory and application of advanced technology*, 1:163–168, 1995.
- [53] Xiaogang WANG, Xiaoxu MA et W.E.L. GRIMSON : Unsupervised Activity Perception in Crowded and Complicated Scenes Using Hierarchical Bayesian Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3):539–555, 2009.
- [54] Y. XIE, Z. YE et Y. ZHANG : Accurate Speed Estimation using Single Loop Detector Data. Rapport technique, Southwest Region University Transportation Center, 2007.
- [55] X. YU, P. XUE, L. DUAN et Q. TIAN : An Algorithm to Estimate Mean Vehicle Speed rom MPEG Skycam Video. *Multimedia Tools and Applications*, 34(1), 2007.
- [56] W. ZHANG, Q.M.J. WU et H.B. YIN : Moving Vehicles Detection Based on Adaptive Motion Histogram. *Digital Signal Processing*, 20(3):793–805, 2010.
- [57] Z. ZIVKOVIC : Improved Adaptive Gaussian Mixture Model for Background Subtraction. *Proceedings of the 17th International Conference on Pattern Recognition*, 2:28–31, 2004.