

Sommaire

I	- Introduction	7
1.	Contexte de la problématique (l'aide à la conduite)	8
2.	Problème de vision face aux scènes à grande plage dynamique.	21
3.	La reconnaissance des voies routières et la transformée de Radon	23
4.	Contributions.....	24
5.	Organisation du mémoire.....	25
II	- Etat de l'art	26
1	La vision face aux scènes à grande plage dynamique.....	28
1.1	Introduction	28
1.2	La luminance et la plage dynamique en photographie	29
1.3	Une limite du capteur visuel.....	30
1.4	Photographie computationnelle	32
1.4.1	Imagerie à grande plage dynamique.....	33
1.4.1.a.	Estimation de la fonction de réponse de la caméra.....	34
1.4.1.b.	Construction de l'image à grande plage dynamique.....	39
1.4.1.c.	Reproduction de tons.....	39
1.4.2	Fusion directe des expositions.....	44
1.4.2.a.	Fusion de blocs.....	44
1.4.2.b.	Algorithme de Tom Mertens	45
1.4.2.c.	Fusion d'expositions basée sur la direction du gradient.....	48
1.5	Conclusion	49
2	La détection des voies routières	51
2.1	Introduction	51
2.2	Etapas de détection des voies routières	53
2.2.1	Prétraitement	53
2.2.2	Extraction de caractéristiques.....	55
2.2.3	Ajustement de modèle	56
2.3	Conclusion	57

III - Contributions	58
3 Fusion d'expositions basée sur la segmentation d'histogrammes	60
3.1 Introduction	60
3.2 La Segmentation d'images	61
3.3 Segmentation d'histogrammes	63
3.4 Fusion des régions segmentées.....	66
3.5 Acquisition des expositions.....	68
3.6 Résultats et comparaisons.....	69
3.7 Conclusion	74
4 Fusion d'expositions basée sur l'apprentissage supervisé	76
4.1 Introduction	76
4.2 Estimation de poids généralisés pour la fusion d'exposition	77
4.3 Expérimentation et comparaisons.....	79
4.4 Discussion des résultats	86
4.5 Estimation de modèle de fusion d'expositions.....	87
4.6 Expérimentation et comparaisons.....	93
4.7 Conclusion	96
5 Détection précise de lignes sur les images	98
5.1 Introduction	98
5.2 L'imprécision de la transformée de Radon.....	99
5.3 Transformée de Radon Gaussienne	101
5.4 Résultats et comparaisons.....	103
5.5 Conclusion.....	106
IV - Conclusion générale et perspectives	108
V - Table des figures	110
VI - Liste des tableaux	112
VII – Bibliographie	113

I- Introduction

1. Contexte de la problématique (l'aide à la conduite)

Communément, la conduite automobile est une activité qui consiste à utiliser un véhicule dans le but de se déplacer d'un emplacement à un autre, en empruntant la route et en respectant des règles qui assurent une interaction sécurisée entre les différents usagers de la route. Lors de la conduite, le chauffeur se trouve inséré dans le système conducteur-véhicule-infrastructure. En fonction des informations qu'il perçoit sur l'environnement qui l'entoure et sur l'état de son véhicule il choisit avec son expérience les commandes à effectuer pour mener à bien la tâche de conduite qu'il s'est fixée. Les informations utiles au conducteur sont celles qui vont le renseigner d'une part sur l'environnement (la topologie de la route, les marquages et les panneaux de signalisation, les obstacles et usagers de la route, condition climatiques...etc.) d'autre part sur l'état et le comportement de son véhicule (Vitesse, direction...etc.).

[Michon, 1985] voit l'activité de conduite comme un ensemble de tâches simultanées ayant des exigences temporelles et cognitives différentes. Il décrit trois niveaux hiérarchiques pour catégoriser les tâches qui composent l'activité humaine en générale: le niveau stratégique, le niveau tactique et le niveau opérationnel. Au niveau stratégique, niveau supérieur du modèle, les tâches exigent un investissement cognitif élevé et des ressources attentionnelles disponibles. En revanche, l'exigence temporelle de ces tâches est relativement faible. Au niveau tactique, niveau intermédiaire du modèle, les tâches relèvent des prises de décision concernant les actions qui vont être mises en œuvre dans les instants qui suivent. Le coût cognitif de la tâche est moins élevé qu'au niveau supérieur, mais son exigence temporelle est relativement forte. Enfin, le niveau opérationnel correspond à l'exécution des actions décidées aux niveaux supérieurs. Situé au plus bas dans la hiérarchie du modèle, les tâches dépendant de ce niveau sont automatisées. Leur coût cognitif est très faible, voire inexistant mais l'exigence temporelle liée à ces tâches est très élevée.

Cette répartition de l'activité humaine en tâches hiérarchiques semble pertinente. Le conducteur accomplit de manière plus ou moins consciente différentes tâches situées sur les trois niveaux hiérarchiques de Michon (Stratégique, tactique et opérationnel). Pour ce qui est du niveau le plus haut, on trouve la navigation. Une tâche peu fréquente où le

conducteur trace son itinéraire en se basant sur des connaissances acquises. Le pilotage du véhicule et une tâche plus fréquente qui consiste à adapter la conduite selon la situation venue (ralentir, garder sa voie et la distance inter véhicules, dépassement...etc.). La commande du véhicule est un ensemble de tâches situées au plus bas niveau hiérarchique du modèle de Michon. Le conducteur s'habitue facilement à ces tâches et les effectuent presque inconsciemment durant la conduite car elles sont fréquentes et demandent très peu d'efforts cognitifs.

Intégration d'un système d'aide à la conduite

Pour exécuter les différentes tâches de conduite citées un peu plus haut, le conducteur fait plus ou moins appel à des sous tâches qui sont la collecte d'informations et la prise de décisions. Un système d'aide à la conduite doit être en mesure de collecter les informations sur la scène routière les analyser pour déterminer la nécessité d'une intervention. Selon le système d'aide à la conduite visé il conviendra de prendre en compte les informations propres au véhicule, ou alors relatives à l'infrastructure. Pour une assistance de bas niveau, seules les données inhérentes au véhicule sont nécessaires. Un régulateur de vitesse n'a besoin que des informations de régime moteur et de freinage. Pour une assistance de plus haut niveau des informations extérieures, souvent perçues visuellement par le conducteur sont indispensables (marquage au sol, carte routière, objets,...etc).

Vers des systèmes avancés d'aide à la conduite (ADAS)

La plupart des véhicules récents disposent de systèmes d'aide à la conduite de base tels que le système de freinage antiblocage, la direction assistée, boîte de vitesse automatique et d'autres. Ce genre de systèmes fournit une assistance de bas niveau au conducteur. La voiture de demain devra être plus intelligente et capable d'offrir une assistance accrue au conducteur face aux situations difficiles. Elle devra contenir des systèmes capables d'interpréter les scènes routières afin d'intervenir d'une façon ou d'une autre pour éviter les drames, d'optimiser la conduite ou d'apporter plus de confort. Ces systèmes d'aide à la conduite de haut niveau, nommé ADAS (Advanced Driver Assistance Systèmes) sont de plus en plus demandés par les constructeurs automobiles. Un système d'aide à la conduite peut être passif et se contenter d'informer le conducteur

d'un éventuel danger par un signal lumineux ou par des bips sonores...etc. Il peut aussi être actif, en prenant des décisions et en interagissant avec le véhicule. Ainsi, dans le cas de dispositifs d'information, de conseil ou d'avertissement seulement, l'action sur le véhicule reviendra toujours au conducteur. Pour des systèmes d'assistance actifs, il conviendra de définir lors de leur conception le niveau de collaboration avec le conducteur. La sous-section suivante présente quelques systèmes avancés d'aide à la conduite les plus prometteurs et déjà commercialisés dans le marché.

La conception de n'importe quel système nécessite une bonne connaissance du contexte dans lequel il va fonctionner. Pour concevoir un système de sécurité routière, il est important de bien comprendre les interactions qui peuvent avoir lieu entre les acteurs : conducteur, véhicule, environnement. La conduite automobile est une activité complexe qui consiste à faire fonctionner un véhicule dans le but de se déplacer d'un endroit à un autre, en respectant certaines contraintes comme celles émanant du code de la route. Pour conduire, le chauffeur a besoin de collecter toutes sortes d'informations et d'avoir des connaissances liées au véhicule et à l'environnement qui l'entoure. Durant la conduite, le chauffeur accomplit des tâches sur des niveaux différents. Dans le niveau le plus haut, on trouve la planification du trajet ou bien la navigation (tâche la moins fréquente). Le conducteur choisit son itinéraire en se basant sur des connaissances sur le parcours. Le guidage du véhicule est une tâche plus fréquente qui consiste à adapter la conduite selon la scène visualisée (garder sa voie de roulement, faire des dépassements, éviter un obstacle, etc.). Puis viennent au plus bas niveau les actions pour commander le véhicule. Pour accomplir les différentes tâches de conduite, le conducteur a besoin d'une quantité d'information spécifique à chacune. Les informations sont analysées dans le cerveau puis des actions de commande sont générées. Pour conclure, trois activités sont donc constamment sollicitées : la perception, la décision et la commande.

L'activité de perception est primordiale pour la conduite car elle permet de collecter les informations nécessaires à l'exécution de toutes les tâches. La vision est le capteur le plus important pour la perception. Le conducteur a besoin de voir l'environnement afin de se repérer géographiquement et utiliser ses connaissances (cartes) pour atteindre la

destination qu'il s'est fixée. Il doit voir la scène routière pour prendre les bonnes décisions afin de guider son véhicule en toute sécurité.

Il est reconnu que le comportement du conducteur est la cause de la plupart des accidents de la circulation (Etat de fatigue, baisse de vigilance, incompetence.). Si on néglige les risques dus à l'infrastructure et à la conception des véhicules (risques faibles) et si on considère la définition de la conduite automobile, un problème de la circulation survient à cause du dysfonctionnement de l'une de ces tâches : perception, analyse, commande. Le conducteur peut ne pas rassembler suffisamment d'informations, ou au contraire être submergé par celles-ci. Il convient de citer aussi les difficultés apparaissant au stade décisionnel (prendre la meilleure décision face à telle situation). Concernant les actions de commande, elles représentent les choix effectués par le conducteur. Elles doivent donc représenter au mieux les décisions prises. Ceci repose sur la bonne conception du véhicule et de l'interface conducteur-véhicule.

L'aide à la conduite est une branche de la sécurité routière active qui tente d'assister le conducteur à accomplir correctement ces tâches de conduite dans les différents niveaux. Tout système allégeant ou facilitant la tâche au conducteur peut être considéré comme une aide à la conduite. Voici des exemples de systèmes d'aide à la conduite:

- ABS (Système anti-blocage des roues), Direction Assistée : appartiennent à l'Assistance à la commande (bas niveau) qui a pour rôle de faciliter l'activité de commande du véhicule.
- Le Guidage Par Satellite : pour une assistance de plus haut niveau telle que la navigation. Des informations telles que les cartes routières et les techniques de localisation sont nécessaires.

- Evitement de collisions, détection de franchissement de voies routières : Pour ce genre de systèmes (guidage) des informations extérieures, souvent perçues visuellement par le conducteur sont indispensables (marquage au sol).

ADAS commercialisés

Adaptive Cruise contrôle : C'est un système de régulation de distance de sécurité. Equipé d'un capteur d'acquisition actif (un radar ou un capteur laser), ce système mesure la distance et la vitesse relative entre le véhicule et celui qui le précède. Si l'écart devient trop faible, le système réduit la vitesse en décélérant ou en sollicitant automatiquement les freins. Lorsque la voie se libère le système ré-accélère jusqu'à la vitesse mémorisée. Mercedes-Benz fut parmi les premiers constructeurs automobiles à introduire un tel système.

Lane Assistance: c'est un système de détection de changement involontaire de voies. En générale, ces systèmes reposent sur une caméra installée dans la partie du rétroviseur interne du véhicule. L'Audi lane assist par exemple est un système qui avertit le conducteur lorsqu'il quitte sa file sans avoir préalablement actionné le clignotant. Ce système utilise une caméra située au-dessus du rétroviseur intérieur qui couvre un angle de 40° sur une distance de 60 mètres. L'application détecte grâce à cette caméra les marquages au sol et calcule la position du véhicule pour respecter ce marquage. Si le véhicule s'en écarte, une vibration se déclenche dans le volant pour attirer l'attention du conducteur. Il existe des systèmes similaires actifs, qui contre-braquent au moment du franchissement de voie.

Navigation Assistance: ce sont des systèmes conçus pour aider le conducteur à se rendre à sa destination en le guidant souvent par une voix. Les moyens techniques proposés de nos jours permettent une aide avancée en proposant un guidage en temps réel (le trajet est recalculé si l'utilisateur n'emprunte pas le chemin proposé). Récemment, ces systèmes se sont vu améliorer par la prise en compte de l'état du trafic en temps réel et

des statistiques d'utilisation des routes selon les heures de la journée. Ils proposent désormais une aide à la conduite encore plus avancée qu'elle n'était avant en fournissant des trajets différents en cas de bouchon et selon l'heure de la journée.

Reconnaissance de panneaux de signalisation : Parfois, les conducteurs ne remarquent pas les panneaux de signalisation, ou oublient la vitesse autorisée dans un tronçon de route. Des systèmes munis généralement d'une caméra montée sur le rétroviseur interne du véhicule permettent de détecter et de reconnaître les panneaux de signalisation afin de prévenir les signaler en permanence au conducteur. Le Traffic Sign Recognition de Mercedes classe S en est un exemple.

Aide au parking : l'Audi Parking System par exemple est conçu avec un ensemble de dix capteurs différents (actifs et passifs). Lorsque la marche arrière est enclenchée, une image s'affiche sur l'écran associée à des lignes virtuelles qui vont aider le conducteur dans ses manœuvres, pour un stationnement en créneau comme pour un stationnement à angle droit par rapport à la chaussée.

Night View Assist : C'est un système qui permet de projeter au conducteur une image en temps réel de l'environnement durant la nuit sur un écran placé dans le tableau de bord. Le système Night Vision de Mercedes par exemple est basé sur une caméra thermique qui produit une image en temps réel de la route et grâce à un algorithme de détection de piétons avertit le conducteur en cas de détection.

Capteurs pour l'aide à la conduite

Il existe aujourd'hui une panoplie de capteurs utilisée pour les systèmes d'aide à la conduite. Deux catégories de capteurs sont généralement distinguées: Ceux qui décrivent l'état du véhicule (capteurs proprioceptifs) et ceux qui se chargent de percevoir l'environnement (capteurs extéroceptifs). Les capteurs proprioceptifs se chargent de mesurer un attribut en fonction de leur propre état. Cet attribut peut être l'accélération, l'orientation ou la vitesse de l'objet sur lequel ils sont montés. Par exemple, les capteurs

d'inclinaison, les accéléromètres et les odomètres. La deuxième classe de capteurs, ceux qui se chargent de percevoir l'environnement, sont capables de mesurer un attribut d'un objet externe présent dans la scène. Les caméras qui fonctionnent dans le spectre visible ou infrarouge, les sonars, les Lidars et les Radars sont des exemples de capteurs extéroceptifs. Le choix de l'un de ces capteurs repose sur leurs propriétés et les objectifs visés par l'application.

Une autre propriété importante qui caractérise particulièrement les capteurs extéroceptifs est le fait qu'ils soient actifs ou passifs. Les capteurs actifs collectent l'information sur l'environnement d'une manière intrusive : ils envoient des signaux dans l'environnement et mesurent ensuite l'interaction de ces signaux avec cet environnement (exemple : Lidar, Radar, sonar). À la différence des capteurs actifs, les capteurs passifs récupèrent l'information contenue dans l'environnement d'une manière non intrusive : Ils n'émettent pas de radiations, ils ne font que recevoir un signal qui peut être réfléchi, émis ou transmis par des sources d'énergie externes. Les caméras visibles et certaines caméras infrarouges sont des exemples de capteurs passifs. Dans ce qui suit sont énumérés quelques capteurs extéroceptifs les plus utilisés par les systèmes d'aide à la conduite :

Camera vision : La caméra est un capteur passif ayant une fonction similaire à la vision chez l'être humain. Elle permet de restituer l'environnement sous forme d'images ou de vidéo. Les images capturées par les caméras visibles (en couleurs ou en niveaux de gris) sont d'un côté très riches en contenu, de l'autre, difficiles à interpréter. Ceux-ci sont les raisons pour lesquelles la recherche s'est beaucoup focalisée sur ce capteur. Il existe dans la littérature un très grand nombre de travaux et d'algorithmes qui ont été proposés pour la détection, le suivi et la reconnaissance d'objets dans les images. De plus, avec l'avancée technologique, les caméras sont devenues peu coûteuses et faciles à installer sur les véhicules. Utiliser un seul capteur de ce genre présente des avantages de coût et de simplicité de mise en œuvre, mais il existe une autre approche qui consiste à utiliser deux ou plusieurs caméras. Ça permet de restituer la profondeur de la scène observée. Cette approche appelée stéréovision, s'agit de combiner deux images prises de points de vue différents puis restituer la scène en trois dimensions. La mise en correspondance entre les points de vue est la phase du traitement la plus difficile et la plus coûteuse.

Caméra thermique: C'est un capteur passif qui permet de mesurer les rayonnements infrarouges émis par les objets chauds. Ce capteur transforme une image captée dans le domaine infrarouge en une image visible par l'œil humain. L'avantage majeur des caméras thermiques est le fait qu'elles peuvent produire des images visibles dans l'obscurité complète. Cependant, leur portée est dépendante des conditions atmosphériques, du type de la caméra et de la différence de température de la cible avec le fond. Le brouillard et la pluie limitent cette portée car le rayonnement infrarouge est affaibli. Plus le contraste thermique est élevé, plus il est facile de détecter des cibles sur un fond de température constante. Récemment, des constructeurs automobiles comme Honda, Mercedes et BMW ont commercialisé un dispositif qui donne une vision routière où les piétons et les animaux présentent les zones les plus claires de l'image. Bien que l'avantage de l'utilisation des caméras infrarouges pour la détection de piétons soit évident, les véhicules et d'autres obstacles n'émettant pas de chaleur, sont difficilement repérables dans ces images.

Radar : C'est un capteur actif qui émet des ondes en une série de pulsions puis reçoit la partie d'énergie réfléchiée par la cible. Le temps nécessaire à l'onde pour voyager de la cible jusqu'à l'objet permet de déterminer la distance ou la vitesse dans le cas de multiples émissions. Les ondes du radar peuvent pénétrer à travers les nuages, la pluie...etc. Ainsi, il est considéré comme un capteur robuste face à plusieurs conditions défavorables. De plus, il a une portée très élevée couvrant la possibilité de détecter des objets très distants. Tous ces avantages ont contribué de façon évidente à l'apparition de systèmes Adaptive Cruise Control. Néanmoins, la faible résolution spatiale (surtout dans le sens latéral) du Radar entraîne des détections moins fiables voir même inexistantes pour les petits obstacles. En outre, les parties métalliques présentent entre autres une réflectivité supérieure comparée aux autres objets comme les êtres humains. Ainsi, les objets qui présentent une forte réflexion minimisent l'effet des autres réflexions moins fortes et conduisent donc à des fausses détections. Enfin, un inconvénient majeur du Radar est le problème d'interférences qui se dégrade en présence de plusieurs voitures utilisant la même technologie dans le trafic.

Lidar : ce capteur actif est basé sur la mesure du temps mis par la lumière réfléchiée sur l'obstacle se trouvant dans l'axe de tir du laser. Les Lidars se basent sur le même principe que les Radars étant donné que la distance aux objets est déterminée par le temps séparant les pulsions transmises et reçues. Ils sont utilisés pour des distances d'environ 40m et ont une grande précision dans les deux directions : longitudinale et latérale. Toutefois, le contenu des images Lidar est différent de celui des images visibles. En effet, le Lidar fournit une image de profondeur, tandis que les caméras captent la réflexion de la lumière visible. Ce capteur est largement utilisé par la communauté robotique pour la navigation de robots en terrain inconnu. Néanmoins, le coût, l'encombrement et la consommation d'énergie élevée limitent son utilisation en embarque sur un véhicule.

Technologies de Communication V2V, I2V : La communication sans fil entre les véhicules et les stations de base cellulaires est devenue une réalité. Les techniques d'échanges d'informations entre véhicules et le monde extérieur deviennent faisables. Les véhicules peuvent communiquer des informations pour éviter des dangers. Les autres usagers de la route tels que les piétons peuvent bénéficier de ces technologies si détectés et évité par un système d'assistance à la conduite intégré dans le véhicule, ou informé du danger par communication sans fil. Les technologies Vehicle-to-Vehicle et Vehicle-to-infrastructure feront peut-être partie de la conduite assistée mais pas dans un futur proche car ça demande de grands investissements concernant l'infrastructure et les équipements de communication.

Capteur cartographique (Horizon électronique): Ce capteur est un peu spécial car il s'agit d'une combinaison d'une base de données cartographique avec un système de localisation GPS (Global Positioning System). Selon la position du véhicule estimée par le GPS, le capteur peut fournir un grand nombre d'informations statiques à un système d'aide à la conduite automobile relative à la route actuelle, mais aussi à l'infrastructure à venir. Les systèmes d'aide à la conduite disponibles aujourd'hui dans le marché se basent sur ce capteur sont dédiés à l'assistance à la navigation routière (niveau stratégique : guider le chauffeur à atteindre sa destination). Des travaux récents commencent à viser des systèmes plus complexes qui se basent sur des informations plus précises sur

l'environnement routier. Il devient donc possible par exemple de proposer au conducteur la vitesse et la trajectoire optimales pour passer le virage suivant en toute sécurité.

La précision du système de localisation demandée est liée à l'application visée. Une application d'aide à la navigation n'a pas besoin d'une grande précision, une marge de quelques mètres est tolérable. Mais pour une application telle que l'évitement de franchissement de lignes, l'application doit bénéficier d'une précision à l'ordre de centimètres ou de décimètres. Cette même précision est exigée généralement par les applications de contrôle latéral du véhicule. Ces dernières années, on trouve de plus en plus de travaux sur l'aide à la conduite basé sur ce capteur. Par exemple [Polychronopoulos, 2005] utilise ce capteur pour l'aide au maintien du véhicule dans la voie via la prédiction de la géométrie à venir de la route. Un autre système propose d'anticiper l'approche des virages afin de définir la vitesse appropriée et d'adapter l'éclairage des phares [Lauffenburger , 2007]. [Revue, 2002] propose d'avertir le conducteur des risques potentiels à l'approche de carrefour en se basant toujours sur l'analyse du graphe routier courant.

Fusion de capteurs pour l'aide à la conduite

Pour interpréter correctement les scènes routières et intervenir en cas de nécessité, le système d'aide à la conduite doit d'abord percevoir correctement les informations sur l'environnement. Car la phase de perception est primordiale dans les processus de reconnaissances et d'interprétations. Pour ceci, la technologie offre un éventail de capteurs ayant des propriétés différentes. Mais il s'avère que chaque capteur a ces faiblesses et des points forts. Pour aboutir à des systèmes fiables et robustes face aux conditions adverses (climat, illumination, occlusion, interférences...etc.), il devient nécessaire de fusionner les informations issues de plusieurs capteurs (mêmes capteurs, ou capteurs de types différents).

Définition de la fusion de données

La fusion de données consiste d'une façon générale à combiner différentes informations relatives à un problème. On trouve plusieurs définitions sur la fusion de données. Voici la définition de [Bloch, 2003] : « La fusion d'informations consiste à

combiner des informations issues de plusieurs sources afin d'améliorer la prise de décision. ». L'homme est un exemple très évocateur de ce qu'est la fusion de données. Celui-ci perçoit les informations autour de lui par ses cinq sens et utilise ses connaissances pour interagir avec les objets, prendre des décisions ou bien améliorer encore sa base de connaissances. Concernant le sujet abordé ici, pour conduire le chauffeur analyse les différentes informations lui provenant de sa vue (frontale, latérale, arrière), son ouïe (bruit du moteur, klaxon, clignotant, frein), son toucher (retour d'effort volant, pédales), son équilibre, ses connaissances (le trajet, le code de la route,...)...etc. Puis prend des décisions sur la manière de conduire.

La fusion de capteurs peut être vue comme l'utilisation de données fournies par plusieurs sources d'informations pour construire une base de connaissance plus précise et plus riche que les informations initiales prises indépendamment les unes des autres. L'intérêt de la fusion est lié à l'amélioration des capacités de décision.

Niveaux de fusion de données :

Le niveau de fusion de données définit le niveau dans lequel les informations provenant des différents capteurs sont fusionnées. On retrouve dans la littérature plusieurs définitions de ces niveaux. [Alexandre, 2009] distingue 3 niveaux de fusion différents:

Niveau signal : A ce niveau, la fusion consiste à combiner des signaux de même nature pour obtenir un signal de meilleure qualité. Tel que le BRACKETING d'expositions (combiner plusieurs images originales différemment exposées).

Niveau attribut : A ce niveau, la fusion permet d'augmenter la confiance et la précision sur les attributs des objets (par exemple la taille), de créer de nouveaux attributs. Il s'agit du premier niveau d'abstraction qui nécessite une première interprétation des informations [Alexandre, 2009].

Niveau symbolique : Ce niveau correspond au plus haut niveau d'abstraction. Les informations fusionnées sont en général très hétérogènes, et correspondent à des zones d'observation, des objets ou des instants différents. La fusion permet alors

d'avoir une vision beaucoup plus globale de la situation. A ce niveau, beaucoup d'informations a priori sont généralement utilisées. On parle parfois de niveau décisionnel [Alexandre, 2009].

Le choix du niveau de fusion dépend des capteurs utilisés. Si les capteurs sont de même nature, la fusion au niveau signal est possible. Pour tenir compte de toutes les informations. Quand les capteurs sont très différents, la fusion au niveau symbolique est plus adaptée. La fusion au niveau attribut est utilisée lorsque les attributs trouvés par le traitement de chaque capteur peuvent être convenablement associés.

L'importance de la vision pour les systèmes d'aide à la conduite

Il existe aujourd'hui une panoplie de capteurs utilisés par les systèmes d'aide à la conduite. Deux catégories de capteurs sont distinguées : Ceux qui décrivent l'état du véhicule (capteurs proprioceptifs) et ceux qui se chargent de percevoir l'environnement (capteurs extéroceptifs). Les capteurs proprioceptifs se chargent de mesurer un attribut en fonction de leur propre état. Cet attribut peut être l'accélération, l'orientation ou la vitesse de l'objet sur lequel ils sont montés. Citons par exemple, les capteurs d'inclinaison, les accéléromètres et les odomètres. La deuxième classe de capteurs, ceux qui se chargent de percevoir l'environnement, sont capables de mesurer un attribut d'un objet externe présent dans la scène. Les caméras qui fonctionnent dans le spectre visible ou infrarouge, les sonars, les lidars et les radars sont des exemples de capteurs extéroceptifs. Le choix de capteurs repose sur ses propriétés ainsi que sur les objectifs visés par l'application.

La vision est sans doute le sens le plus important pour la conduite automobile. L'information visuelle est primordiale pour le conducteur. Elle lui permet de percevoir l'environnement pour pouvoir accomplir les tâches de conduite normalement. La caméra est un capteur passif ayant une fonction similaire à la vision chez l'être humain. Elle permet de restituer l'environnement sous forme d'images ou d'une vidéo. Les images capturées par les caméras visibles (en couleurs ou en niveaux de gris) sont d'un côté très riches en contenu, de l'autre, difficiles à interpréter. Ce sont les raisons pour lesquelles on remarque d'énormes efforts de la recherche scientifique pour l'exploitation de la caméra comme capteur. Il existe dans la littérature un très grand nombre de travaux et d'algorithmes proposés pour la détection et le suivi d'objets dans les images. De plus, avec l'avancée technologique, les caméras sont devenues peu coûteuses et faciles à installer sur les véhicules.

2. Problème de vision face aux scènes à grande plage dynamique.

En photographie, la plage dynamique d'une scène naturelle ou d'une photo se réfère à l'intervalle des valeurs d'intensités lumineuses qui les caractérise. Une scène à grande plage dynamique est une scène qui contient une grande différence entre les plus fortes et les plus faibles intensités lumineuses. Avec des appareils photos ordinaires, l'image numérique obtenue est composée de pixels représentés sur 256 valeurs d'intensité. Ce qui fait que la plage dynamique de cette photo se résume sur l'intervalle $[0, 255]$ (photo à petite plage dynamique). Or, le monde naturel contient un nombre impressionnant de rayonnements lumineux différents (des millions).

Face à une scène à grande plage dynamique les appareils photo sont incapables de restituer convenablement toutes les différentes valeurs des rayonnements lumineux qui s'abattent sur sa lentille. Certaines parties de la photographie seront sous exposées, ou bien surexposées. Face à cette situation, le photographe a la laborieuse tâche de choisir les meilleurs paramètres d'exposition de l'appareil pour capturer la plage dynamique qui l'intéresse. Pour obtenir la photographie sur la figure1 à gauche, l'appareil photo est réglé d'une telle manière à bien capturer le ciel. La lumière du soleil étant très forte, le temps d'exposition du capteur de la caméra à la lumière doit être très bref pour ne pas saturer la photo. Mais en choisissant ces paramètres, le reste de la scène sera sous exposé, résultant à des zones très sombres dans l'image. Ce problème est connu aussi par l'effet du contre-jour. Afin de voir les détails assombris de la scène, on doit utiliser un temps d'exposition plus lent. Ainsi, on remarque la saturation de la région du ciel (figure1 à droite). Ce genre de scènes qui comporte à la fois de forts rayonnements lumineux et en même temps des zones faiblement illuminées s'appelle : scènes à grande plage dynamique. Quels que soient les paramètres choisis (Temps d'exposition, ISO, filtres,..., etc.), on ne pourra jamais rassembler tous les détails de la scène dans la même capture. Ce qui induit une perte considérable d'informations.



Figure 2-1: Scène à grande plage dynamique. Impossibilité de capturer tous les détails de la scène dans une seule photographie.

La solution à ce problème consiste à utiliser des algorithmes d'analyse et de traitement d'images appliqués sur plusieurs photographies différemment exposées. Des techniques de photographies computationnelles existent pour créer des images qui vont au-delà des capacités d'un système d'imagerie ordinaire, tel que montré dans la figure 2. Quelques-unes sont maintenant directement incorporées dans les appareils photos. Nous allons aborder dans la deuxième partie de ce mémoire les approches les plus importantes parmi celles qui existent dans la littérature.

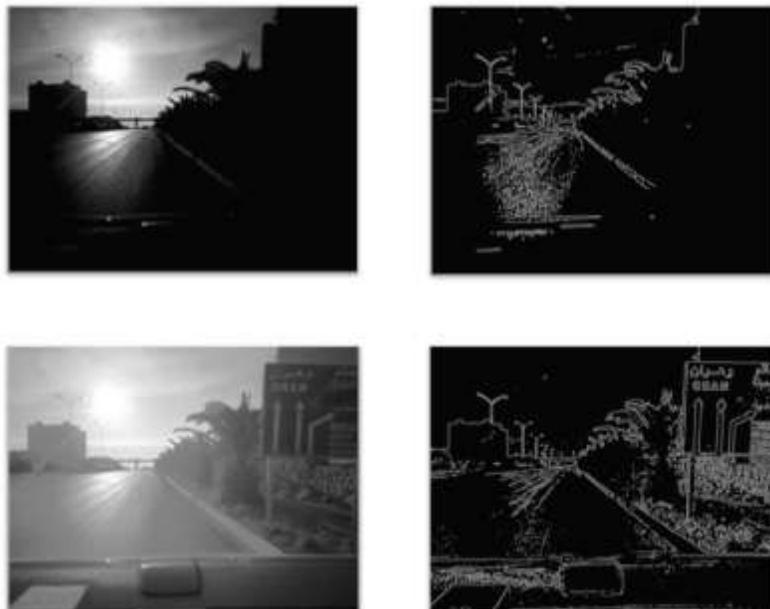


Figure 2-2 Acquisition face à une scène à grande plage dynamique. (Haut) : Capture automatique de la caméra. (Bas) Capture avec la méthode de fusion d'expositions proposée dans le chapitre 3.

3. La reconnaissance des voies routières et la transformée de Radon

Parmi les composantes fondamentales d'une scène routière qu'un système d'aide à la conduite doit prendre en considération, les voies routières (marquages des voies routières au sol). C'est une information nécessaire pour le bon déroulement de l'activité de conduite, car cette information veille à ce que chaque usager de la route garde sa voie de circulation. La baisse de vigilance du conducteur ou de mauvaises conditions de vision peuvent engendrer une sortie de voie qui peut aboutir à un drame. D'ailleurs, l'assistance au changement de voies de circulation est l'un des systèmes avancés d'aide à la conduite les plus prometteurs, et qui sans doute sollicite de plus en plus d'efforts de la part des chercheurs. Plusieurs systèmes avancés d'aide à la conduite, ayant des buts différents reposent d'une manière ou d'une autre sur l'information des voies routières [Nooij , 2011].

Dans la majorité des scènes naturelles, on retrouve des structures linéaires qui se transforment en lignes droites dans les images. Ces lignes droites sont des caractéristiques intéressantes qui jouent un rôle important dans la reconnaissance d'objets et la compréhension d'images. La transformée de Radon est une technique très utilisée pour la détection de ces caractéristiques particulières. Cette solution suppose que les lignes sont minces et formées par des pixels colinéaires. Cependant, les lignes dans une image peuvent être formées par des bandes de pixels pouvant avoir différentes épaisseurs. Cette épaisseur peut dépendre de la résolution de l'image, la nature des lignes, la perspective. Dans les images binaires, les pixels formant une ligne ne sont pas tous colinéaires. La Figure3 montre l'espace de recherche ainsi que la ligne détectée après l'application de la transformée de Radon sur des lignes épaisses ou bruitées. Nous avons en face un problème de sélection de piques. La zone rouge dans l'espace de recherche représente les lignes ayants de forts accumulateurs. Il y a une multitude de choix pour l'algorithme de la transformée de Radon. Ceci est dû au fait que les lignes sont épaisses. L'algorithme retourne automatiquement celle qui détient le plus grand accumulateur, souvent la diagonale, qui ne représente pas toujours le meilleur choix. Ceci est une erreur

de précision fréquente de l'application de la transformée de Radon à la détection de lignes dans les images.

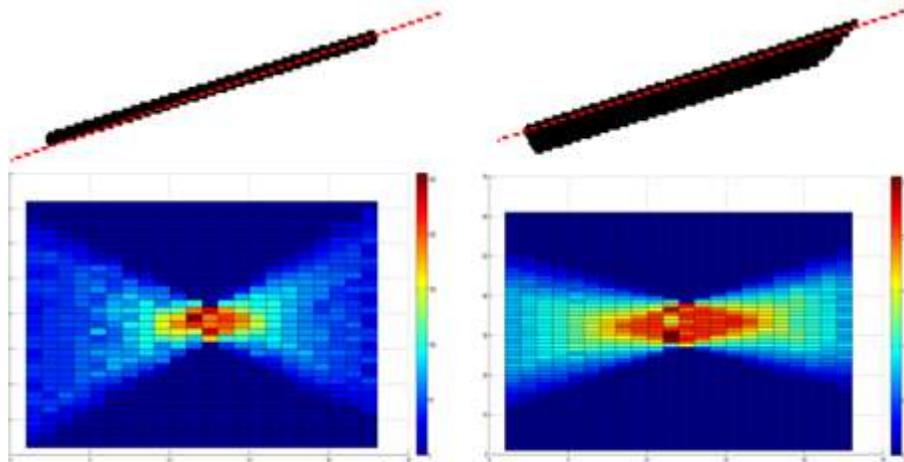


Figure 3: Comportement de la Transformée de Radon appliquées à des lignes épaisses.

4. Contributions

Nos efforts de recherche se sont concentrés sur deux problématiques différentes dont les solutions participent à la réalisation de systèmes d'aide à la conduite. Concernant la première problématique, celle de la vision face aux scènes à grande plage dynamique expliquée en haut, les solutions existantes sont gourmandes en temps d'exécution. Visant une solution dédiée à des systèmes d'aide à la conduite, notre objectif est de développer une technique de fusion d'expositions qui soit la plus légère possible afin de satisfaire la contrainte de temps réel. Celle-ci doit bien entendu délivrer des images de qualité. Dans ce sens, nous avons proposé trois méthodes originales différentes. La première méthode est basée sur une phase de segmentation rapide d'histogrammes d'images [Moumene , 2014a]. La deuxième ainsi que la troisième méthode sont basées sur un apprentissage supervisé [Moumene , 2014b], [Moumene , 2017].

La deuxième problématique concerne l'imprécision de la transformée de Radon quand elle est utilisée pour détecter des lignes dans les images. Etant l'approche la plus utilisée pour la détection des marquages routiers, elle génère souvent une erreur de précision importante car les lignes à détecter sont épaisses. Nous démontrons que

l'utilisation d'une transformée de Radon Gaussienne est plus adéquate pour la détection de ce genre de structures linéaires.

5. Organisation du mémoire

Cette thèse est organisée en deux parties principales:

Partie II : État de l'art

Le premier chapitre consiste en une introduction au problème de la vision face aux scènes à grande plage dynamique et un état de l'art expliquant des solutions qu'on retrouve dans la littérature.

Le deuxième chapitre est dédié à l'état de l'art sur la détection des voies routières sur les images. Nous y présentons une introduction ainsi que les différents modules qui constituent un système de détection des marquages routiers.

Partie III : Contributions

Le chapitre trois est consacré à la première méthode que nous proposons pour la fusion des expositions. Celle-ci est basée sur la segmentation rapide d'histogrammes.

Dans le chapitre quatre nous introduisons les deux autres techniques originales pour la fusion d'expositions. Celles-ci sont basées sur l'apprentissage supervisé.

Dans le chapitre cinq nous introduisons l'utilisation d'une généralisation de la transformée de Radon, une variante de la transformée de Radon basée sur la fonction Gaussienne qui produit plus de précision pour la détection de lignes en général.

Les résultats, les comparaisons ainsi que les conclusions sont présentés pour chacune des méthodes dans leurs chapitres respectifs. La fin du mémoire est dédiée à une conclusion générale ainsi que nos perspectives de recherche.

II - Etat de l'art

Partie II

Chapitre 1 : La vision face aux scènes à grande plage dynamique

1 La vision face aux scènes à grande plage dynamique

Ce chapitre introduit la problématique de la capture d'images face à des scènes de grande plage dynamique et présente les solutions principales connues de la littérature. Les sections 1.1 et 1.2 expliquent la notion de grande plage dynamique et la limite du capteur visuel face à ce phénomène. La section 1.3 présente les deux approches principales de photographie computationnelle qui tentent de résoudre le problème de la capture d'images face à des scènes de grande plage dynamique.

1.1 Introduction

En photographie, la plage dynamique d'une scène naturelle ou d'une photo se réfère à l'intervalle des valeurs d'intensités lumineuses qui les caractérise. Avec des appareils photos ordinaires, l'image numérique obtenue est composée de pixels représentés sur 256 valeurs d'intensité. La plage dynamique de cette photo se résume donc en l'intervalle $[0; 255]$. Le monde naturel contient un nombre impressionnant de rayonnements lumineux. L'intervalle qui regroupe toutes ces valeurs d'intensités devrait être très grand. La plage dynamique des appareils photo est incapable de restituer convenablement toutes les différentes valeurs des rayonnements lumineux qui s'abattent sur sa lentille. Si la scène à photographier contient une petite plage dynamique, le problème ne se pose pas, car toutes les intensités seront capturées. Mais si la scène contient un grand intervalle d'intensités, certaines parties de la photographie seront sous ou surexposées. Le photographe a donc la laborieuse tâche de choisir les meilleurs paramètres de l'appareil (exposition) pour capturer la plage dynamique qui l'intéresse. Ceci n'évite quand même pas l'apparition de zones sombres ou saturées dans la photographie. Ce problème est connu aussi par l'effet du contre-jour. Avec n'importe quelle valeur d'exposition choisie, l'appareil photo ne pourra jamais joindre les deux bouts dans un seul intervalle LDR.

1.2 La luminance et la plage dynamique en photographie

La luminance est la mesure de la lumière rayonnante d'une source. Elle est perçue par le spectateur humain comme l'éclat d'une source lumineuse et elle est mesurée en candela par mètre carré (cd/m^2). La luminance peut être mesurée par un luminance mètre comme celui montré sur la figure1-1.



Figure 1-1: Mesurer la lumière avec un luminance mètre

On retrouve dans le monde physique des lumières de différentes intensités allant de la lumière des étoiles ($10^{-4} cd/m^2$) jusqu'à celle du soleil ($10^6 cd/m^2$) comme illustré dans la figure1-2. La plage dynamique en photographie, nommée aussi parfois contraste de la scène représente l'ensemble des valeurs d'intensité lumineuse que l'on retrouve sur cette scène. C'est le ratio entre la plus grande et la plus petite valeur de luminance. Une scène dite à grande plage dynamique est une scène qui contient une différence significative entre les valeurs de luminance des zones sombres et claires.

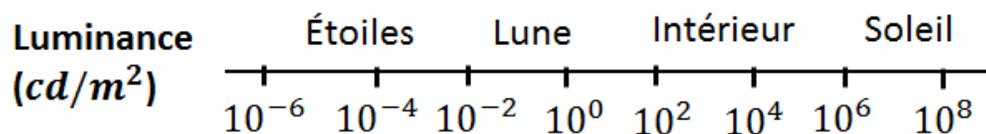


Figure 1-2: Les différentes intensités lumineuses retrouvées dans une scène réelle.

1.3 Une limite du capteur visuel

Les valeurs des pixels d'une image acquise par une caméra numérique sont proportionnelles à la luminance perçue par la scène. La lumière est un signal électromagnétique qui lorsqu'il s'abat sur le capteur photographique, en passant par la lentille de la caméra, produit des électrons (voir figure1-3). Chaque pixel d'une image reflète la quantité d'électrons emmagasinée dans un point du capteur photographique.

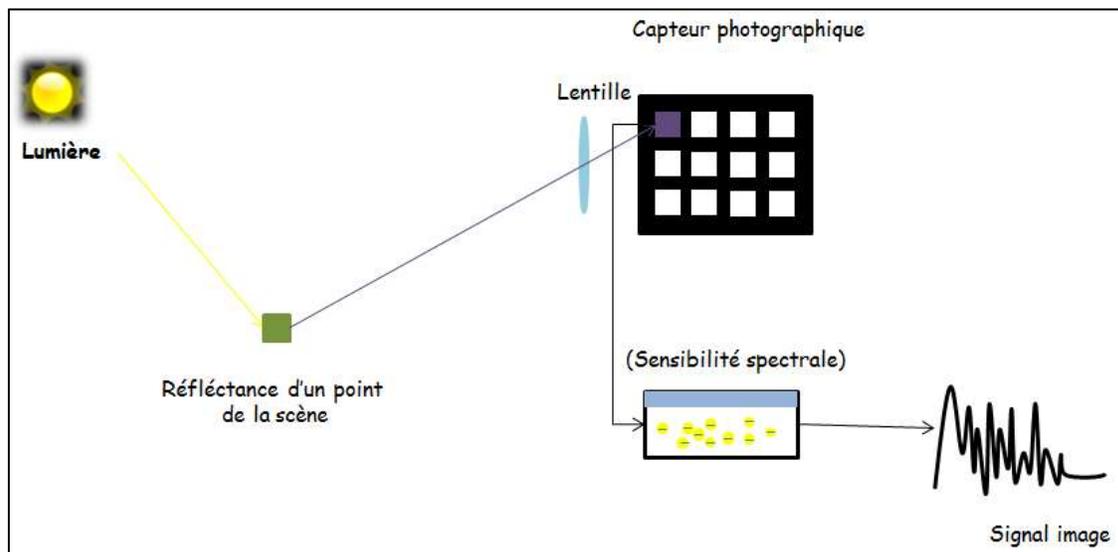


Figure 1-3: Formation du signal image.

Le système visuel humain traite la lumière entrante en changeant dynamiquement la contraction ou l'expansion de la pupille. Les caméras aussi sont équipées de fonctions permettant de contrôler automatiquement la quantité de lumière qui s'abat sur le capteur photographique. Face à une lumière forte, l'algorithme d'exposition automatique de la caméra fait en sorte que le temps d'exposition du capteur photographique soit très court. Ceci en jouant sur l'ouverture du diaphragme et la vitesse de l'obturateur. De cette manière le capteur photographique ne sera pas submergé par les électrons et donc ne fournira pas une image saturée comme on peut le voir dans la photographie de droite dans la figure1-4. Face à une lumière faible, l'algorithme d'exposition automatique de la caméra augmente le temps d'exposition du capteur photographique afin que les électrons collectés puissent être significatifs. Dans une scène telle que montrée par la figure1-4, on

retrouve des régions sombres ainsi que des régions très lumineuses. On appelle ce genre de scènes, scènes à grande plage dynamique.

Cependant, une caméra conventionnelle ne contient qu'une plage dynamique de 256 (8 bits par pixel), qui est pour la plupart du temps insuffisante pour représenter la grande plage dynamique d'une scène naturelle. Une scène contenant des régions sombres ainsi que des régions très lumineuses peut provoquer une perte significative de détails dans la photo capturée, quels que soient les paramètres d'exposition choisis. Choisir certains paramètres d'exposition permet de capturer une petite partie de la grande plage dynamique. En choisissant n'importe quels paramètres d'exposition, des zones sous exposées et/ou surexposées vont apparaître sur les photos.



Figure 1-4: Photographies capturées face à une scène à grande plage dynamique.

1.4 Photographie computationnelle

La photographie computationnelle est une branche qui utilise le calcul numérique au lieu de processus optiques pour surmonter les limites de la caméra traditionnelle. Des exemples d'algorithmes de photographie computationnelle sont la génération d'images à grande plage dynamique ou bien la fusion d'expositions. Ce sont deux approches utilisées pour palier au problème de scènes à grande plage dynamique qui a été cité précédemment.

La première approche, l'imagerie à grande plage dynamique, essaie de capturer la lumière aussi fidèlement que possible afin de préserver une information maximale face aux scènes à grande plage dynamique. Grâce à l'estimation de la fonction de réponse de la caméra conventionnelle, une image à grande plage dynamique est fabriquée avec la capture de plusieurs expositions. Ensuite un algorithme de reproduction de tons est utilisé pour créer une image à petite plage dynamique affichable.

La fusion d'expositions directe consiste plutôt à fusionner directement plusieurs expositions en une seule image qui contient tous les détails, sans passer par une image à grande plage dynamique. La figure1-5 résume les étapes importantes des deux approches de photographie computationnelle.

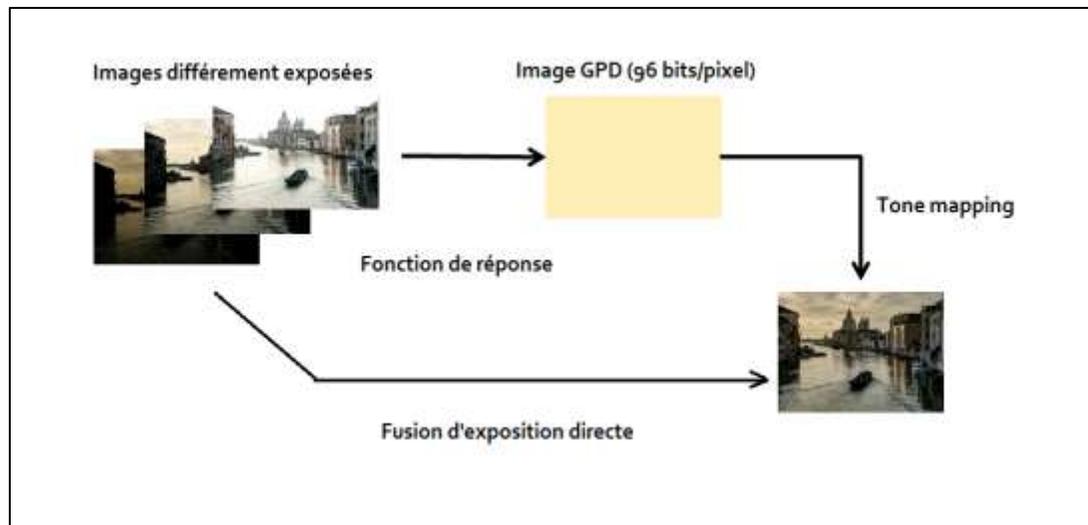


Figure 1-5: Imagerie à grande plage dynamique et fusion d'exposition directe.

1.4.1 Imagerie à grande plage dynamique

L'imagerie à grande plage dynamique (GPD) essaie de résoudre le problème rencontré par les caméras conventionnelles en produisant des images à grandes plages dynamique regroupant une large plage de lumière entrante tel que montré par la figure suivante. La méthode la plus couramment utilisée pour capturer des images GPD est de prendre plusieurs images à différentes expositions de la même scène pour capturer les détails des zones les plus sombres jusqu'aux zones les plus brillantes. Ensuite, il faut retrouver pour chaque pixel de chaque image l'irradiance (le rayonnement), sachant que le temps d'exposition de chaque image est connu ainsi que la fonction de réponse de la caméra [Mann , 1995] [Debevec , 1997](figure1-6).

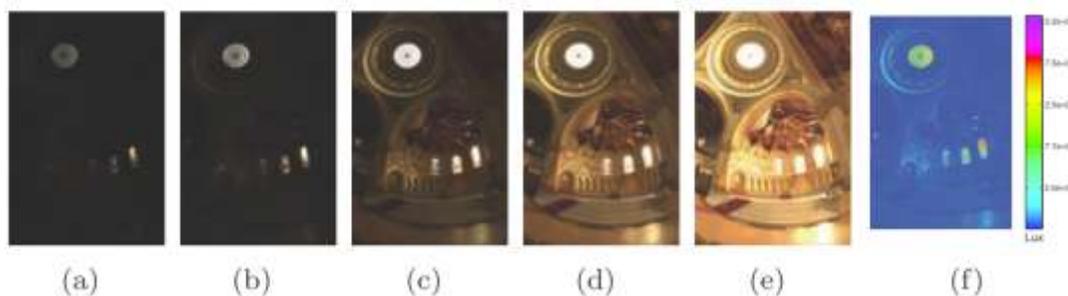


Figure 1-6: Un exemple de capture d'expositions et obtention d'une image GPD. Les expositions sont prises à différents temps d'exposition. (a) 1/250 sec. (b) 1/30 sec. (c) 1/4 sec. (d) 2 sec. (e) 8sec. (f) Est un rendu de la luminance de l'image GPD obtenue. [Debevec , 1997]



1.4.1.a. Estimation de la fonction de réponse de la caméra

La première étape à la formation d'images à grande plage dynamique est l'estimation de la fonction de réponse de la caméra. Si la caméra a une réponse linéaire, les valeurs de radiance stockées dans chaque exposition pour chaque canal de couleur peuvent être combinées pour récupérer l'irradiance E , comme suit :

$$E(x) = \frac{\sum_{i=1}^{N_e} \frac{1}{\Delta t_i} w(I_i(x)) I_i(x)}{\sum_{i=1}^{N_e} w(I_i(x))} \dots \dots \dots (1)$$

Avec I_i est l'image de la i ème exposition, Δt_i est le temps d'exposition pour I_i , N_e est le nombre d'images à différentes expositions, et $W(I_i(x))$ est une fonction de pondération utilisée pour réduire le bruit. Les valeurs centrales des expositions sont les plus crédibles car les grandes valeurs d'intensité peuvent représenter des régions saturées, quant aux petites valeurs, elles sont saturées ainsi qu'influencées par le bruit.

Malheureusement, les caméras numériques n'ont pas de réponse linéaire mais une fonction plus générale f . Quand on photographie une scène et on digitalise le contenu pour obtenir une image numérique (tableau à deux dimensions de valeurs qui représentent la luminosité), la valeur d'un pixel retrouvée est rarement la mesure réelle du rayonnement reçu par la lentille. Il y a généralement une correspondance non linéaire qui détermine comment un rayonnement de la scène devient une valeur d'un pixel. La fonction de réponse de la caméra (FRC) est la fonction qui définit cette correspondance. Cette fonction non linéaire est difficile à connaître à l'avance car c'est la composition de plusieurs opérations non linéaires qui interviennent dans le processus de photographie [Richard Szeliski, 2010] (figure1-7). En premier lieu l'ouverture de la camera et la vitesse de l'obturateur (l'exposition). Puis la conversion analogique numérique qui applique souvent un gain contrôlé par le paramètre ISO de la camera. Si l'image est sauvegardée sous le format JPEG elle subit plusieurs autres traitements qui peuvent influencer d'une manière non linéaire les valeurs des intensités. (Les traitements effectués

durant tout le processus de la capture de l'image numérique sont expliqués en détails dans [Debevec , 1997] et sont résumés par la Figure 1-7.

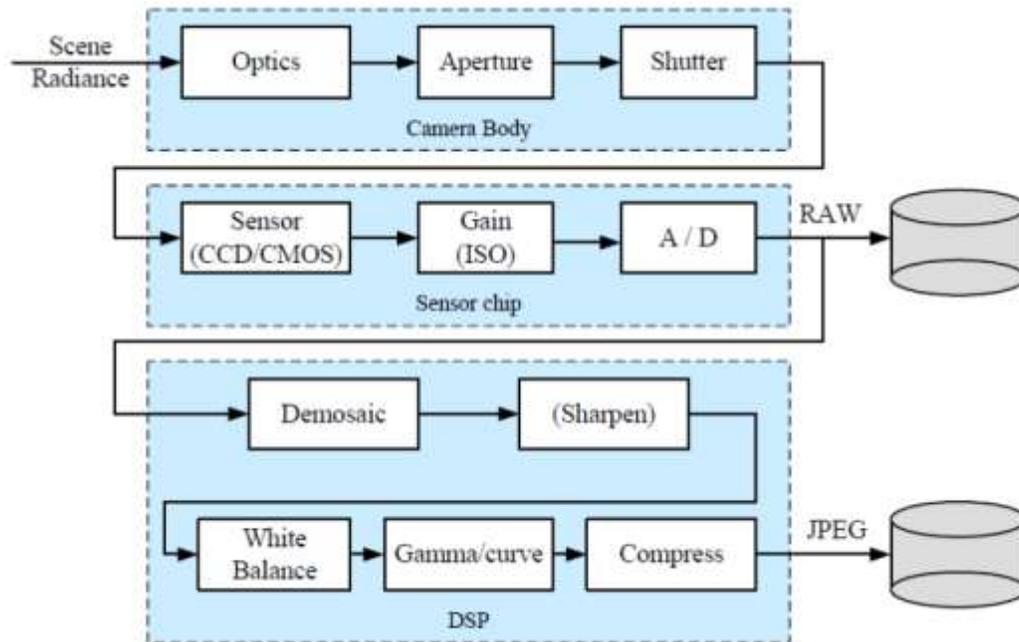


Figure 1-7: Diagramme montrant les différentes sources de bruit ainsi que les près traitement typiques [Richard Szeliski, 2010].

Mann et Picard [Mann , 1995] ont proposé une méthode simple pour calculer f . Elle consiste à ajuster un modèle sur les valeurs de pixels se trouvant dans les expositions. Le modèle serait $(x) = ax^\gamma + b$. Ce modèle paramétrique reste très limité et ne supporte pas la plupart des fonctions de réponse des caméras.

[Debevec , 1997] ont proposé une méthode simple pour estimer la fonction de réponse de la caméra. Les différentes méthodes seront présentées pour les images en niveau de gris. La valeur d'un pixel dans une image est donnée par l'application de la fonctions de réponse de la caméra à l'irradiance avec en facteur le temps d'exposition :

$$I(x) = f(E(x) \Delta t_i) \dots \dots (2)$$

Après le réarrangement des termes et l'introduction du logarithme aux deux cotés on obtient :

$$\ln (f^{-1}(I(x))) = \ln E_i(x) + \ln \Delta t_i (3)$$

En supposant que f est une fonction lisse et monotone croissante, f et E peuvent être calculées par minimisation de l'erreur dérivée de l'équation (3) en utilisant les pixels des différentes expositions :

$$O = \sum_{i=1}^{N_e} \sum_{j=1}^M (w(I_i(x_j)) [g(I_i(x_j)) - \ln E(x_j) - \ln \Delta t_i])^2 + \lambda \sum_{x=T_{min}+1}^{T_{max}-1} (w(x)g''(x))^2 (4)$$

Avec $g = \ln f^{-1}$ est la fonction inverse de la fonctions de réponse de la caméra, M est le nombre de pixels utilisés pour la minimisation, T_{max} et T_{min} sont respectivement le maximum et le minimum des valeurs entières retrouvées dans les expositions I_i . La deuxième partie de l'équation (4) est un terme de lissage utilisé pour réduire le bruit. W est définie comme :

$$w(x) = \begin{cases} x - T_{min} & \text{if } x \leq \frac{1}{2} (T_{max} + T_{min}) \\ T_{max} - x & \text{si non} \end{cases} (5)$$

Il est à noter que la minimisation est effectuée seulement sur un sous ensemble de M , car il est très lourd en terme de calcul de le faire sur l'ensemble des pixels. Le sous ensemble de pixels est pris des différentes régions de l'image.

Mitsunaga et Nayar [Mitsunaga , 1999] ont amélioré l'algorithme de [Debevec , 1997] avec une méthode plus robuste basée sur une représentation polynomiale de f . Ils affirment que toute fonction de réponse peut être modélisée en utilisant un polynôme d'ordre supérieur:

$$I_i(x) = f(E(x)\Delta t_i) = \sum_{k=0}^P c_k (E(x)\Delta t_i)^k \dots \dots \dots (6)$$

A ce stade, le processus de calibration peut être réduit à l'estimation de l'ordre P du polynomial et des coefficients C_j . En prenant deux images d'une scène avec deux temps d'exposition différents Δt_1 et Δt_2 , le rapport R peut être écrit comme :

$$R = \frac{\Delta t_1}{\Delta t_2} = \frac{I_1(x)}{I_2(x)} \dots \dots \dots (7)$$

La mesure de luminosité $I_i(x)$ produite par un système d'imagerie est liée à la radiance de la scène $E(x \Delta t_i)$ à l'instant i via la fonction de réponse $I_i(x) = f(E(x \Delta t_i))$. De ce fait, $I_i(x)$ peut être écrite $E(x \Delta t_i) = g(I_i(x))$ avec $g = f^{-1}$. Puisque la fonction de réponse d'un système d'imagerie est liée au rapport d'exposition, l'équation (7) peut être réécrite comme :

$$R_{1,2}(x) = \frac{I_1(x)}{I_2(x)} = \frac{\sum_{k=0}^P c_k I_1(x)^k}{\sum_{k=0}^P c_k I_2(x)^k} \dots \dots \dots (8)$$

Les images sont ordonnées de tel sorte que $\Delta t_1 < \Delta t_2$ pour que $R \in (0, 1)$. Le nombre de paires f-R qui satisfont l'équation 8 est infini. Cette ambiguïté est atténuée par l'utilisation du modèle polynomial. La fonction de réponse peut être récupérée en formulant une fonction d'erreur telle que :

$$\varepsilon = \sum_{i=1}^{N_e} \sum_{j=1}^M \left(\sum_{k=0}^P c_k I_i(x_j)^k - R_{i,i+1}(x_j) \sum_{k=0}^P c_k I_{i+1}(x_j)^k \right)^2 \dots \dots (9)$$

Toutes les mesures peuvent être normalisées de façon que $I_i(x)$ soient dans $[0, 1]$. Les coefficients de la fonction de réponse sont déterminés par la résolution d'un système linéaire :

$$\frac{d\varepsilon}{dc_k} = 0$$

Robertson et al [Robertson , 1999, Robertson , 2003] ont proposé une méthode qui estime la fonction de réponse ainsi que l'irradiance $E(x)$ par l'utilisation de l'approche de maximum de vraisemblance, où la fonction objective à minimiser est :

$$O(I, E) = \sum_{i=0}^{N_e} \sum_{j=0}^M w_{i,j} (I_i(x_i) - \Delta t_i E(x_j))^2 \dots \dots \dots (10)$$

Avec w le poids défini par une fonction gaussienne qui représente le bruit dans le système d'imagerie utilisé pour capturer les images. Noter que toutes les méthodes présentées pour récupérer la fonction de réponse peuvent être étendues à des images en couleur en appliquant la méthode séparément sur chaque canal de couleur.

1.4.1.b. Construction de l'image à grande plage dynamique

Une fois que la fonction de réponse g récupérée, elle peut être utilisée pour convertir les valeurs de pixel en valeurs de radiance, en supposant que l'exposition Δt_i est connue. La fonction g peut maintenant être utilisée pour déterminer les valeurs de radiance de n'importe quel image acquise par la camera associée à cette fonction. De l'équation (3) on obtient :

$$\ln E_j = g(I(x)) - \ln \Delta t_i \dots (11)$$

Pour plus de robustesse, la radiance est calculée à partir de toutes les expositions disponibles sur un pixel. Pour y arriver, une fonction de pondération est utilisée :

$$\ln E_j = \frac{\sum_{i=1}^{N_e} w(I_i(x))(g(I_i(x)) - \ln \Delta t_i)}{\sum_{i=1}^{N_e} w(I_i(x))} \dots (12)$$

1.4.1.c. Reproduction de tons

Les images à grande plage dynamique ne peuvent pas être visualisées dans des écrans conventionnels à 8bit/canal. Le processus de compression du contenu GPD en images affichables à petite plage dynamique est appelé la reproduction de tons. Le défi de la reproduction des tons consiste à compresser la plage dynamique de telle sorte que l'information rejetée ne soit pas perceptuellement pertinente. La sensibilité du système visuel humain à la lumière dépasse de loin la sensibilité des capteurs visuels et donc l'opération de reproduction de tons finira par dégrader la qualité de l'image GPD originale. Cette dégradation doit être contrôlée pour maximiser la qualité de l'image résultante.

Les opérateurs de reproduction de tons sont divisés en deux groupes, opérateurs globaux ou locaux. Les opérateurs globaux, aussi connus comme étant spatialement invariables [Ledda , 2004][Wang , 2007], utilisent les mêmes fonctions pour tous les pixels sans

prendre en considération les valeurs des pixels voisins. Leurs avantages sont la simplicité et la rapidité d'exécution. Cependant, l'utilisation d'opérateurs globaux peut causer une perte de détails. Pour résoudre ce problème, les opérateurs de reproduction de tons locaux sont développés. L'opérateur local utilise les informations de voisinage du pixel pour définir la fonction de transformation adéquate [Ashikhmin , 2002] [Fattal , 2002]. La fonction de transformation utilisée change localement dépendamment de la position du pixel pour produire à la fin le plus de contraste. Le problème des opérateurs locaux sont la grande consommation de temps de calcul et la difficulté de déterminer la taille du voisinage.

L'opérateur de reproduction de tons le plus simple utilisé pour transformer l'image GPD en une image affichable est linéaire. Toute la plage dynamique est transformée d'une manière linéaire sans se soucier de la distribution des détails dans la plage dynamique. Cependant, elle ne préserve pas les détails de l'image GPD dans l'image résultante. La fonction linéaire est donnée par :

$$L_d(x, y) = \frac{L(x,y)-L_{min}}{L_{max}-L_{min}} \dots\dots(13)$$

Avec $L(x, y)$ la radiance de la scène et $L_d(x, y)$ est la luminance affichable. L_{max} et L_{min} sont respectivement la radiance maximale et la radiance minimale trouvées dans la scène.

La transformation logarithmique est une autre manière simple pour la compression de la grande plage dynamique. L'opérateur n'effectue pas la compression de façon uniforme pour toute la plage dynamique. L'information dans la carte GPD est transférée avec un minimum de perte de détails. La compression logarithmique de la plage dynamique est donnée par :

$$L_d(x, y) = \frac{\log_{10}(L(x,y)+1)}{\log_{10}(L_{max}+1)} \dots\dots(14)$$

Les opérateurs de reproduction de tons globaux sont souvent des fonctions logarithmiques qui opèrent sur chaque pixel individuellement. Ils ne prennent pas en considération la relation spatiale entre pixels et donc tendent à générer du flou dans les contours de l'image résultante. L'un des opérateurs de reproduction de tons importants a été introduit par [Tumblin , 1993] :

$$n = \left[\frac{L_w^{\frac{\alpha(w)}{\alpha(d)}}}{L_{d_max}} * 10^{\frac{\beta(w)-\beta(d)}{\alpha(d)}} - \frac{1}{C_{max}} \right]^{\frac{1}{\gamma}} \dots (15)$$

$$\alpha(L) = 0.4 * E(\log(L)) + 2.92 \dots (16)$$

$$\beta(L) = -0.4 * (E(\log(L)))^2 - 2.58 * E(\log(L)) + 2.02 \dots (17)$$

Avec :

γ = Gamma de l'affichage.

C_{max} = Le contraste maximal de l'affichage.

d = Luminance de l'affichage (constant).

w = Luminance de la scène.

L_w = Luminance de l'arrière-plan de la scène.

L_{d_max} = Luminescence maximale de l'affichage.

$E(\log(L))$ = Espérance mathématique de Log(L).

Les équations mentionnées en haut représentent l'opérateur de reproduction de tons de [Tumblin , 1993]. Cet opérateur a été conçu pour des images en niveau de gris seulement il n'est donc pas utilisable pour des images en couleur. Aussi, cet opérateur global de reproduction de tons requiert des paramètres spécifiques d'affichage qui pourraient ne pas être disponibles dans de nombreuses situations.

Plus tard, [Ward , 1997] ont introduit un opérateur de reproduction de tons basé sur l'ajustement d'histogrammes. Une nouvelle technique d'ajustement d'histogramme a été proposée qui est basée sur l'adaptation locale de population. L'opérateur incorpore des modèles de sensibilité de contraste chez l'humain, l'éblouissement, l'acuité spatiale, et la sensibilité aux couleurs. L'opérateur basé sur l'ajustement d'histogramme est partiellement local car il choisit ses paramètres en se basant sur des attributs locaux.

[Schlick , 1995] a proposé une autre fonction de reproduction de tons :

$$F(I) = \frac{I}{I + \frac{I_{max} - I}{p}} \dots \dots (18)$$

Avec :

I= Luminance de la scène.

I_{max}= Luminance maximale de la scène.

P= Contraste ajustable.

L'un des premiers opérateurs de reproduction de ton locaux a été proposé par [Chiu , 1993]. Ils tentent de compresser la plage dynamique en utilisant une version floue de l'image d'entrée :

$$L_d(x, y) = \frac{L(x,y)}{k * L_w^{blur}(x,y)} \dots \dots \dots (19)$$

Avec :

$L_d(x, y)$ = Image à petite plage dynamique résultante.

$L_w^{blur}(x, y)$ = Image qui a subi un filtrage Gaussien.

L'opérateur de reproduction de tons le plus populaire est celui de [Reinhard , 2002]. L'opérateur de Reinhard détecte d'abord combien la scène est lumineuse et produit ce qu'on appelle la clé de l'image. Par exemple, une scène capturée la nuit va avoir une petite clé comparée à une scène de jour. La clé de la scène est estimée en utilisant l'équation (20). La deuxième étape de l'opérateur consiste à utiliser l'équation (21) pour remettre l'image à une échelle appropriée. L'image à petite plage dynamique finale est calculée en utilisant l'équation (22).

$$key = \frac{1}{N} \exp\left(\sum \log(\delta + L_w(x, y))\right) \dots \dots \dots (20)$$

$$L(x, y) = \frac{\alpha}{key} L_w(x, y). \dots \dots \dots (21)$$

$$L_d(x, y) = \frac{L(x,y) \left(\frac{L(x,y)}{L_{white}}\right)}{1+L(x,y)} \dots \dots \dots (22)$$

Avec:

N= Le nombre total de pixels.

α = Paramètre ajustable.

L_{white} = Balance des blancs

1.4.2 Fusion directe des expositions

1.4.2.a. Fusion de blocs

La fusion d'exposition directe est une technique alternative à l'imagerie à grande plage dynamique expliquée dans la section 1.3.1. En fusion d'exposition, une image à petite plage dynamique affichable est produite directement à partir des expositions capturées. Les étapes d'estimation de la fonction de réponse, création de l'image GPD et la reproduction de tons sont complètement évitées.

Le travail pionnier de la fusion de l'exposition peut être considéré comme étant celui de de [Goshtasby, 2005]. Cette méthode consiste à diviser les expositions d'entrée en blocs de taille fixe et à sélectionner le bloc ayant l'entropie (mesure de qualité) la plus élevée pour l'image finale :

$$E_c = \sum_{i=0}^{255} -P_i^c \log(P_i^c) \dots \dots (23)$$

Avec P_i^c la probabilité des pixels dans l'ensemble i . Chaque image en entrée est divisée en blocs $d \times d$, qui contiennent 16 pixels au minimum. Le bloc qui a la plus grande entropie parmi tous les autres blocs des autres images en entrée est sélectionné pour l'image finale. Les blocs sélectionnés sont fusionnés ensemble pour éviter les discontinuités sur les bords. Pour cela, la fonction suivante est utilisée :

$$O(x, y) = \sum_{j=1}^{n_r} \sum_{k=1}^{n_c} W_{jk}(x, y) I_{jk}(x, y) \dots \dots (24)$$

Avec $I_{jk}(x, y)$ est la valeur du pixel ayant l'index (x, y) dans les blocs (j, k) . $W_{jk}(x, y)$ est le poids de l'index (x, y) dans les blocs d'images (j, k) . La valeur d'intensité est la somme pondérée des pixels ayant le même index (x, y) dans les expositions. Le poids du bloc sélectionné est de 0.8 et celui des autres blocs a la somme de 0.2.

La fonction de pondération attribue un poids maximum aux centres des blocs sélectionnés. Ce poids est décroissant relativement à la distance du centre. Une fonction Gaussienne est utilisée pour cela :

$$W_{jk}(x, y) = \frac{G_{jk}(x, y)}{\sum_{m=1}^{n_r} \sum_{n=1}^{n_c} G_{mn}(x, y)} \dots \dots (25)$$

n_r et n_c sont les nombres de blocs verticaux et horizontaux. $G_{jk}(x, y)$ est définie comme suit :

$$G_{jk}(x, y) = \exp\left\{-\frac{(x - x_{jk})^2 + (y - y_{jk})^2}{2\sigma^2}\right\} \dots \dots (26)$$

où (x_{jk}, y_{jk}) sont les coordonnées du centre du bloc jk . σ est l'écart type de la Gaussienne ou bien la largeur de la fonction de fusion.

1.4.2.b. Algorithme de Tom Mertens

Il s'agit de l'algorithme de fusion d'exposition le plus utilisé dans les logiciels de traitement d'images ainsi que le plus cité dans la littérature. Les expositions sont fusionnées en une image à petite plage dynamique affichable en se basant sur une phase

de pondération des pixels en correspondance. Trois critères de pondération (qualité de mesures) sont utilisés : l'exposition, la saturation et le contraste [Mertens , 2007].

Pour le critère d'exposition, des poids faibles sont attribués aux zones sous exposées ou surexposées et les poids les plus forts sont attribués aux pixels ayant une intensité qui s'approchent de la moyenne. Tout cela est accompli en utilisant une Gaussienne :

$$g(i) = \exp\left(-\frac{(i - 0.5)^2}{2\sigma^2}\right) \dots \dots (27)$$

Avec i la valeur d'intensité d'une couleur et $\sigma = 0.2$. La Gaussienne est appliquée à chaque canal de couleur séparément puis leur produit est pris pour calculer le poids final. La mesure du contraste est déterminée en appliquant un filtre Laplacien aux images en niveau de gris, puis en prenant la valeur absolue de la réponse du filtre. La saturation est déterminée en prenant l'écart type dans les valeurs R,G,B de chaque pixel. Les trois mesures de qualités sont combinées par la multiplication suivante :

$$W_{i,j,k} = (C_{i,j,k})^{wc} (S_{i,j,k})^{ws} (E_{i,j,k})^{we} \dots \dots (28)$$

C, S et E sont respectivement les poids du contraste de la saturation et de l'exposition. Les valeurs w sont des exposants de pondération. i, j, k est le pixel ayant la position (i, j) dans l'image k . Par défaut, $wc = ws = we = 1$.

Les poids d'un pixel ayant la position (i, j) sont normalisés par la formule :

$$\widehat{W}_{i,j,k} = \left[\sum_{k'=1}^N W_{i,j,k'} \right]^{-1} W_{i,j,k} \dots \dots (29)$$

L'image finale est obtenue par la fusion pondérée :

$$R_{i,j} = \sum_{k=1}^N \widehat{W}_{i,j,k} I_{i,j,k} \dots \dots \dots (30)$$

Cette approche conduit à des artefacts dans l'image finale car la carte de poids varie soudainement d'une région à l'autre en raison de la différence des temps d'exposition. Le lissage de la carte des poids par des filtres gaussiens fait apparaître des halos autour des bords et l'utilisation de filtrage bilatéral a des problèmes avec le choix des paramètres optimaux. C'est pourquoi la méthode utilise une décomposition pyramidale de chaque exposition et de chaque carte de poids normalisés :

$$L\{R\}_{i,j}^l = \sum_{K=1}^N G\{\widehat{W}\}_{i,j,k}^l L\{I\}_{i,j,k}^l \dots \dots \dots (31)$$

$L\{A\}$ est la décomposition Laplacienne en pyramide de l'image A. $G\{B\}$ est la pyramide Gaussienne de l'image B. l est le niveau de la pyramide. Dans cette formule, chaque niveau de la pyramide laplacienne résultante est exprimé avec une moyenne pondérée de la pyramide laplacienne d'image d'exposition. La pyramide résultante est combinée pour obtenir l'image finale. L'opération est appliquée à chaque niveau séparément.

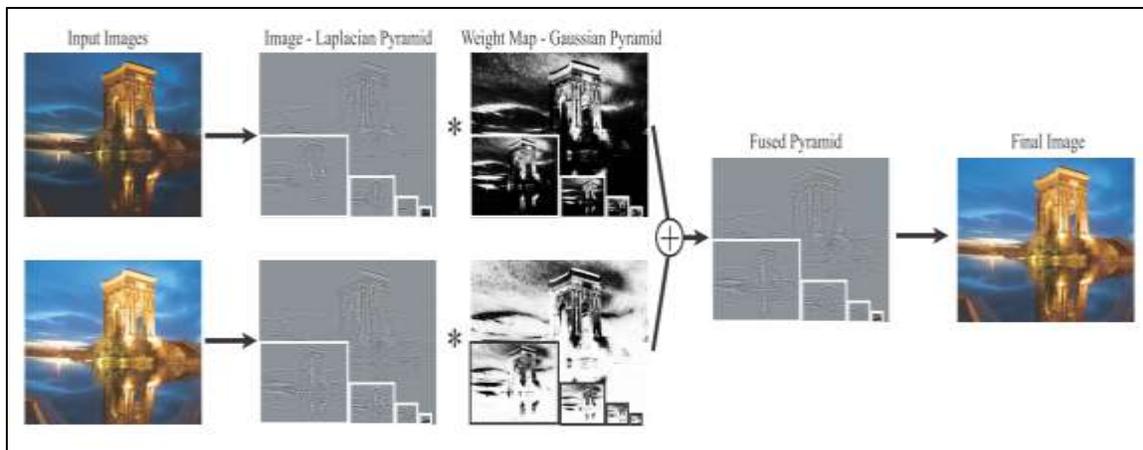


Figure 1-8: Fusion d'images différemment exposées à l'aide d'une décomposition laplacienne et d'une pyramide gaussienne des cartes de poids, qui représentent les mesures de contraste, saturation et exposition [Mertens , 2007].

1.4.2.c. Fusion d'expositions basée sur la direction du gradient

Dans une série d'expositions d'une scène à grande plage dynamique, il a été observé que la magnitude du gradient est élevée pour les régions bien exposées et elle est basse pour les régions sous exposées [Zhang , 2012]. De plus, si des objets en mouvement existent dans la scène, des changements de direction de gradient se produisent dans les images d'exposition. Par conséquent, la magnitude et la direction du gradient sont utilisées comme des métriques pour fusionner les expositions afin de produire des images bien exposées sans artefacts. Les expositions sont combinées par la formule générale :

$$H(x, y) = \sum_{i=1}^N W^i(x, y) I^i(x, y) \dots \dots (32)$$

N est le nombre d'expositions, I et W sont respectivement l'intensité et la fonction de pondération. Le résultat dépend du poids W qui est calculé par la mesure de la qualité basée sur le gradient. L'image résultante est produite avec un processus de reconstruction multi-résolution utilisant la pyramide de Laplacien.

Les gradients des images d'exposition sont calculés par la première dérivée du noyau Gaussien 2D. Finalement, le résultat du gradient est amélioré par un filtrage bilatéral

croisé [Paris, 2009]. On s'attend à ce que la visibilité d'un pixel soit élevée où l'amplitude de gradient devient plus grande puisque le gradient diminue graduellement si un détail est surexposé ou sous-exposé. Par conséquent, l'évaluation de la visibilité est développée comme suit:

$$V^i(x, y) = \frac{m^i(x, y)}{\sum_{i=1}^N m^i(x, y) + \epsilon} \dots \dots (33)$$

Avec $m^i(x, y)$ est la magnitude du gradient à la position (x, y) de la i ème exposition. ϵ est une petite valeur pour éviter la singularité. Dans des scènes statiques $V^i = W^i$ et ce réglage produit des résultats agréables dans la préservation des détails dans toutes les régions, dans toutes les images d'exposition.

1.5 Conclusion

Pour capturer des photos face à une scène à grande plage dynamique avec une moindre perte d'informations (perte de contours, couleurs, ...etc.), il faut mobiliser la plage dynamique nécessaire. Pour le faire, le photographe a besoin de plusieurs photographies différemment exposées. Une fois rassemblées, les expositions sont exploitées pour générer une image à grande plage dynamique ou bien fusionnées directement en une seule image à petite plage dynamique contenant tous les détails. La fusion d'expositions paraît être une solution assez simple et moins coûteuse en temps de calculs.

Partie II

Chapitre2 : La détection des voies routières

2 La détection des voies routières

Ce chapitre commence par une introduction présentant l'importance de la détection des voies routières pour les systèmes d'aide à la conduite ainsi que les différents capteurs utilisés pour cela. La Section 2.2 présente un état de l'art sur les différents modules de détection des marquages routiers ainsi que les méthodes utilisées dans chaque module.

2.1 Introduction

Les systèmes avancés d'aide à la conduite sont de plus en plus intégrés dans les nouveaux véhicules. Que ce soit des systèmes qui alertent le conducteur de situations dangereuses ou bien qui interviennent avec une action sur la conduite. Ces systèmes sont supposés se complexifier et s'améliorer jusqu'à aboutir à des voitures complètement autonomes. Le défi majeur dans le développement de ce genre de systèmes est le problème de perception. Il s'agit en majeure partie de la perception des marquages routiers, la perception des obstacles tels que les autres véhicules, pétons...etc, dans les différentes situations de conduite, condition d'éclairage et météorologique.

Pour se localiser sur la route, le conducteur compte sur des indices visuels spécifiques tels que la couleur, la texture ainsi que les marquages peints sur la route. Les systèmes d'aide à la conduite ou de conduite autonome sont supposés partager la route avec les conducteurs humains et devraient donc compter sur les mêmes indices de perception utilisés par l'humain. Il pourrait y avoir un moyen de communication entre le véhicule et l'infrastructure mais ça ferait l'objet d'un énorme investissement. La perception des marquages routiers via les indices visuels reste donc l'approche préférée pour la réalisation de systèmes d'aide à la conduite ou de conduite autonome.

Les systèmes de transports intelligents sont supposés détecter la route avec le nombre des voies routières, la position actuelle, la fusion des voies, la fin d'une voie, sa séparation en deux, dans un environnement urbain, rurale ou dans des autoroutes. Ce type de compréhension de la scène routière est encore complexe pour les systèmes de perception

actuels. Il existe plusieurs capteurs utilisés comme compléments pour aider la compréhension de la scène routière: la vision, la stéréoscopie, LIDAR, GPS et carte numérique ainsi que le radar :

La vision : Positionner une caméra sur l'avant du véhicule est la modalité la plus utilisée pour la perception de la route et des voies routières. Ceci est due au fait que l'information visuelle est la principale modalité utilisée par le conducteur humain, et les marquages routiers sont fait de telle sorte qu'ils restent visibles dans les différentes conditions de conduite. Les marquages routiers ont une signature dans le domaine visuel, et donc aucun système de détection des voies routières ne peut être réalisé sans le capteur visuel. En plus, la caméra est actuellement la modalité de perception la moins coûteuse.

LIDAR : Light Detection and Ranging, représente un autre capteur important pour la détection de la route et des voies routières. Il permet de donner la structure 3D de l'environnement qui entoure le véhicule. En plus, la majorité des LIDAR peuvent donner aussi l'intensité réfléchi ce qui peut être un substitut de la camera visuelle avec l'avantage d'être un capteur actif, et donc indépendant des problèmes liés à l'éclairage. Puisque le marquage routier a uniquement une information d'intensité et pas de relief 3D, la mesure d'intensité est nécessaire si on veut utiliser uniquement un LIDAR comme modalité de perception.

La vision stéréoscopique : Il s'agit d'utiliser deux cameras pour obtenir de l'information 3D. Cette technique est beaucoup moins coûteuse que l'utilisation de LIDAR. D'un autre côté, la vision stéréoscopique ne peut généralement pas atteindre la précision et la robustesse d'un LIDAR. De plus, ça pose un défi computationnel avec une probabilité d'erreur.

Positionnement GPS et cartographie routière : L'utilisation d'une base de données géographique avec la position du véhicule est devenue une activité effectués tous les jours par les automobilistes. Les systèmes de navigation commercialisés sont utilisés

partout dans le monde pour guider les conducteurs. Il est tout à fait concevable que le positionnement GPS couplé à une carte peut être utilisé pour la conduite autonome sans l'utilisation d'autres capteurs. Cette réalisation dépendra de la précision et la robustesse de la localisation GPS et l'information de la carte. Les GPS commercialisé actuellement produisent une précision de 5-10 mètres.

Radar : Le radar est un capteur incapable d'observer les marquages routiers, ni de restituer une scène 3D précise. Il est par contre utilisé pour détecter les obstacles qui obstruent les voies, et distinguer entre la route et la région non route en se basant sur la large différence de réflectivité.

La vision est le l'outil le plus important pour la détection de la route en raison du fait que les marques sont faites pour la vision humaine. Alors que le LIDAR et le positionnement par satellite sont d'importants compléments. Des bénéfices évidents peuvent être obtenus grâce à la combinaison de plus d'un capteur. Les faiblesses d'un capteur peuvent être comblées par un autre. La fusion de données issues de plusieurs capteurs produit des systèmes robustes face aux différentes conditions de conduite difficile.

2.2 Etapes de détection des voies routières

Dans cette section nous détaillons les différents modules de détection des voies routières et énumérons des méthodes utilisés dans chaque module. En scrutant la littérature concernant la détection des voies routières on remarque que celle-ci est basée principalement sur des modules qui sont : le prétraitement, l'extraction de caractéristiques et l'ajustement de modèle.

2.2.1 Prétraitement

Il y a plusieurs opérations qui peuvent être appliquées à l'image avant la phase d'extraction de caractéristiques, afin de réduire le bruit et améliorer les caractéristiques recherchées. Les obstacles peuvent être identifiés et enlevés. Les ombres peuvent être

réduits grâce à un prétraitement appliqué à toute l'image. La surexposition ou sous exposition peuvent être réglées par normalisation de l'image, par le contrôle de l'exposition de la camera ou bien par des algorithmes de fusion d'expositions tel que montré dans le chapitre 1 de la partie II. On peut aussi définir les zones d'intérêt en supprimant par exemple la région au-dessus de l'horizon. Les méthodes de prétraitement peuvent être classées en deux familles : amélioration de la qualité d'image en réglant l'effet de l'illumination, et la suppression de parties de l'image qui sont sans importance.

Un aspect de gestion des effets liés à l'illumination est d'adapter la plage dynamique du capteur visuel. Un système de détection de voies routière doit être capable de gérer les différentes conditions d'illumination. La plupart des systèmes ne développent pas leur propre mécanisme pour adapter la plage dynamique de la camera et se base uniquement sur les capacités standards de la caméra qui consistent au contrôle de l'iris ou du gain. Dans le système de détection des voies [Mobileye], quatre images avec différent expositions sont simultanément capturées. Un mécanisme permet le contrôle de l'exposition de chaque image et leur fusion produit une image a grande plage dynamique qui peut gérer les changements d'illumination abruptes et maintien une grande qualité d'image sous différents scénarios. Les ombres générées sur la surface de la route sont une source majeure de bruit, due aux contours qu'ils génèrent. Dans le but de contourner cet effet d'illumination, certains travaux font une variété de transformations dans les espaces de couleur HSL, LAB, YCbCr [Cheng, 2006] [Katramados, 2009] [Alvarez, 2007]. Par la combinaison de canaux de couleurs ils réussissent à obtenir une image invariante à l'illumination, où les zones illuminés ou dans l'ombre produisaient les mêmes intensités.

La deuxième catégorie de prétraitement contient des méthodes qui tentent d'éliminer les parties de l'image qui sont inintéressantes ou bien qui peuvent engendrer une information erronée pour la tâche de détection des voies routières. Des obstacles tels que les voitures, les piétons sont la source majeure de bruits et différentes approches sont utilisées pour les détecter et les enlever de l'image. Dans [Yamaguchi, 2008], le mouvement 2D de chaque point de l'image est suivis et une technique est développée pour générer si le mouvement appartient au sol ou bien à des obstacles. Les mêmes étapes ont été suivis par [Cheng,

2006]. Dans les travaux [Hernandez , 2009], [Huang , 2009], ils utilisent un capteur LIDAR pour générer le scène en 3D, puis la détection des objets sur la route. Une autre technique pour rejeter les parties inutiles de l'image est par la définition de régions d'intérêts. Seulement les régions d'intérêts sont traitées par la phase d'extraction de caractéristiques. La méthode la plus simple pour ça définit la partie basse de l'image comme étant la région d'intérêt [Zhang, 2009]. Dans des travaux comme [Huang, 2009] [Pomerleau, 1995] [Broggi, 2006], la correspondance entre l'environnement 3D et l'image 2D est estimée, puis certaines lois ont été établies pour définir la région d'intérêt. La connaissance de la position des voies routières détectées dans l'image à l'instant t-1 permet de définir la région d'intérêt où les voies routières sont supposés être à l'instant t [Ogawa, 2006], [Wu, 2008].

2.2.2 Extraction de caractéristiques

Une fois que les prétraitements sont effectués et les régions d'intérêt définies, plusieurs caractéristiques peuvent être extrait. Ces caractéristiques devrait contenir l'information nécessaire pour l'étape d'ajustement du modèle de voie routière qui a suivre. Les voies routières sont définies par les marquages routiers qui peuvent être détectés par leur forme ou bien par leur couleur. La supposition la moins restrictive sur les marquages routiers est le fait qu'ils ont une apparence différente comparé à la route. Cette supposition conduit à l'apparition de plusieurs méthodes basées sur le gradient. Un simple gradient a été calculé dans [Samadzadegan, 2006] [Sawano, 2006] [Nieto, 2008] à partir de l'image originale ou bien de sa version lissée. La direction du gradient a été utilisée par [McCall, 2006] en convoluant l'image avec seulement trois filtres. Cette propriété a été utilisée pour l'extraction des contours ayant une certaine direction. Dans les travaux [Huang, 2009] [Labayrade, 2006] [Wu, 2008], la forme des marquage et la supposition que ceux-ci ont une intensité supérieur à leur entourage a motivé les auteurs à chercher dans les lignes de l'image le modèle basse-grande-basse intensité.

Une autre approche pour la détection des marquages routiers suppose que leur intensité lumineuse ou leur couleur est connue. Dans [Ogawa, 2006] le seuillage d'une carte de

reflectance de LIDAR est effectuée. Dans [Borkar, 2009] ils effectuent une moyenne entre les trois images successives, puis un seuillage adaptative est appliqué. La moyenne entre les images permet de remplir le vide entre les marquages discontinus. Dans [Cheng, 2006] [Kornhauser, 2007], l'apprentissage de la distribution de couleur des marquage routier est effectué en amont. Après, les pixels de l'image sont classifiés selon la vraisemblance d'appartenir à la classe marquage. Les caractéristiques résultantes de toutes les procédures susmentionnées sont des blocs segmentés. L'ensemble des blocs extraits sont susceptible de contenir du bruit et ces algorithmes sont souvent augmentés avec un filtrage supplémentaire. Dans [Cheng, 2006] par exemple, les blocs ont été filtrés par leur taille ainsi que leur forme.

2.2.3 Ajustement de modèle

Le principal objectif de cette phase est d'extraire une représentation de haut niveau de la voie routière qui puisse être utilisée pour prendre des décisions. Cette représentation se traduit généralement par un modèle de voie en 2D composée des deux bordures droite et gauche. Le modèle géométrique le plus simple utilisé pour cette représentation est les lignes droites [Rasmussen, 2005], [Kong, 2009], [Alon, 2006], [Zhang, 2009], [Borkar, 2009]. Ce sont une bonne approximation des voies routières dans les zones proches du véhicule et surtout quand il s'agit d'auto routes. Pour représenter des voies routières courbées [Wu, 2008] [Samadzadegan, 2006] utilisent des paraboles et [Labayrade, 2006] utilise des courbes polynomiales. Pour la plupart des méthodes, les données en entrée pour l'ajustement de modèle est un ensemble de points extrait des marquages routiers durant la phase d'extraction de caractéristiques. D'autres informations peuvent être calculées pour guider l'ajustement du modèle. Dans [Kim, 2008], les points sont regroupés en des lignes. Dans [Huang, 2009], un filtre Hessian permet d'ajouter l'information de direction aux points des marquages.

Plusieurs méthodes sont utilisées pour ajuster les modèles paramétriques des données issues de la phase d'extraction de caractéristiques. [McCall, 2006] et [Jiang, , 2010] utilise une optimisation au moindre carrés. [Huang, 2009] [Nieto, 2008] [Borkar, 2009]

utilise aussi les moindres carrés avec l'algorithme de RANSAC pour améliorer la robustesse face au bruit. [Labayrade, 2006] utilise les moindres carrés pondérés [Tarel, 2002] pour obtenir une estimation robuste de la courbe. [Samadzadegan, 2006] la courbure et l'orientation de la voie routière sont retrouvés en utilisant une généralisation de la transformée de Hough. La transformée de Hough est spécialisée pour l'ajustement de modèle linéaire. Un algorithme relativement simple basé sur la transformée de Hough permet de résoudre le problème de détection de voies routière à 90% des cas d'autoroutes [Borkar, 2009]. La transformée de Hough, une forme discrète de la transformée de Radon s'avère être la technique la plus utilisée dans la littérature pour la détection de lignes.

2.3 Conclusion

Parmi les composantes fondamentales d'une scène routière qu'un système d'aide à la conduite doit prendre en considération, les limites des voies de circulation. C'est une information nécessaire pour le bon déroulement de l'activité de conduite. Cette information veille à ce que chaque usager de la route garde sa voie de circulation. Aujourd'hui, plusieurs constructeurs automobiles dotent leurs véhicules haut de gamme avec des systèmes de détection de franchissement de lignes. Ces systèmes se basent généralement sur la vision comme capteur d'acquisition. Malgré commercialisé, ce genre de systèmes, reste peu fiable face à certaines conditions. La vision, bien que riche en informations demeure plus sensible aux conditions d'éclairages (nuit, éblouissement, ombres...etc.), condition climatiques (brouillard, pluie, neige...etc.) ou les occlusions...etc. Les systèmes d'aide à la conduite doivent être capables de faire face à ces situations, car c'est aussi durant ces conditions que le conducteur a le plus besoin de ces systèmes.

III- Contributions

Partie III
Chapitre 3 : Fusion
d'expositions basée sur la
segmentation
d'histogrammes.

3 Fusion d'expositions basée sur la segmentation d'histogrammes

Ce chapitre présente la première méthode de fusion d'expositions que nous avons proposée. Les sections 3.1 et 3.2 présentent respectivement ce qu'est la segmentation d'images ainsi que la technique de segmentation que nous avons proposée pour classifier les différentes régions de l'image qui nous intéressent. La fusion des régions segmentées est ensuite présentée dans la section 3.3. Nous présentons dans la section 3.4 une méthode d'acquisition des expositions qui est basée sur les régions segmentées. Enfin, les résultats de la méthode proposée, les comparaisons ainsi qu'une conclusion sont présentés respectivement dans les sections 3.5 et 3.6.

3.1 Introduction

La fusion d'exposition est communément réalisée en tant qu'une moyenne pondérée entre les pixels correspondants. La pondération de tous les pixels du jeu d'expositions ralentit le processus de fusion et rend l'acquisition en temps réel difficile. Dans ce chapitre, nous présentons une méthode de fusion d'expositions qui est rapide. Celle-ci est basée sur la segmentation des histogrammes. La phase de segmentation réduit considérablement les calculs tout en préservant une qualité de fusion compétitive. Cette phase a pour but de détecter les zones sombres, les zones claires et les régions intermédiaires de la scène. Une telle classification permet d'éviter beaucoup de calculs. La fusion n'est effectuée que sur les zones intermédiaires. Aussi, la segmentation des expositions permet de réduire l'influence des pixels saturés ou assombris sur la qualité de l'image fusionnée. Nous présentons également dans ce chapitre un moyen automatique de prendre des expositions pour la fusion, en utilisant les régions segmentées. Des comparaisons subjectives et objectives sont menées pour prouver l'efficacité de notre méthode.

3.2 La Segmentation d'images

La segmentation d'images, est une opération très importante en traitement d'image qui a pour but de séparer les différentes régions homogènes d'une image. Il s'agit de regrouper les pixels de l'image ayant en commun diverses propriétés (intensité, couleur, texture, ...etc.) en groupes (clusters). On peut distinguer deux catégories de méthodes de segmentation: la segmentation non supervisée, qui divise l'image automatiquement en régions, sans aucune connaissance préalable des classes; et la segmentation supervisée, qui démarre à partir de la connaissance de chacune des classes définies. La figure suivante montre un exemple de segmentation d'image à l'aide de la méthode d'Otsu. Cette méthode suppose que l'image contient deux classes, puis calcule le seuil optimal séparant ces deux classes. On arrive à distinguer clairement les pièces de monnaie du fond.

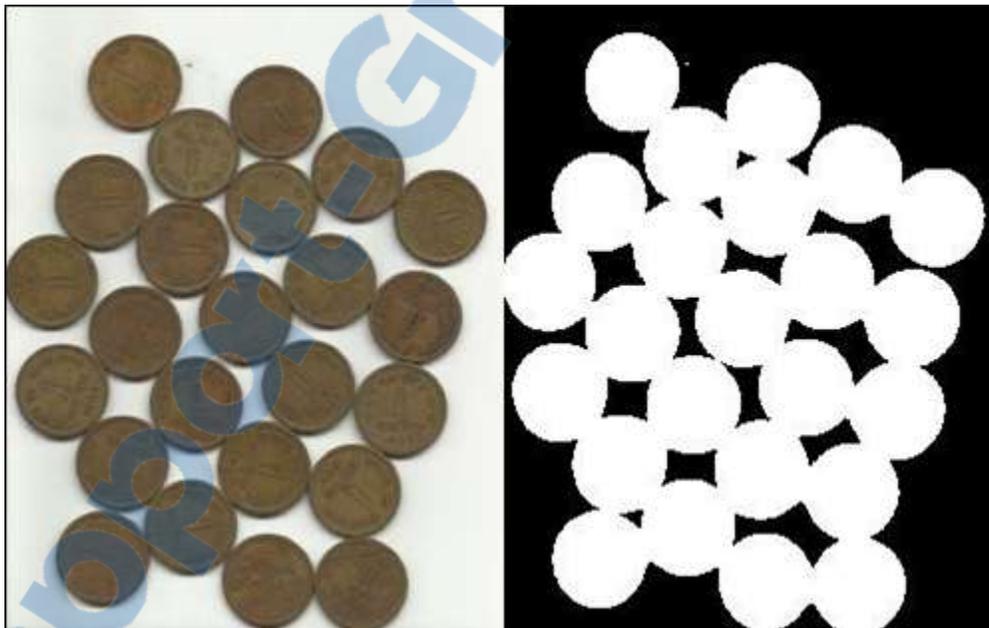


Figure 3-1: Segmentation d'images en régions avec l'algorithme d'Otsu.

La figure 3-2 montre deux expositions capturées face à une scène à grande plage dynamique. On remarque que la scène contient trois types de régions qui produisent trois types de groupes de pixels. Le premier groupe concerne les pixels sombres. Ce sont les pixels sous exposés que l'on retrouve dans l'image à exposition rapide (Le sable et les rochers). Le deuxième groupe de pixels est celui des zones claires. Ce sont les zones très

lumineuses qui deviennent saturées rapidement en augmentant le temps d'exposition (Le ciel, les nuages). Le troisième groupe de pixels est celui des régions intermédiaires. Celles qui sont moyennement exposées et qu'on arrive à distinguer dans les deux expositions.

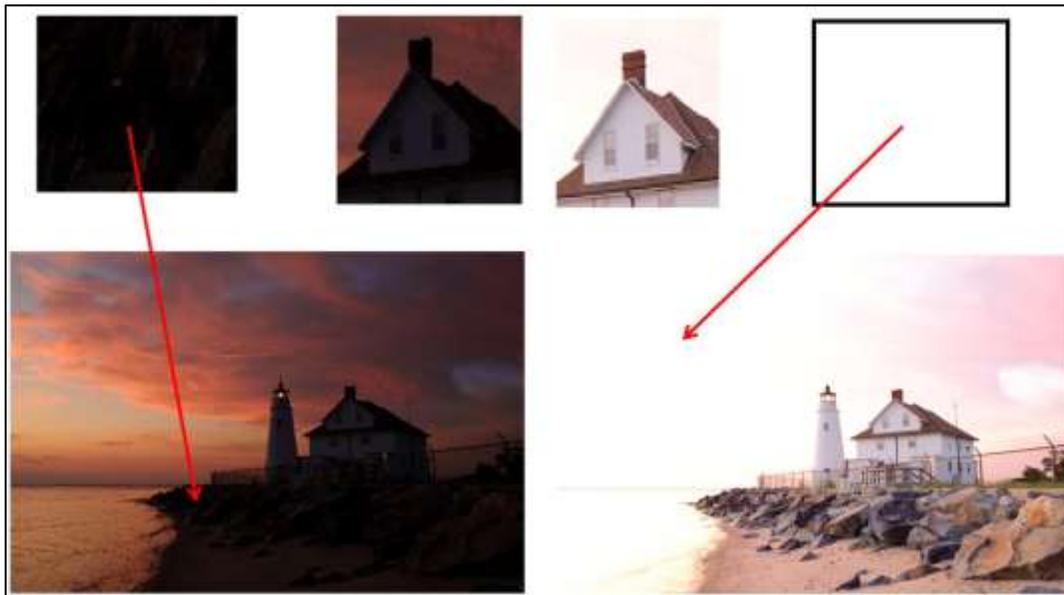


Figure 3-2: Les trois types de régions qu'on retrouve dans un jeu d'expositions.

En se basant sur ces observations, l'idée que nous proposons consiste en la segmentation des expositions en zones claires, sombres et intermédiaires puis la reconstitution d'une image de fusion à partir de ces régions. Cette segmentation va non seulement éviter de faire des calculs inutiles sur les pixels sous exposés/saturés (estimations de mesures de qualité) mais aussi éviter que ceux-ci influencent la qualité de l'image de fusion. Les pixels correctement exposés sont pris directement de l'une des expositions, sans fusionner leurs valeurs avec celles des pixels correspondants dans les autres expositions.

3.3 Segmentation d'histogrammes

Une image est composée de pixels ayant différentes valeurs. La distribution de ces valeurs dans l'image constitue une caractéristique importante qui est l'histogramme. Dans une image en niveau de gris, chaque pixel est représenté par une valeur entre 0 (noir) et 255 (blanc). Un histogramme est une simple table qui représente le nombre de pixels ayant une valeur donnée dans l'image. Un histogramme d'une image en niveau de gris, comme le montre la figure 3-3 contient 256 entrées (valeur de niveau de gris), où chaque entrée est représentée par le nombre d'occurrence dans l'image.

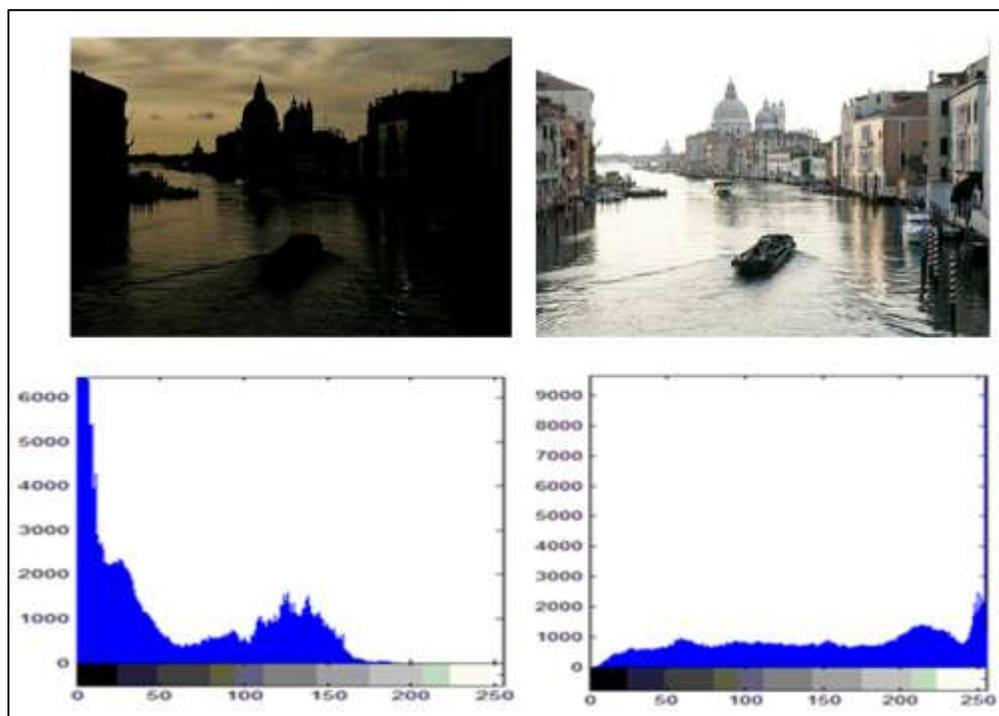


Figure 3-3: Histogrammes des expositions.

La majorité des travaux de fusion d'exposition qui visent une fusion en temps réel considère que l'utilisation de deux expositions est suffisant [Chapiro, 2011] [Kang, 2003] [Kao.,2007]. L'utilisation de deux expositions réduit le nombre d'opérations à effectuer et couvre généralement la majorité des détails de la scène. Nous visons aussi une fusion d'exposition rapide dédiée à la capture de vidéo. C'est pour cela que nous considérons aussi deux expositions : une image surexposée O et une autre image U qui est sous exposée. Plus de détails sur la manière de capture de ces expositions sont



introduits dans la section 3.4 de ce chapitre. La phase de segmentation d'expositions proposée dans notre travail est basée sur les histogrammes notés $H_k(L)$. $H_k(L)$ représente le nombre de pixels trouvés dans l'exposition $k \in \{U, O\}$ ayant l'intensité L . Les zones sombres sont des pixels ayant de petites intensités dans l'image sous exposée. Ils forment un grand mode dans l'histogramme H_u positionné à gauche (figure3-3). Les zones claires sont les pixels ayant de fortes intensités dans l'image surexposée. Ceci est reflété par l'apparition d'un autre grand mode cette fois ci à droite de l'histogramme H_o (figure3-3). Ces suppositions sont réalistes et démontrés par les histogrammes des expositions. Les zones claires et sombres sont donc deux modes dans les histogrammes qui peuvent être représentés par des fonctions Gaussiennes, tel que montré dans la figure3-4. La phase de segmentation d'expositions est donc faite comme l'estimation des paramètres de ces deux Gaussiennes qui sont les maximums et les écart types (M_u, T_u, M_o, T_o) . M_u et M_o sont considérés comme les indices d'intensité liés aux plus grandes valeurs d'histogrammes respectifs H_u et H_o . Les écarts type T_u et T_o sont estimés en utilisant les dérivées première et seconde des histogrammes après un lissage comme suit :

$$T_u = \min\{i \in [0,255] \text{ where } H_u''(i) \times H_u''(i-1) \leq 0 \text{ and } H_u'(i) < 0\} \dots \dots (34)$$

$$T_o = \max\{i \in [0,255] \text{ where } H_o''(i) \times H_o''(i-1) \leq 0 \text{ and } H_o'(i) > 0\} \dots \dots (35)$$

Une fois que les écarts types sont estimés, on peut sélectionner les régions qui nous intéressent pour la fusion à l'aide des règles suivantes :

$$S = \{p_{x,y} \in O \text{ where } U_{x,y} \leq T_u\} \dots \dots (36)$$

$$C = \{p_{x,y} \in U \text{ where } O_{x,y} \geq T_o\} \dots \dots (37)$$

$$I = \{p_{x,y} \in \{U \cup O\} \text{ where } U_{x,y} > T_u \text{ and } O_{x,y} < T_o\} \dots \dots (38)$$

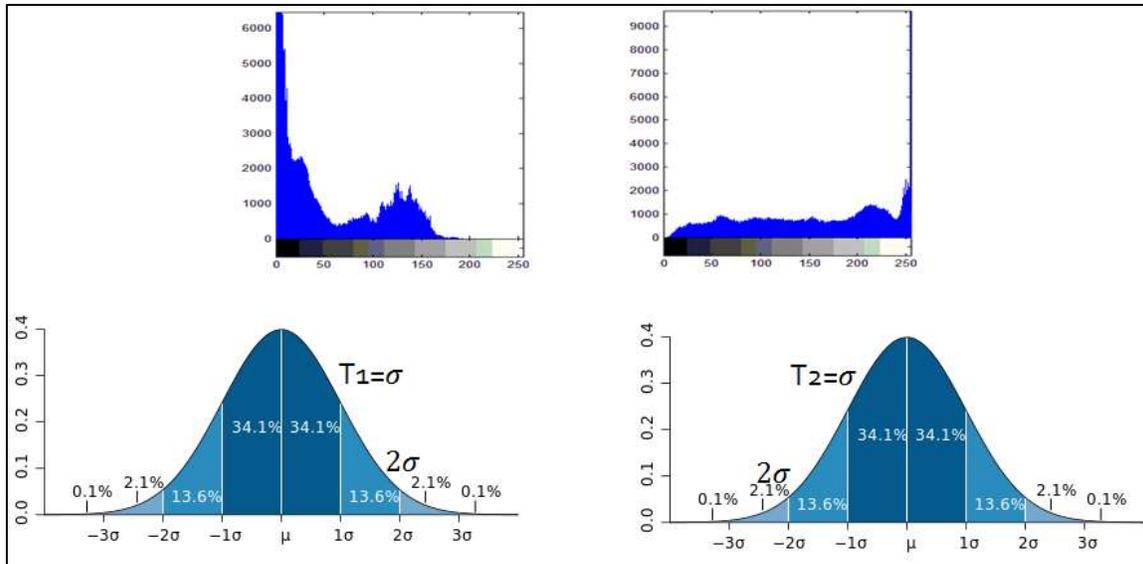


Figure 3-4: Représentation des zones recherchées par des Gaussiennes.

La figure 3-5 montre les résultats de la phase de segmentation. Les trois types de régions définies sont bien localisées grâce à la segmentation d'histogrammes proposée. La prochaine section montre la manière d'obtenir une seule image regroupant tous les détails.

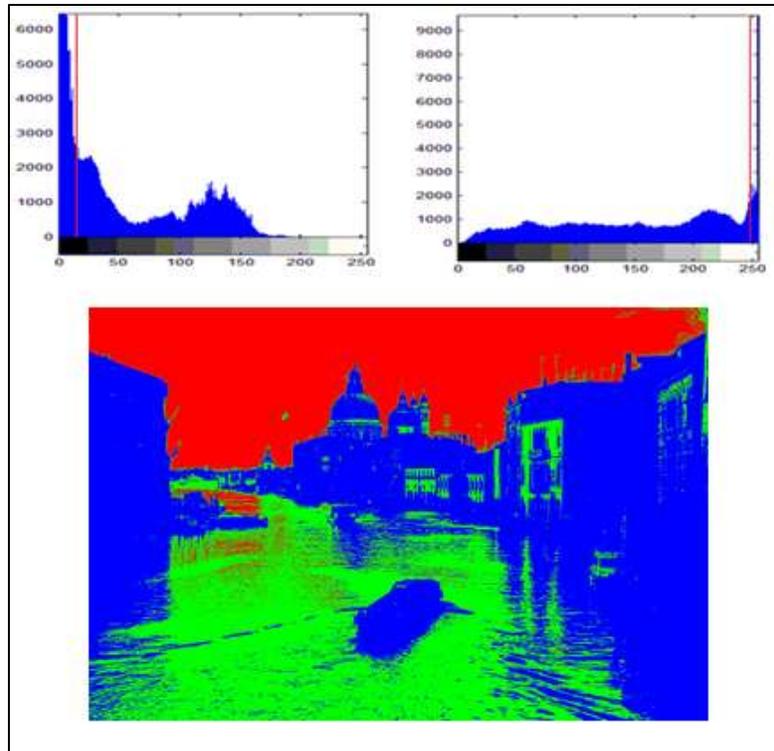


Figure 3-5: Segmentation d'histogrammes. (Bleu) : Zones sombres, (Rouge) : zones claires, (Vert) : zones intermédiaires.

3.4 Fusion des régions segmentées

Une fois les régions S, C, I segmentées, il ne reste plus qu'à les assembler en une seule image R qui contient tous les détails. Les zones sombres sont prises directement de l'image surexposée car elles sont complètement noires dans l'image sous exposée. Les zones claires sont tirées de l'image sous exposée car elles sont saturées dans l'autre exposition. Seul les zones intermédiaires nécessitent une fusion avec pondération des pixels. Pour cela nous avons supposé que plus les pixels sont claires dans la photo sous exposée, plus ils sont pondérés. Dans l'image surexposée, plus les pixels sont claires, moins ils sont pondérés. Cette manière de donner les poids aux zones intermédiaires est formalisée par deux simples fonctions basées sur les écarts types estimés (W_1, W_2) :

$$W_1(U_{x,y}) = U_{x,y} - T_U \dots \dots (39)$$

$$W_2(O_{x,y}) = T_O - O_{x,y} \dots \dots (40)$$

Les valeurs finales des zones intermédiaires dans l'image R sont calculées avec l'équation suivante:

$$R_{x,y} = \frac{U_{x,y} \times W_1(U_{x,y}) + O_{x,y} \times W_2(O_{x,y})}{W_1(U_{x,y}) + W_2(O_{x,y})} \dots \dots (41)$$

La figure3-6(a) montre un exemple de résultat insatisfaisant de la fusion d'expositions. Il y a une différence de tons entre les régions fusionnées car elles sont issues d'images différentes. Pour résoudre ce phénomène, [Mertens , 2007] utilisent la décomposition en Laplacien. Cette technique est très couteuse en temps de calcul. Une autre solution consiste à traiter les expositions avec une égalisation d'histogrammes [Kang ,2012]. Cependant cette technique produit une grande déformation des intensités, produisant parfois une image fusionnée de moindre qualité (comme montré dans la figure3-6(c)). Pour obtenir une fusion cohérente avec un minimum de temps de traitement, nous utilisons deux fonctions linéaires qui rapprochent les tons des deux expositions fusionnées (Réduire la luminance de O et augmenter la luminance de U). On a posé ($d = 2.3$) par expérimentation. La figure3-6(b) montre l'image de fusion R après l'ajustement proposé.

$$u(x, y) = \frac{U_{x,y} + (255 - U_{x,y})}{d} \dots \dots (42)$$

$$o(x, y) = O_{x,y} - \frac{O_{x,y}}{d} \dots \dots (43)$$

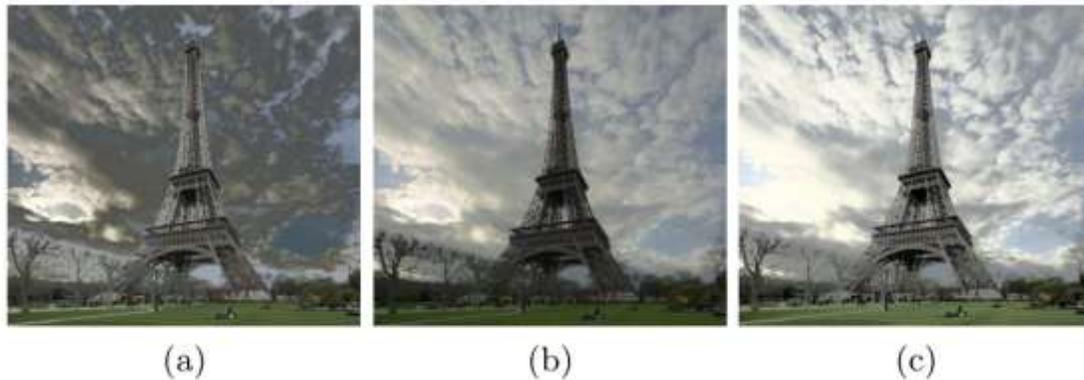


Figure 3-6: Reconstruction de l'image de fusion avec les régions segmentées. (a) sans aucun ajustement. (b) : Ajustement proposé. (c) Egalisation d'histogrammes.

3.5 Acquisition des expositions

L'acquisition des expositions pour la fusion est une partie importante du problème. Usuellement, les expositions sont capturées avec des paramètres ou seuils fixés à l'avance. Ceci peut engendrer une perte d'information dans les zones sombres et claires. [Gelfand , 2010] supposent que O doit être capturée quand le nombre des pixels sombres (luminosité < 16) est inférieur à 10%. U est capturée quand moins de 10% de pixels ont une intensité plus grande que 239. [Bilcu , 2008], [Tico , 2010] proposent aussi une capture des expositions basée sur des seuils fixes. Une technique plus sophistiquée pour capturer les expositions est d'ajuster les paramètres d'exposition en fonction de la qualité de régions d'intérêt. [Kao , 2007] propose un algorithme qui ajuste les expositions après avoir détecté la couleur de la peau humaine. Dans notre travail, les premières exposition sont capturées par des paramètres fixés tel que fait par [Gelfand , 2010]. Après, U et O sont segmentées en régions comme expliqué dans la section 3.2. Les régions détectées S et C sont utilisées avec l'algorithme d'auto exposition de la caméra pour produire respectivement une image sous exposée et une autre surexposée. Pour cela nous avons utilisé l'architecture du Nokia N900 pour l'implémentation [Adam 2010]. L'algorithme d'auto exposition implémenté par [Adam 2010] ajuste le temps d'exposition et le gain de la caméra en fonction de l'histogramme de l'image. Nous avons adapté cet algorithme pour qu'il ne comptabilise que les pixels des zones S et C à tour de rôle. La figure 3-7 montre les expositions obtenues. Notre technique d'acquisition d'exposition capture plus de détails comparée à celle de l'application Fcam du Nokia N900.

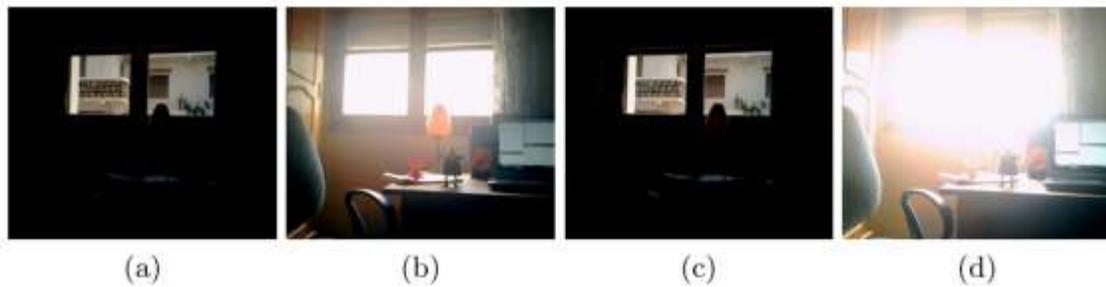


Figure 3-7: Capture automatique d'exposition pour la fusion. (a,b) : Zones sombres et claires acquises avec notre algorithme. (c,d) : Fcam acquisition. (temps d'exposition, gain) : a(1666,1), (b)(40000,1.58),c(2397,1),(d)(40000,6.68).

3.6 Résultats et comparaisons

Pour prouver l'efficacité de la méthode de fusion d'exposition proposée, nous la comparons avec les résultats obtenus avec l'algorithme de Mertens [Mertens , 2007] et ceux obtenus avec le logiciel de fusion d'exposition Photomatix [Photomatix]. Les figure3-8 et 3-9 montrent les sorties des trois méthodes. Subjectivement, on peut voir que notre méthode apporte plus de détails sur certaines régions comparée à l'algorithme de Mertens (le bateau sur la scène « venise » et les contours sur la scène « tour Eiffel »). Cette amélioration est due au fait que les zones sombres et claires sont segmentées et tirées des bonnes expositions sans les fusionner avec les zones qui leur correspondent dans les expositions saturées ou sous exposées. L'algorithme de Mertens utilise une fonction de pondération Gaussienne qui attribue des poids similaires à ces régions dans les deux expositions. Une comparaison visuelle entre notre méthode et la fusion de Photomatix montre que les deux résultats sont proches.

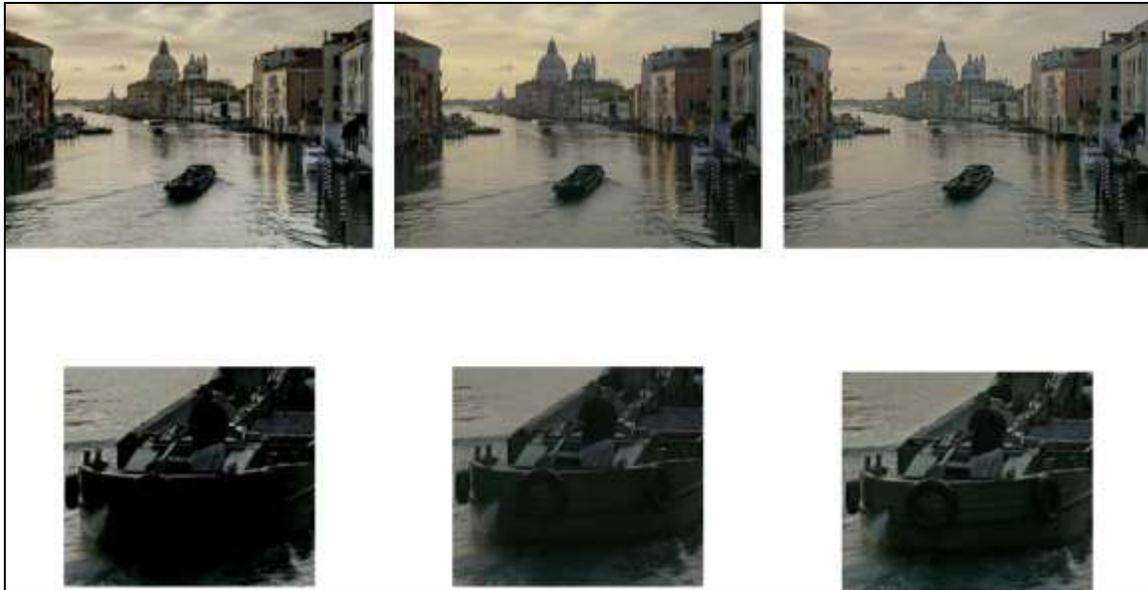


Figure 3-8: Résultats obtenus par les trois méthodes : (Gauche) Algorithme de Mertens, (Milieu) notre méthode, (Droite) : Photomatix.

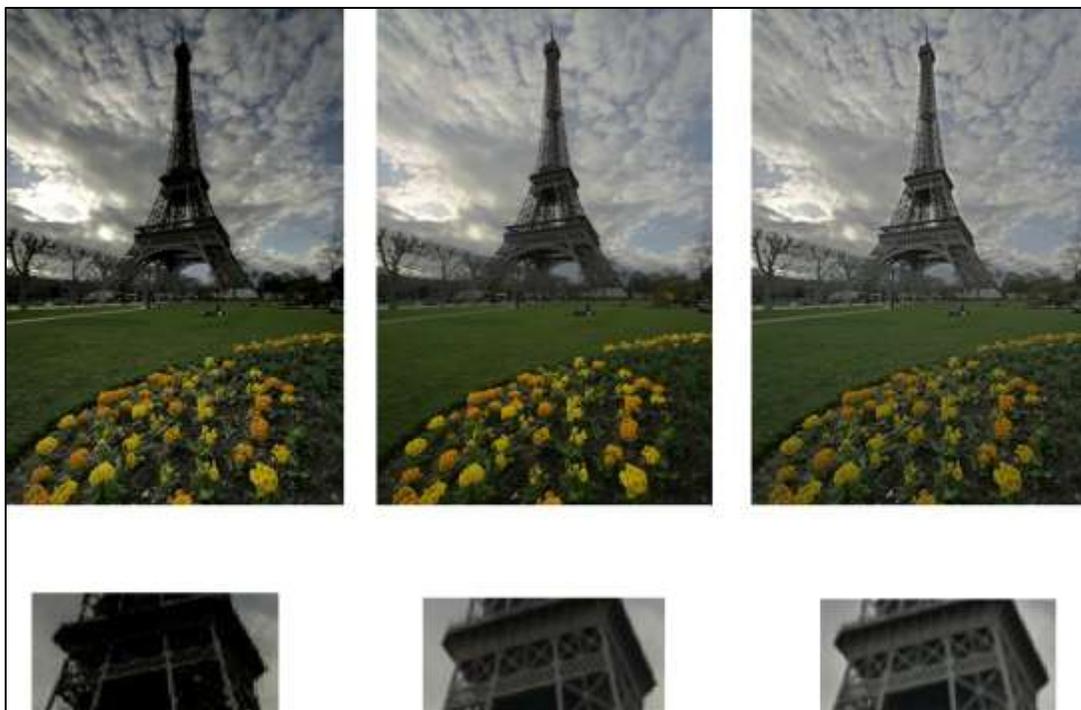


Figure 3-9: Résultats obtenus par les trois méthodes : (Gauche) Algorithme de Mertens, (Milieu) notre méthode, (Droite) : Photomatix.

Nous avons utilisé trois mesures de qualité pour effectuer une comparaison objective entre les résultats obtenus et ceux des deux autres méthodes. La première mesure de qualité est l'information mutuelle, adaptée pour l'évaluation de la performance de fusion d'images [Guihong , 2002]. Par la suite nous avons calculé la différence de couleur estimée par le critère CIEDE2000 [Sharma , 2005] ainsi que la différence de contraste (différence de la magnitude du gradient). Les images de vérité terrain sont construites manuellement avec les meilleures parties des expositions en entrée sans déformer les valeurs des pixels. Les résultats sont présentés dans le Tableau 1, ils montrent que notre méthode produit généralement les meilleures performances en termes d'information mutuelle, CIEDE2000 et de différence de contraste.

Metrique	Méthodes	Venise	Eiffel	LightHouse	Landscape
Contrast	Mertens	1,15	0,85	0,32	1,53
	Notre Méthode	1,05	0,71	0,27	1,34
	Photomatix	1,13	0,72	0,3	1,4
Couleur	Mertens	1,56	2,89	1,17	1,26
	Notre Méthode	0,99	1,99	1,1	1,18
	Photomatix	0,94	1,79	1,63	1,57
IM	Mertens	2,93	3,42	2,77	4,55
	Notre Méthode	6,69	7,46	5,57	7,73
	Photomatix	5,2	6,29	6,3	6,88

Tableau 1 : Comparaison objective : Contraste (différence en magnitude de gradient), Couleur (CIEDE2000 [Sharma , 2005]) et l'information mutuelle [Guihong , 2002].

Dans le tableau2 on peut voir que les zones sombres et claires segmentées représentent une large partie des expositions fusionnées. Ces régions sont directement sélectionnées pour former l'image de fusion sans passer par leur pondération. Ce qui fait que notre méthode est très peu gourmande en temps d'exécution (voir tableau2). Une version C++ de l'application de fusion exécutée sur un processeur 2.0GHZ et 3GO de mémoire

produit en moyenne des images fusionnées en couleurs avec une vitesse de 29 images/seconde, ayant une résolution de 640×480.

Scène	Venise	Eiffel	Lighthouse	Landscape
Résolution	640 480	530 795	1024 682	1800 1198
S+C (%)	63,99	69,51	60,28	46,53
Mertens	2,1	2,5	3,4	10,4
Notre méthode	0,3	0,5	1,1	2,4

Tableau 2 : Temps de fusion d'exposition en secondes (Implémentation Matlab) et pourcentage des zones segmentées.

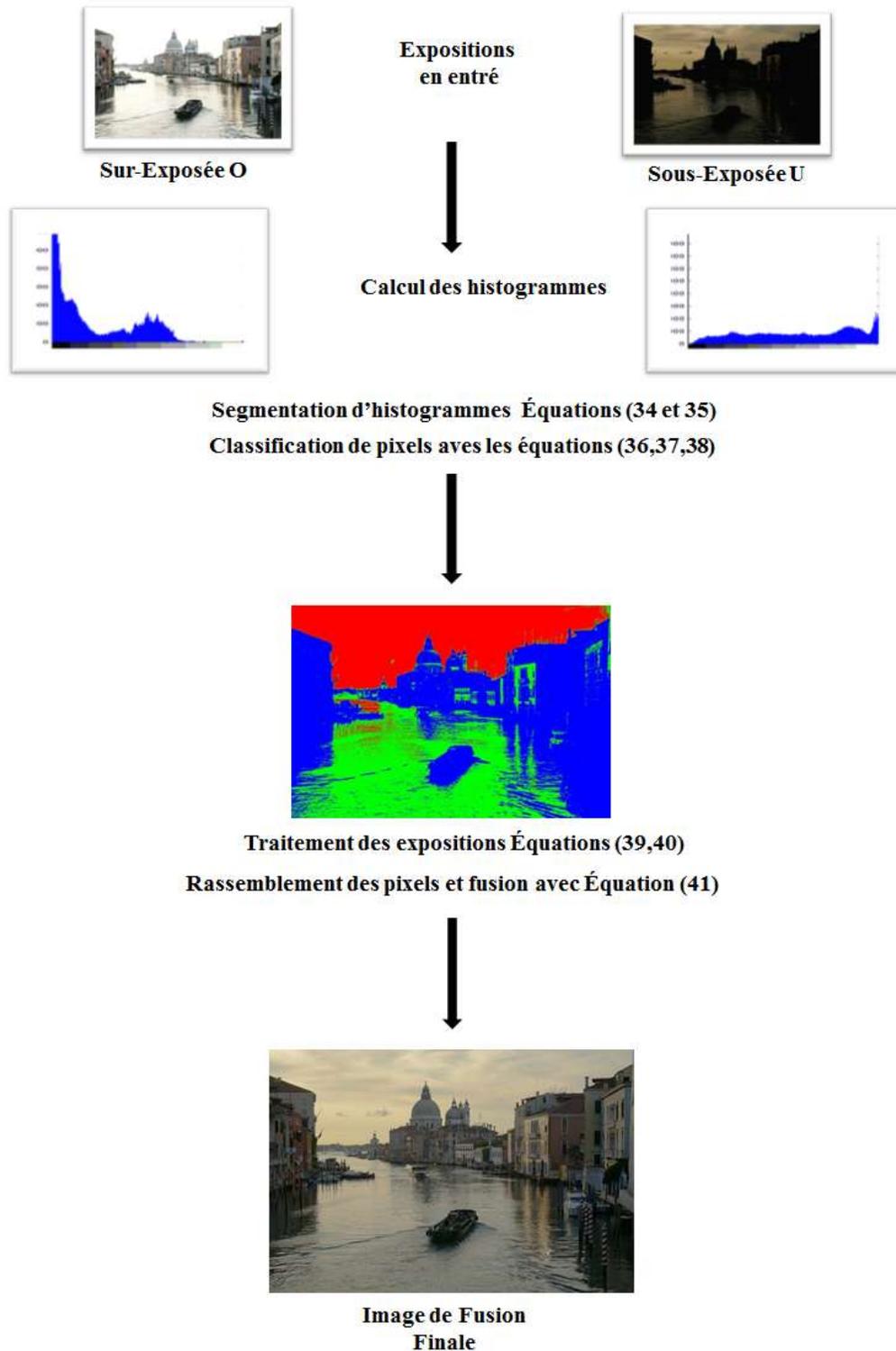


Figure 3-10: Organigramme de la méthode fusion d'expositions basée sur la segmentation d'histogrammes.



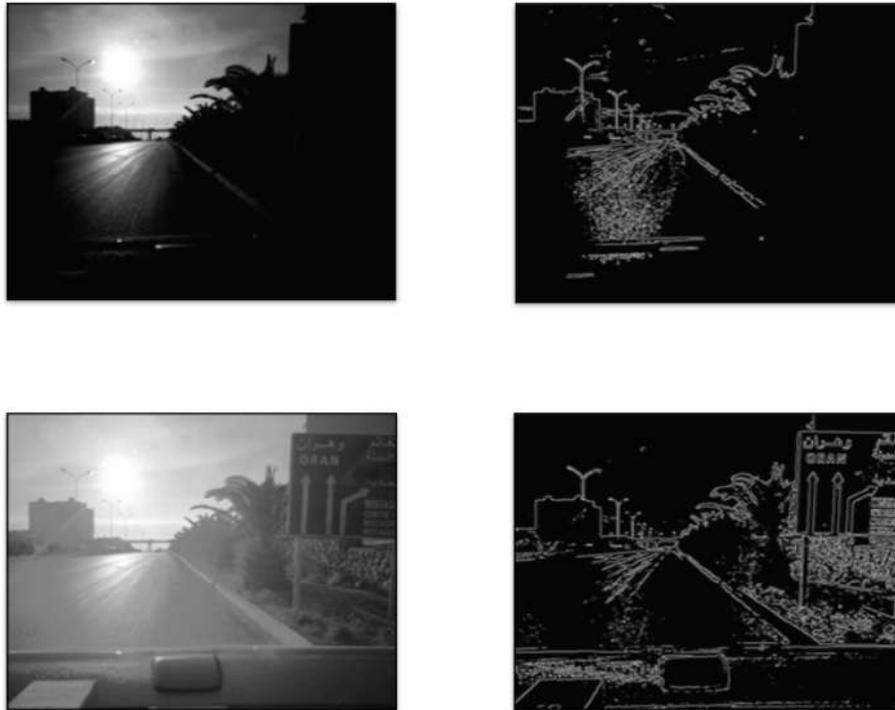


Figure 3-11: L'apport de la fusion d'exposition face à une scène de conduite à grande plage dynamique. (Haut) Algorithme original de la caméra. (Bas) Image résultante de la fusion d'expositions.

3.7 Conclusion

Dans ce chapitre nous avons présenté une méthode de fusion d'exposition qui est non seulement très rapide mais qui préserve aussi une qualité de fusion d'images similaires à certaines méthodes de fusion d'exposition pertinentes. Notre méthode produit 29 images en couleur par seconde, des images riches en détails face à des scènes à grande plage dynamique. La phase de segmentation d'exposition proposée peut être le point de départ de n'importe quel autre algorithme de fusion d'exposition afin de diminuer le temps de calculs mais aussi diminuer l'influence des pixels saturés ou sous exposés sur la qualité de l'image de fusion. Nous avons aussi démontré dans ce chapitre qu'utiliser l'algorithme d'auto exposition sur les régions segmentées peut produire de meilleures expositions pour la fusion. La figure suivante montre une scène de conduite obtenue à l'aide de caméras normales face à un soleil bas, puis la version obtenue à l'aide de fusion d'exposition. L'utilisation de la fusion d'expositions apporte un gain appréciable en informations pertinentes pour les systèmes d'aide à la conduite.

Partie III

Chapitre 4 : Fusion d'expositions basée sur l'apprentissage supervisé.

4 Fusion d'expositions basée sur l'apprentissage supervisé.

Dans ce chapitre nous présentons deux nouvelles méthodes de fusion d'expositions qui sont basées sur l'apprentissage supervisé. La phase de pondération des pixels pour la fusion est complètement évitée. L'idée est d'utiliser une collection d'images d'apprentissage composée d'expositions en entrée et leur fusion pour définir à la fin des lois pour la fusion d'exposition. La première méthode, introduite dans la section 4.1 consiste à retrouver à partir de la base d'apprentissage des poids de fusion généralisés qui peuvent être utilisés par la suite pour fusionner n'importe quels expositions avec un minimum d'opérations. La deuxième méthode est présentée dans la section 4.4. Basée sur la même collection d'images d'apprentissage, cette méthode consiste à estimer un modèle de régression pour la fusion d'expositions.

4.1 Introduction

La fusion d'exposition est principalement réalisée en tant que moyenne pondérée entre les pixels correspondants. Le défi consiste à trouver les poids de pixels qui produisent la meilleure qualité d'image fusionnée et dans un minimum d'opérations pour répondre aux exigences en temps réel. Dans ce chapitre nous présentons deux méthodes basées sur un apprentissage supervisé. L'idée consiste à estimer les poids de fusion d'expositions ou bien un modèle de fusion qui peuvent être utilisés pour fusionner des expositions très rapidement. Ceci est une solution pour éviter les opérations de pondération qui sont habituellement effectuées pour chaque image fusionnée. L'idée que nous proposons est d'utiliser une collection d'images composées d'entrées et de leurs fusions pour définir des poids de fusion généralisés ou bien un modèle de fusion.

4.2 Estimation de poids généralisés pour la fusion d'exposition

Tel que montré dans l'état de l'art, la fusion d'exposition est souvent réalisée comme étant une moyenne pondérée entre les pixels correspondants. La fusion de N expositions en une seule image R consiste à effectuer la somme pondérée des pixels correspondants [Mertens , 2007] :

$$R_{i,j} = \sum_{k=1}^N I_{i,j,k} W_{i,j,k} \dots \dots (44)$$

Avec I_k est la $k^{ième}$ exposition en entrée utilisée pour la fusion, $W_{i,j,k}$ est le poids attribué au pixel (i, j) dans l'image k . Comme mentionné précédemment, deux expositions sont souvent suffisantes pour acquérir la majorité des détails d'une scène à grande plage dynamique. De ce fait l'équation de fusion devient :

$$R_{ij} = U_{ij}W_{ij,u} + O_{ij}W_{ij,o} \dots \dots (45)$$

Habituellement, tous les pixels appartenant aux expositions fusionnées doivent être pondérés grâce à des fonctions de pondération (mesures de qualité). Cette phase de pondération demande beaucoup de calculs, ce qui ralentit considérablement le processus de fusion. Notre travail consiste à éliminer cette phase de pondération en définissant des poids généralisés de fusion qui peuvent être utilisé pour fusionner n'importe quelles expositions. Pour cela nous supposons que les poids attribués aux pixels sont reliés à leur luminance. Plusieurs travaux supposent que les intensité du milieu de l'histogramme sont les meilleures et utilisent des Gaussiennes comme fonction de pondération [Mertens , 2007] [Zhang , 2012] [Gelfand , 2010] [Jinno , 2012]. Les paramètres de ces fonctions de pondération sont choisis par expérimentation ce qui amène à une perte de détails dans les valeurs extrêmes. Pour retrouver la relation entre la luminance et les poids de fusion nous avons utilisé une méthode d'apprentissage. A partir d'un ensemble d'images composé

d'expositions (sous exposées U_k , et surexposées O_k) et leurs fusion R_k , l'estimation des poids est cherchée comme étant la solution au problème de minimisation :

$$E(W_{ij,u}, W_{ij,o}) = (R_{ij} - U_{ij}W_{ij,u} - O_{ij}W_{ij,o})^2 \dots \dots (46)$$

La luminance d'un pixel est une mixture des luminances dans l'image sous exposée et surexposée. Si la luminance du pixel est pareille dans les deux expositions et qu'elle est égale à une valeur v , alors la luminance du pixel fusionné est aussi v . Nous avons besoin de résoudre le problème de minimisation dans l'équation sous la contrainte

$$W_{ij,u} + W_{ij,o} = 1 \dots \dots (47)$$

On peut formuler alors la nouvelle fonction à minimiser à l'aide des multiplicateurs de Lagrange

$$E(W_{ij,u}, W_{ij,o}, \lambda) = (R_{ij} - U_{ij}W_{ij,u} - O_{ij}W_{ij,o})^2 + 2\lambda(W_{ij,u} + W_{ij,o} - 1) \dots \dots (48)$$

En calculant les dérivées partielles de l'équation précédente, on retrouve le système d'équation suivant :

$$\begin{cases} \frac{\sigma E}{\sigma W_u} = U^2W_u + UOW_o - UR + \lambda = 0 \\ \frac{\sigma E}{\sigma W_o} = UOW_u + O^2W_o - OR + \lambda = 0 \dots \dots (49) \\ \frac{\sigma E}{\sigma \lambda} = W_u + W_o - 1 = 0 \end{cases}$$

Après une résolution complète du système d'équations on obtient :

$$W_{ij,u} = \frac{(R_{ij} - O_{ij})}{(U_{ij} - O_{ij})} \cdot \dots \quad (50)$$

$$W_{ij,o} = \frac{(U_{ij} - R_{ij})}{(U_{ij} - O_{ij})} \cdot \dots \quad (51)$$

Quand $U_{ij} = O_{ij}$, il n'y a pas suffisamment d'information présente dans les expositions pour produire une meilleure image que les entrées ($R = U = O$). Dans ce cas le poids 0.5 est affecté aux deux pixels dans les deux expositions. Une fois que les poids de tous les pixels sont calculés à l'aide des equations 3.1 et 3.2, nous devons définir les poids liés aux luminances pour qu'ils puissent être utilisés pour fusionner d'autres expositions. Le poids de la luminance noté $W_k[L]$ représente le poids attribué à un pixel dans l'exposition k ayant la luminance L . $W_u[L]$ est estimé comme étant la moyenne des poids attribués aux pixels appartenant aux expositions sous exposées, et ayant la luminance L . Avec le même principe, $W_o[L]$ est estimé comme étant la moyenne des poids attribués aux pixels appartenant aux expositions surexposées, ayant l'intensité L . Les deux tableaux qui résultent de cette opération $W_u[L]$ et $W_o[L]$ ayant chacun la taille de 256, représentent les poids de fusion qui devrait être attribués aux pixels en fonction de leur luminances et leur image source. Pour des images en couleur, l'estimation des poids généralisés et le processus de fusion sont fait sur chaque canal séparément.

4.3 Expérimentation et comparaisons

Afin d'expérimenter la méthode proposée nous avons utilisé différentes collections d'images composées d'images en couleurs issues de 17 différentes scènes d'extérieur et d'intérieur. Les expositions en entrée de ces 17 scènes viennent de différentes sources : notre acquisition, images fournies par Jacques Joffre [Photomatix], ainsi que des images fournies par Sebastian Negraszus et Daniel Pirch [Negraszus , 2010]. Les images fusionnées en sortie (vérité terrain) sont construites en utilisant les différentes méthodes de fusion d'exposition : la première est l'algorithme de Mertens [Mertens , 2007], la deuxième méthode est celle que nous proposons dans le chapitre 1, et

la troisième méthode est la fusion d'exposition issue du logiciel Photomatix [Photomatix]. La figure4-1 montre les poids estimés à partir de collection d'apprentissage composés seulement d'une seule scène. Puis, l'utilisation de ces poids pour la fusion des expositions de cette même scène. Les images de fusion utilisées sont issues de la méthode proposée basée sur la segmentation d'histogrammes. La figure4-1 montre les poids estimés des scènes « Eiffel » et « Landscape » respectivement. On peut voir que notre méthode produit des poids qui permettent une fusion correcte.

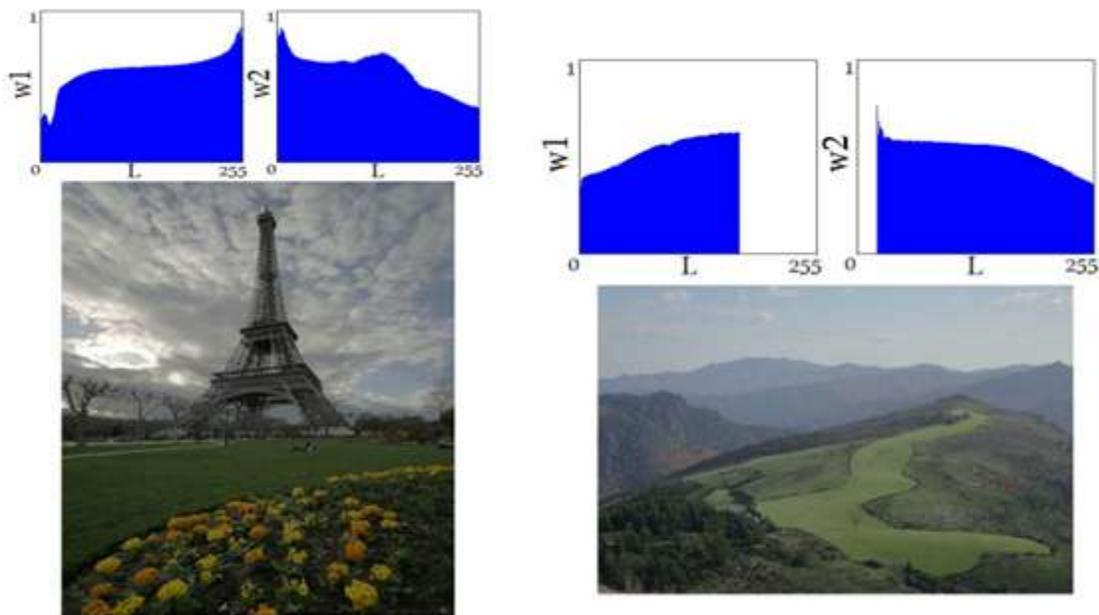


Figure 4-1: Estimation des poids à partir de collection d'images.

Dans la figure4-2 on voit les poids estimés à partir de la scène « Eiffel », où les images de vérité terrain utilisées sont produites respectivement par l'algorithme de Mertens et le logiciel de traitement d'image Photomatix. On peut voir que l'utilisation de ces poids ne permet pas une restitution parfaitement correcte de l'image de fusion. La collection d'images basée sur vérité terrain issue de notre méthode (segmentation d'histogrammes) est plus adéquate pour l'estimation des poids de fusion. Ceci est du au fait que notre méthode se base uniquement sur la luminance des pixels pour les pondérer.

Les images fusionnées à l'aide de mesures de qualités telles que le contraste, ou bien celles qui passent par un processus de fusion multi résolution, peuvent amener à des poids généralisés de fusion qui sont erronés.

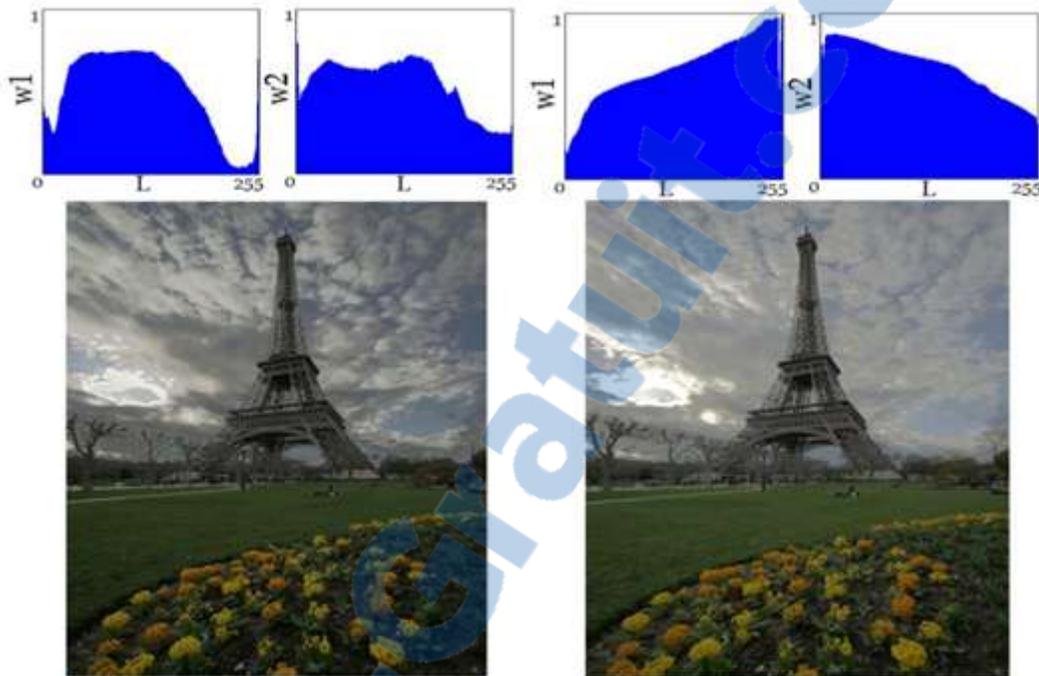


Figure 4-2: Estimation des poids de luminance à partir de la scène « Eiffel ». (Gauche) collection d'images basée sur l'algorithme de Mertens. (Droite) : collection d'image basée sur la fusion d'exposition de Photomatix.

En se basant sur les expérimentations effectuées, quelques observations sont à noter. Les poids estimés de la scène « Landscape » ne peuvent pas être utilisés pour fusionner des expositions issues de scène ayant d'hautes intensités dans l'image sous exposée comme « Eiffeltour ». Parce que dans la scène « Landscape » on ne retrouve pas de pixels ayant une grande intensité dans l'image sous exposée, et donc on ne pourra pas estimer les poids des grandes luminances (voir figure4-1). L'estimation des poids généralisés doit être effectuée à partir d'une collection d'images qui contient différentes scènes et couvre toutes les valeurs de luminance. Il est préférable d'utiliser dans la collection d'images les mêmes résolutions (taille des images) car des images à grande résolution peuvent influencer les poids estimés. Quand nous avons estimé les poids issus des scènes « Eiffel » et « Landscape » ensemble, nous avons obtenu des poids qui ont permis une

fusion correcte de toutes les autres scènes (Voir figure4-3). Changer ou ajouter d'autres scènes produit approximativement les mêmes poids et les mêmes images fusionnées. Nous concluons que les poids estimés sont stables et peuvent être utilisés dans l'équation pour fusionner n'importe quelles autres expositions.

Nous avons comparé la qualité des images fusionnées par les poids généralisés estimés avec celle des trois méthodes de fusion d'expositions citées en haut. La figure4-4 montre les résultats des quatre méthodes. Subjectivement les images paraissent être proches les unes des autres et on remarque que l'algorithme de Mertens perd des détails dans les zones sous exposées et saturées. Dans la figure4-4 (b) le bateau est sous exposé et les fenêtres du building sont saturées alors qu'on peut trouver des détails sur ces zones dans les images en entrée. Ceci est dû au fait de l'utilisation de la fonction Gaussienne pour la pondération des intensités. Les mêmes poids sont attribués aux zones extrêmes des deux expositions (les pixels sous exposés ou saturés sont pondéré de la même manière que leur correspondants dans les autres expositions. Dans les figures4-4 (c) on confirme que les poids généralisés estimés permettent d'éviter ce problème. Les valeurs extrêmes ne sont pas pondérées de la même manière, ce qui permet de produire plus de détails que l'algorithme de Mertens dans ces régions.

Nous avons effectué une comparaison objective entre les quatre méthodes en utilisant les mesures de qualités introduites dans le chapitre 3. Les images de fusion produites par notre méthode ont montrés en général les meilleures performances en termes d'information mutuelle et des performances compétitives en termes de différence de contraste (voir tableau 3). Concernant le critère basé sur la couleur, Photomatix et la méthode proposée dans le chapitre 3 sont meilleures. Nous concluons que la méthode proposée ici préserve mieux les détails et introduit une petite distorsion de la couleur.



Figure 4-3: (a) Poids estimés et leur utilisation pour la fusion de différentes scènes. A : les poids estimés à partir des scènes « Eiffel » et « landscape ». (b) Scènes fournies par [Photomatix] (c) : notre acquisition (d) : Sebastien Negraszus.



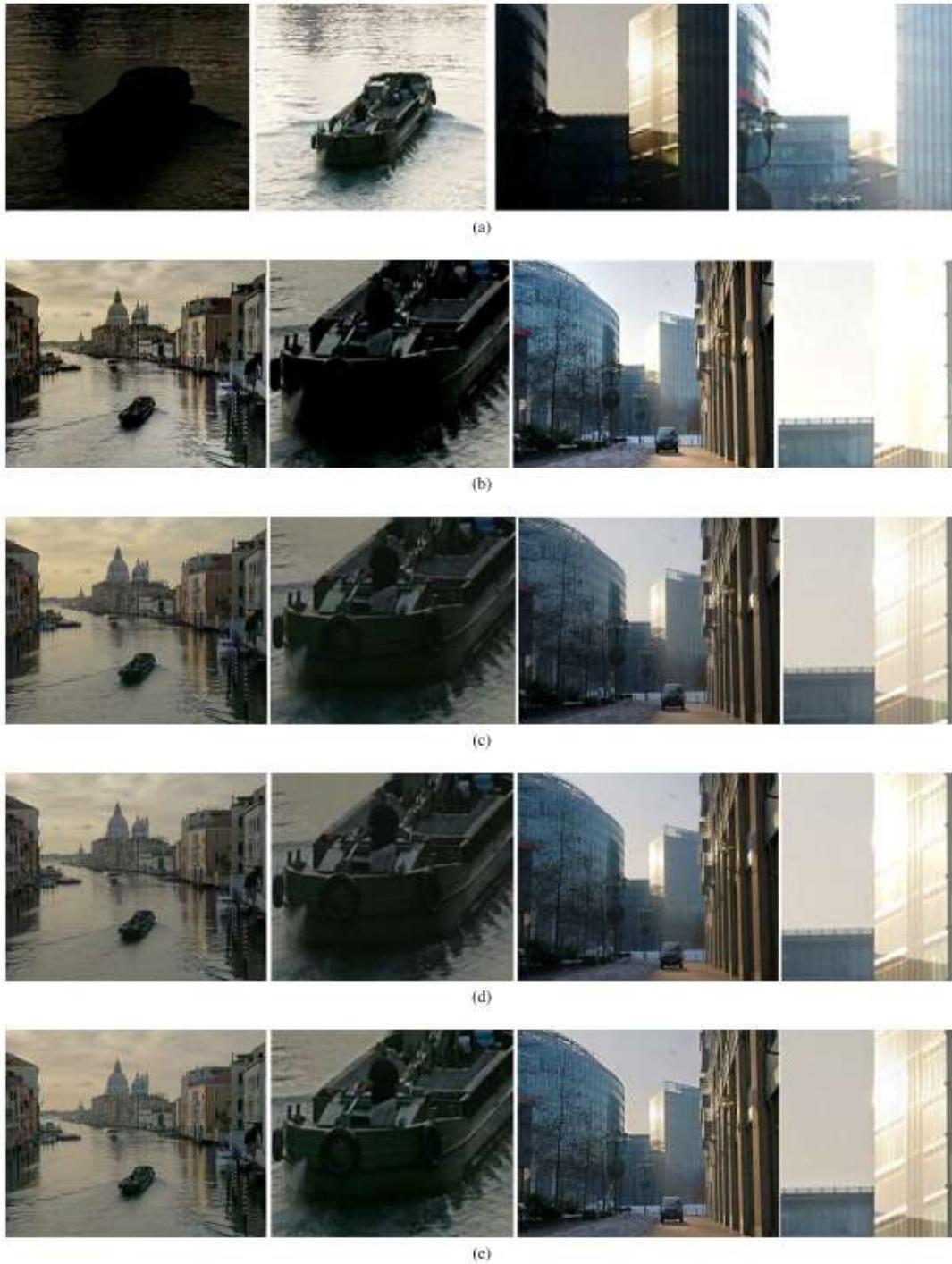


Figure 4-4: Comparaison entre les quatre méthodes. (a) : Parties des entrées. (b): fusion d'exposition de Mertens (c): Méthode proposée dans le chapitre 3. (d) Fusion d'exposition avec la méthode proposée dans 3.1. (e) Fusion d'exposition de Photomatrix.

Metrique	Méthodes	Venise	Eiffel	LightHouse	Landscape
Contraste	Mertens	1,15	0,85	0,32	1,53
	Méthode 1	1,05	0,71	0,27	1,34
	Photomatix	1,13	0,72	0,3	1,4
	Méthode2	1,11	0,7	2,28	1,45
Couleur	Mertens	1,56	2,89	1,17	1,26
	Méthode 1	0,99	1,99	1,1	1,18
	Photomatix	0,94	1,79	1,63	1,57
	Méthode2	1,15	2,3	1,62	1,95
IM	Mertens	2,934	3,42	2,779	4,55
	Méthode 1	6,69	7,46	5,57	7,73
	Photomatix	5,2	6,29	6,3	6,88
	Méthode2	6,48	8,33	7,35	8,83

Tableau 3 : Comparaison objective. Différence de contraste (magnitude de gradient), Différence de couleur (CIEDE2000) et l'information mutuelle.

Le tableau4 présente une comparaison en terme de temps d'exécution. Les trois algorithmes comparés sont implémentés avec le même langage (Matlab) et testés sur la même machine (2GHZ de CPU et 3Go de mémoire). Les méthodes de fusion d'exposition basées sur la phase de pondération comme l'algorithme de Mertens sont lourds. La méthode que nous proposons est très rapide car cette phase est éliminée complètement. Les poids sont définis au préalable dans une phase d'apprentissage et puis utilisés par la suite avec de simples opérations arithmétiques pour produire les images de fusion.

	Venise	Eiffel	Lighthouse	Landscape
Résolution	640 × 480	530 × 795	1024 × 682	1800 × 1198
Mertens	2,1	2,5	3,4	10,4
Méthode 1	0,3	0,5	1,1	2,4
Méthode 2	0,1	0,2	0,4	1,2

Tableau 4 : Temps d'exécution en secondes des méthodes de fusion d'expositions

4.4 Discussion des résultats

Nous avons proposé une technique très rapide et efficace pour la fusion d'expositions basée sur l'estimation de poids généralisés. La définition de ces poids au préalable permet la fusion de n'importe quel ensemble d'expositions sans passer par une phase de pondération, ce qui aboutit à un système de fusion d'exposition très rapide.

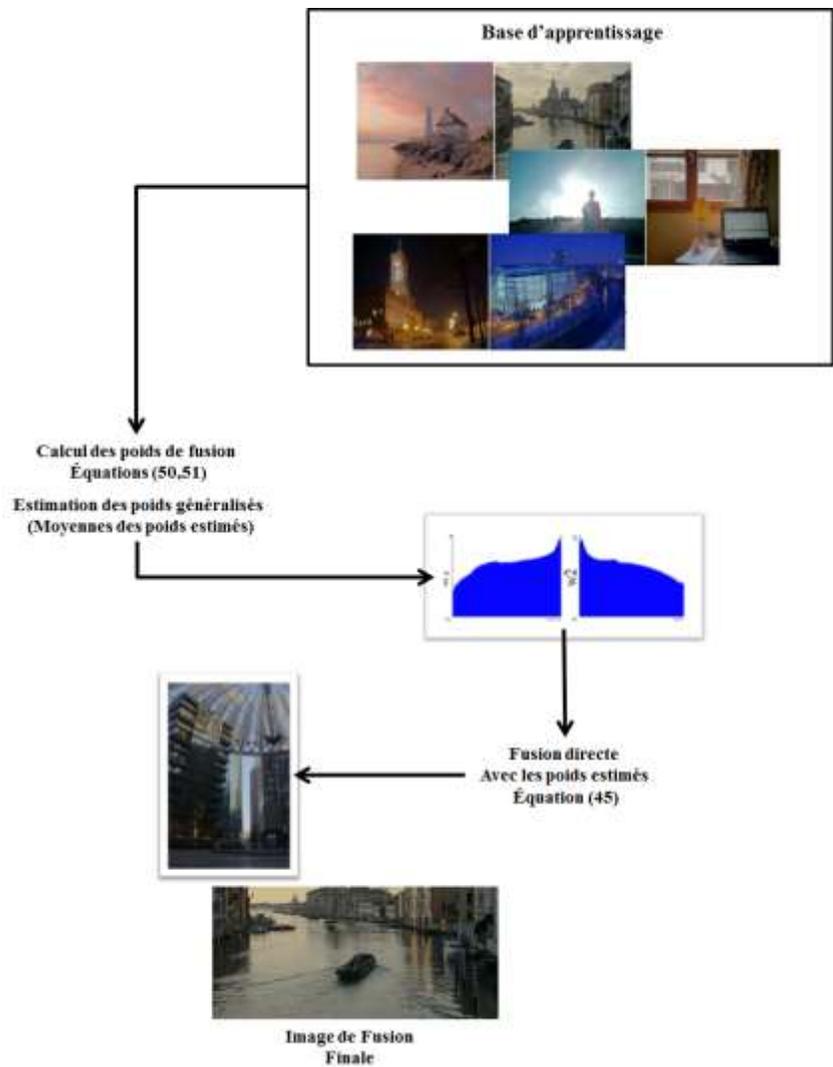


Figure 4-5 Organigramme de la fusion d'exposition basée sur les poids généralisés

4.5 Estimation de modèle de fusion d'expositions

Dans les sections suivantes nous proposons une autre méthode de fusion d'exposition basée sur l'apprentissage supervisé. A partir de la collection d'apprentissage on cherche cette fois-ci un modèle de régression qui s'ajuste sur les données. On considère qu'une méthode de fusion d'expositions est une fonction à multiples variables M qui prend en entrée les valeurs d'expositions (x_1, x_2, \dots, x_j) et produit en sortie la valeur de fusion (f). Pour la production d'images à grande plage dynamique, M consiste

à créer la carte des radiances puis effectuer l'algorithme de reproduction de tons. Quand l'algorithme de Mertens est utilisé, M devient l'estimation de plusieurs mesures de qualité (contraste, saturation exposition) sur chaque pixel puis la fusion multi échelle est effectuée pour aboutir à f . Dans notre travail nous évitons toutes ces opérations en supposant qu'il y ait une relation entre l'exposition et les valeurs de fusion que l'on peut retrouver par des analyses de régression sur la collection d'images. Nous cherchons le modèle qui s'ajuste sur l'ensemble des observations, où chaque observation i représente les pixels correspondants dans j expositions et leur fusion $(x_{i,1}, x_{i,2}, \dots, x_{i,j}, f_i)$. Nous commençons l'expérimentation en estimant un modèle linéaire (ML) qui contient $j + 1$ paramètres :

$$ML_i = a_0 + \sum_{k=1}^j a_k \times x_{i,k} \dots \dots (52)$$

Avec a_0, a_1, \dots, a_j , sont les paramètres recherchés. Nous considérons cette fois ci 3 expositions pour la fusion ($j=3$). Une image sous exposée, une surexposée et une exposition moyenne. La fusion d'images en couleur est visée, donc le processus d'apprentissage est effectué sur les trois canaux de l'image séparément. Nous avons expérimenté les deux espaces de couleur RGB et l'espace YCrCb. Un modèle de régression est estimé pour chaque canal séparément. Trois collections d'images sont utilisées pour l'estimation des modèles. Elles sont construites par trois méthodes de fusion d'exposition différentes : la première est l'algorithme de Mertens. La deuxième est construite par le software Photomatix. La troisième collection est construite à l'aide de la méthode introduite par [Debevec , 1997]. Le tableau5 montre les paramètres estimés du modèle linéaire (ML) sur ces trois collections d'images. Les paramètres sont estimés par la minimisation de :

$$\text{Min} \sum_i (f_i - a_0 - \sum_{k=1}^j a_k \times x_{i,k})^2 \dots \dots (53)$$

L'erreur quadratique moyenne (EQM) est calculée pour témoigner de la fiabilité de l'ajustement du modèle. Aussi, l'index de similarité structurelle (SSIM) est calculé pour mesurer la similarité entre les images fusionnées dans la collection d'images et celles fusionnées à l'aide du modèle estimé. Comme on peut le voir dans le tableau5, la luminance et la chrominance ne s'ajustent pas sur le modèle de régression de la même manière. Les expérimentations ont montré que l'estimation du modèle est meilleure quand la luminance et la chrominance sont ajustées séparément (dans l'espace de couleur YCrCb). L'espace RGB produit de plus grandes erreurs et donc un modèle de fusion de moindre qualité. Les résultats montrent aussi qu'un simple modèle linéaire est suffisant pour la fusion de chrominance. D'un autre côté, l'estimation du modèle linéaire sur la luminance n'est pas aussi bonne. De ce fait nous avons essayé un modèle non linéaire (MNL) de fusion de la luminance :

$$\text{MNL}_i = \sum_{k=1}^j (a_k \times x_{i,k}) + (a_{k+j} \times x_{i,k}^2) \dots \dots (54)$$

Nous avons choisi un polynôme de degré deux car il est simple à estimer et il produit des résultats satisfaisants. L'estimation des paramètres du modèle MNL est faite à l'aide de l'algorithme de Levenberg-Marquardt. Une nette amélioration est perçue comparée au modèle linéaire pour les trois collections d'images. La qualité de l'estimation du modèle diffère d'une collection d'image à une autre. Celle utilisant la fusion de Photomatix semble être la meilleure collection d'apprentissage, elle montre une meilleure qualité d'ajustement que les autres collections. Ceci est confirmé par la figure4-5 qui montre les images de fusion de chaque méthode et images fusionnées à l'aide des modèles estimés. L'objectif est de pouvoir utiliser les modèles de fusion d'exposition estimés pour éviter les calculs effectués habituellement pour chaque méthode de fusion. Les

expérimentations ont montrés que le modèle estimé sur les données de Photomatix est le plus fiable. La luminance est fusionnée à l'aide du modèle non linéaire (MNL) montré dans le tableau5. Cependant, le modèle linéaire est suffisant pour la fusion de chrominance. La figure4-6 montre la fusion d'expositions issues de scènes différentes de celles utilisées pour l'apprentissage. On en conclut que le modèle de régression estimé des données de Photomatix est valide pour la fusion d'expositions.

	Channel	Paramètres	EQM	SSIM
Photomatix	Y(ML)	19,67 0,28 0,26 0,38	77,62	0,9
	Y(MNL)	-0,36 0,02 1,06 0,0013 0,0025 - 0,0024	25,62	0,96
	Cb(ML)	-42,94 0,17 0,74 0,41	2,05	0,99
	Cr(ML)	-18,41 0,05 0,87 0,21	2,09	0,99
	R(ML)	19,15 0,25 0,29 0,41	103,91	0,9
	G(ML)	26,07 0,29 0,31 0,32	131,52	0,89
	B(ML)	21,97 0,13 0,42	150,34	0,9
Mertens	Y(ML)	10,27 0,01 0,25 0,41	366,59	0,8
	Y(MNL)	-1,31 0,50 0,86 0,0053 0,0005 0,0016	285,19	0,85
	Cb(ML)	-87,17 0,83 0,09 0,75	3,42	0,98
	Cr(ML)	-77,78 0,85 0,10 0,65	4	0,99
	R(ML)	-0,44 0,30 0,02 0,56	479,37	0,73
	G(ML)	10,55 0,03 0,28 0,35	598,46	0,74
	B(ML)	7,32 -0,23 0,43 0,30	549,13	0,74
HDR (Debevec)	Y(ML)	42,24 0,18 0,09 0,30	948,74	0,64
	Y(MNL)	-1,64 0,40 1,43 0,0072 0,0013 - 0,0042	704,18	0,78
	Cb(ML)	-101,64 1,09 0,11 0,57	6,03	0,98
	Cr(ML)	-78,74 0,75 0,37 0,48	9,68	0,98
	R(ML)	34,99 0,47 -0,06 0,42	162400	0,61
	G(ML)	48,87 0,09 0,23 0,18	127500	0,62
	B(ML)	35,69 -0,30 0,25 0,28	101800	0,61

Tableau 5 : Estimation des modèles de régression à partir de trois collections d'images

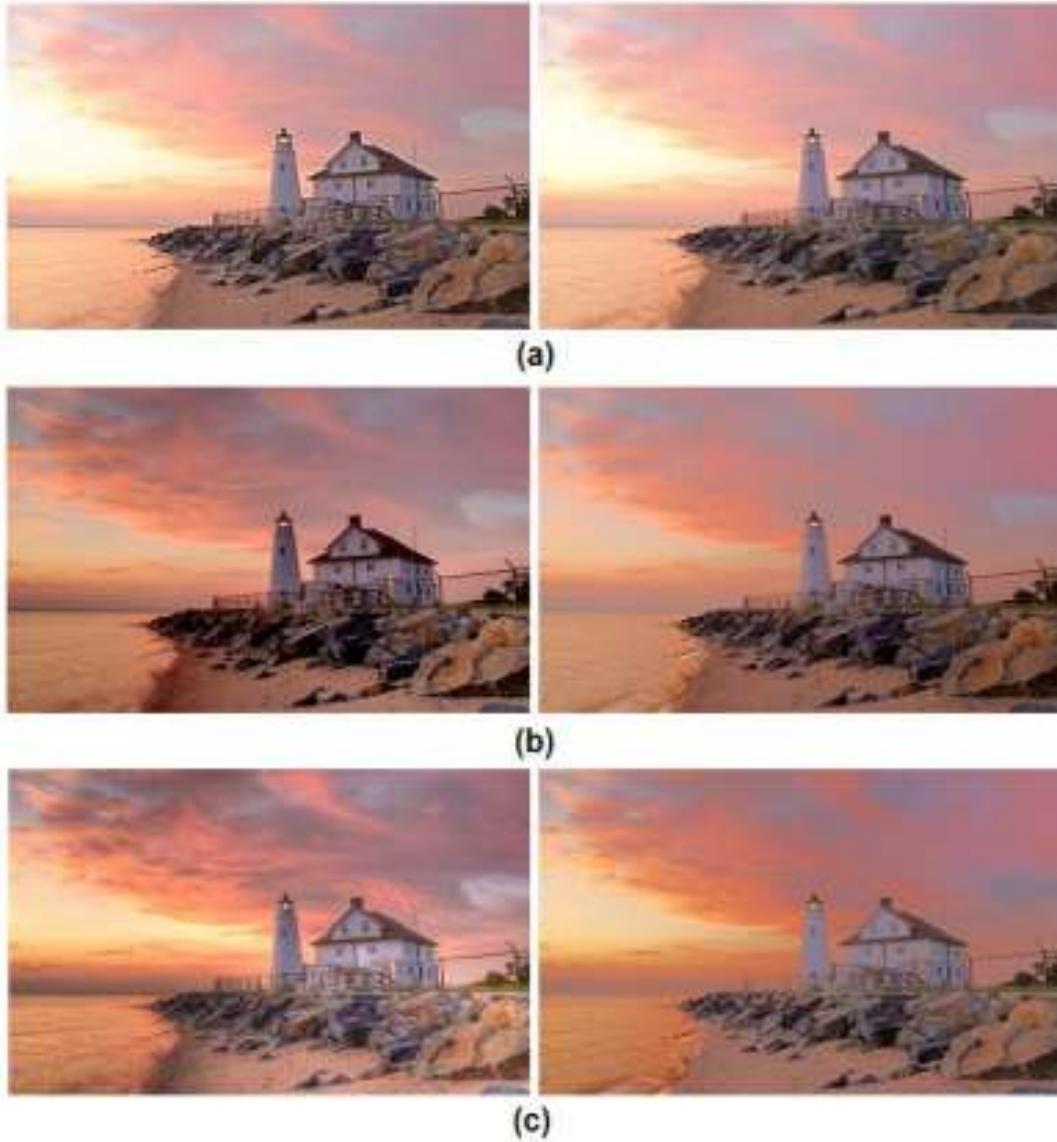


Figure 4-6: Estimation des modèles de régression. (Gauche) Méthodes originales, (Droite) Fusion utilisant les modèles estimés. (a) Photomatix software. (b) Algorithme de Mertens (c) HDR + reproduction de tons.



Figure 4-7: Fusion d'expositions à l'aide du modèle estimé de la collection d'images Photomatix.

4.6 Expérimentation et comparaisons

Nous avons comparé la qualité des images issues de la fusion utilisant le modèle estimé avec les quatre autres méthodes : méthode de fusion de Mertens, de Photomaix ainsi que les deux méthodes que nous avons déjà proposées. La méthode 1 (fusion d'exposition basée sur la segmentation d'histogrammes) et la méthode 2 (estimation de poids généralisés pour la fusion d'expositions). La figure4-7 montre les résultats obtenus par les différentes méthodes. Les trois méthodes que nous avons proposées ainsi que la fusion de Photomatix se ressemblent et apportent plus de détails que l'algorithme de Mertens. Le tableau6 expose les comparaisons objectives grâce aux mesures de qualités introduites précédemment.



Metrique	Méthodes	Venise	Eiffel	LightHouse	Landscape
Contraste	Mertens	1,15	0,85	0,32	1,53
	Méthode 1	1,05	0,71	0,27	1,34
	Photomatix	1,13	0,72	0,3	1,4
	Méthode 2	1,11	0,7	2,28	1,45
	Méthode 3	1,12	0,71	0,29	1,44
Couleur	Mertens	1,56	2,89	1,17	1,26
	Méthode 1	0,99	1,99	1,1	1,18
	Photomatix	0,94	1,79	1,63	1,57
	Méthode 2	1,15	2,3	1,62	1,95
	Méthode 3	0,98	1,86	1,59	1,55
IM	Mertens	2,934	3,42	2,779	4,55
	Méthode 1	6,69	7,46	5,57	7,73
	Photomatix	5,2	6,29	6,3	6,88
	Méthode 2	6,48	8,33	7,35	8,83
	Méthode 3	5,71	7,15	5,93	7,5

Tableau 6 : Comparaison entre les différentes méthodes à l'aide du contraste, la couleur et l'information mutuelle.

Le tableau suivant montre le temps de traitement de chaque méthode. On en conclue que les méthodes basées sur l'apprentissage supervisé sont très rapides comparées aux autres méthodes. Ceci est dû au fait qu'il n'y ait aucune opération à faire concernant la pondération de pixel. Il s'agit de fusionner directement les expositions à l'aide de poids de fusion, ou modèle estimé au préalable.

	Venise	Eiffel	Lighthouse	Landscape
Résolution	640 × 480	530 × 795	1024 × 682	1800 × 1198
Mertens	2,1	2,5	3,4	10,4
Méthode 1	0,3	0,5	1,1	2,4
Méthode 2	0,1	0,2	0,4	1,2
Méthode 3	0,1	0,3	0,5	1,7

Tableau 7 : Temps d'exécution des différentes méthodes de fusion d'expositions (Implémentation Matlab)



Figure 4-8: les résultats des trois méthodes : Mertens, notre méthode, Photomatix.



Figure 4-9: Organigramme de la fusion d'exposition basée sur un modèle de régression.

4.7 Conclusion

Nous avons présenté deux méthodes de fusion d'expositions basée sur un apprentissage supervisé sur une collection d'expositions et leur fusion. Nous avons estimé les poids généralisés de fusion d'exposition qui ont permis d'obtenir une fusion d'image de qualité. Nous avons aussi dégagé un modèle de régression qui reproduit le processus de fusion dans l'espace de couleur YCrCb. Ces techniques permettent une fusion de n'importe quelles expositions sans passer par les opérations habituelles de pondération grâce à des mesures de qualité. Les résultats obtenus grâce aux deux méthodes ont montré que la qualité des images de fusion obtenues est comparable à celle produite par Photomatix ou bien l'algorithme de Mertens.

Partie III

Chapitre 5 : Détection précise de lignes sur les images.

5 Détection précise de lignes sur les images

Ce chapitre explique à travers la section 5.1 une difficulté rencontrée par la transformée de Radon classique, quand elle est utilisée pour la détection de lignes sur les images. Puis, dans la section 5.2 nous proposons l'utilisation d'une variante de la transformée de Radon qui remédie à cette difficulté et permet d'obtenir une détection de lignes plus précise. Celle-ci s'appelle la transformée de Radon Gaussienne. Nous présentons les comparaisons entre l'utilisation de la transformée de Radon classique et Gaussienne pour la détection des lignes dans la section 5.3. Nous concluons à la fin de ce chapitre dans la section 5.4.

5.1 Introduction

Le motif de ligne est présent dans presque toutes les scènes. Son extraction de l'image est une étape clé pour de nombreuses applications de vision par ordinateur, telles que le système d'avertissement de sortie de voie. Dans ce chapitre, nous introduisons une généralisation de la transformée de Radon qui fournit une amélioration appréciable de la précision de détection de ligne. Nous l'appelons transformée de radon gaussienne.

L'une des approches les plus populaires pour détecter les lignes en général est basée sur la transformation de Radon, en gardant à l'esprit que la transformation de Hough est une forme discrète de celle-ci [Nourine, 2004], [Nourine, 2006a], [Nourine, 2013], [Nourine, 2006b]. Cette approche échantillonne l'image en lignes candidates, puis les intégrales sur celles-ci fournissent un espace de paramètres où les lignes pertinentes sont pointées par des pics. Mais en pratique, les lignes d'image sont sur un domaine discret et elles sont formées par des pixels qui ne sont pas nécessairement colinéaires. Ce fait rend le processus d'échantillonnage de transformation de Radon brut, conduisant à des pics biaisés dans l'espace des paramètres.

5.2 L'imprécision de la transformée de Radon

Dans la majorité des scènes naturelles, on retrouve des structures linéaires qui se transforment en lignes droites dans les images. Ces lignes droites sont des caractéristiques intéressantes qui jouent un rôle important dans la reconnaissance d'objets et la compréhension d'images. La transformée de Radon est une technique classique très utilisée pour la détection de ces caractéristiques particulières. Sachant que la transformée de Hough est la version discrète de la transformée de Radon.

Avec la transformée de Radon, les lignes sont représentées en utilisant l'équation suivante :

$$\rho = x \cos \theta + y \sin \theta \dots \dots (55)$$

Le paramètre ρ est la distance perpendiculaire entre la ligne et l'origine de l'image (en haut à gauche), et θ est l'angle formé par la perpendiculaire. Avec cette représentation, comme on peut le voir sur la figure 5-1, une ligne verticale comme la ligne 1 a un angle θ ayant la valeur de zéro. Alors qu'une ligne horizontale comme la ligne 5 a un angle θ ayant la valeur de $\pi/2$. La ligne 3 a un angle $\theta = \pi/4$.

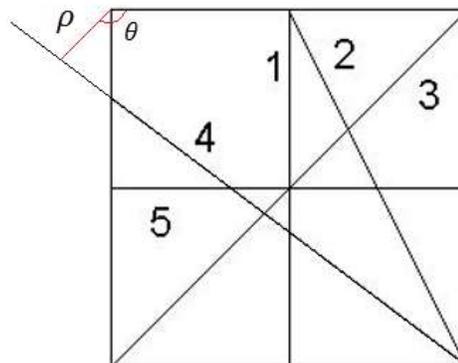


Figure 5-1 : Représentation des lignes pour la transformée de Radon.

La transformée de Radon tente d'échantillonner l'image en des lignes candidates, puis les intégrales sur les lignes produit un espace de paramètres la ou les lignes les plus prononcées sont pointées par des piques. Théoriquement, le processus d'échantillonnage repose sur la multiplication du signal image par des impulsions de Dirac. La Dirac produit une réponse pour les points de l'image qui satisfont les paramètres recherchés. Celle-ci sélectionne des points colinéaires sur l'image.

Pour résumer, la transformée de Radon transforme les points du plan des coordonnées en un espace de paramètres de fonction définie par l'expression $f(X, a) = 0$. Où X est le vecteur contenant les des coordonnées des points et a représente un vecteur constant contenant les paramètres de la fonction recherchée. La fonction recherchée dans notre cas est une ligne droite donc :

$$f(X, a) = \rho - x \cos \theta + y \sin \theta \dots \dots (56)$$

Dans ce cas, $a = (\rho, \theta)$ et $X = (x, y)$. La transformée de Radon de l'image $I(x, y)$ pour ce modèle est :

$$Radon(\rho, \theta) = \iint I(x, y) \delta(f(X, a)) dx dy \dots \dots (57)$$

L'impulsion de Dirac 2D sélectionne les points qui appartiennent à la ligne investie par les paramètres a . L'expression $\delta(f(X, a))$ réagit comme une mesure binaire qui théoriquement prend une valeur nulle pour $f(X, a)$ pour que le point X puisse être sélectionné et ajouté à la somme (l'intégrale). Cependant, le domaine de l'image est discret et les pixels formant les lignes ne sont pas nécessairement colinéaires. Une valeur nulle de f n'est pas garantie pour tous les pixels formant une ligne. La figure5-1 montre comment se passe l'échantillonnage de l'image avec la transformée de Radon. Il s'agit de sélectionner des lignes dans un fond noir en utilisant différentes tolérances à la mesure $f(X, a) = 0$. Avec $|f(X, a)| < 0.1$, les lignes sélectionnées ne sont pas représentatives

et sont à peine formées. Dans les implémentations courantes tel que la transformée de Hough sous OpenCv, les pixels sont sélectionnés après un simple arrondi de la valeur $f(X, a)$. Ce qui veut dire que $|f(X, a)| < 0.5$ est utilisée pour sélectionner les lignes. Dans la transformée de Radon de Matlab, les pixels sont divisés en quatre sous-pixels pour aboutir à un vote plus précis sur les paramètres [Mathworks].



Figure 5-2: L'échantillonnage de la transformée de Radon en utilisant différentes tolérances à l'expression $f(X,a)=0$. $|f(X,a)|<0.1$ pour la ligne la plus mince, $|f(X,a)|<1$ pour la ligne épaisse.

5.3 Transformée de Radon Gaussienne

Les solutions existantes supposent donc que les lignes sont minces et formées par des pixels colinéaires. Cependant, les lignes dans une image peuvent être formées par des bandes de pixels pouvant avoir différentes épaisseurs. Cette épaisseur peut dépendre de la résolution de l'image, la nature des lignes, la perspective. Dans les images binaires, les pixels formant une ligne ne sont pas tous colinéaires.

Il est nécessaire durant l'échantillonnage de la transformée de Radon d'utiliser une fonction qui prend en considération ces faits, pour aboutir à une meilleure sélection des lignes candidates. Il existe une panoplie de fonctions qui peuvent être utilisées à la place de la fonction de Dirac. Nous cherchons la fonction qui assure une détection précise des lignes à la fin de la transformée de Radon. Par détection précise des lignes nous voulons dire la détection du centre de la bande qui représente une ligne dans l'image.

La figure5-3 montre la convolution d'une image contenant une ligne avec des filtres Gaussiens. On peut voir que le filtre Gaussien a une nature qui lisse les lignes et produit des maxima au centre de la bande tant que la taille du filtre Gaussien est relativement large par rapport à l'épaisseur de la ligne. La convolution avec un filtre Gaussien tend à tirer les maxima vers le milieu même si le profil de la bande est asymétrique.

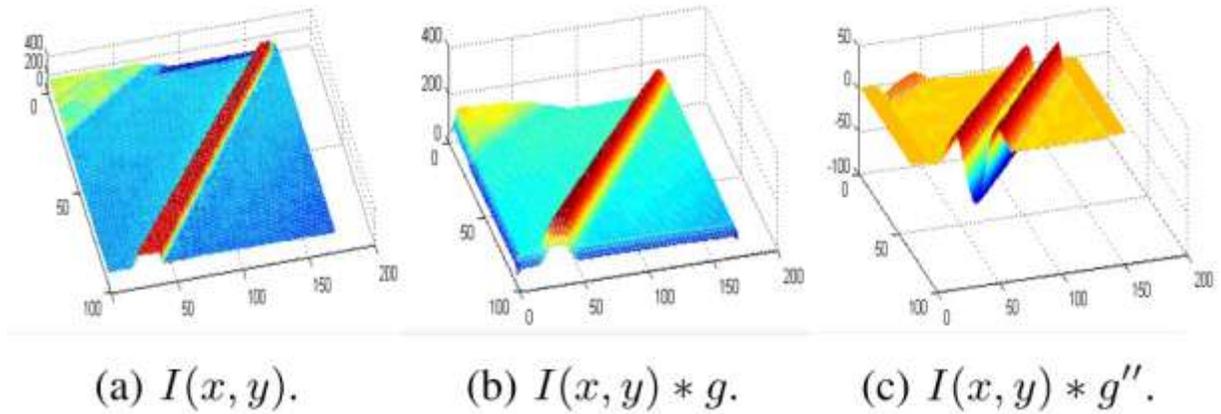


Figure 5-3:Le comportement d'une ligne quand l'image est convoluée avec un filtre Gaussien g et sa deuxième dérivée g'' .

Sachant que la corrélation et la convolution sont des opérations équivalentes pour les filtres symétriques, le changement de l'impulsion de Dirac dans l'équation56 par un filtre Gaussien symétrique permettra de bénéficier des bonnes propriétés de convolution citées en haut. L'échantillonnage qui résulte de ce changement considèrera la dispersion des pixels formant une ligne en produisant des maxima au milieu. Nous formulons donc une transformée de Radon lisse :

$$SmoothRadon(\rho, \theta) = \iint I(x, y) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{f(x, \rho)^2}{2\sigma^2}} dx dy \dots \dots \dots (58)$$

Au lieu que l'échantillonnage soit fait avec l'impulsion de Dirac, à l'aide de la fonction, l'échantillonnage est effectué avec une fonction Gaussienne qui respecte mieux les spécificités des lignes dans les images. La largeur de la Gaussienne est gérée par le paramètre σ .

5.4 Résultats et comparaisons

La figure 5-4 montre les espaces des paramètres obtenus par les deux méthodes : transformée de Radon et transformée de Radon Gaussienne. On peut voir clairement que la fonction Gaussienne produit une transformation lisse avec un maximum précis, alors que la Dirac produit un problème de sélection de piques.

Nous avons fait une comparaison entre la transformée de Radon et la version lisse proposée sur deux applications différentes : la première est l'extraction de réseaux routiers dans des images satellitaire de GoogleMap, après un simple seuillage Otsu. La deuxième application est la détection des marquages routiers. Les figures 5-5 et 5-6 montrent que l'utilisation de la transformée de Radon lisse produit plus de précision pour la détection des lignes face au bruit. La ligne verte étant la vérité terrain, la ligne rouge est la ligne détectée par la méthode.

Nous avons aussi fait une comparaison quantitative à l'aide d'une vérité terrain construite manuellement. Il s'agit d'une base de données qui contient les images et le positionnement manuel des lignes. Le tableau 8 montre la précision de chaque méthode en terme de variance entre la vérité terrain et les lignes détectées. Nous concluons aussi que la transformée de Radon Gaussienne produit une nette amélioration dans la précision concernant la détection des lignes, qui est relative à l'épaisseur de la ligne recherchée.

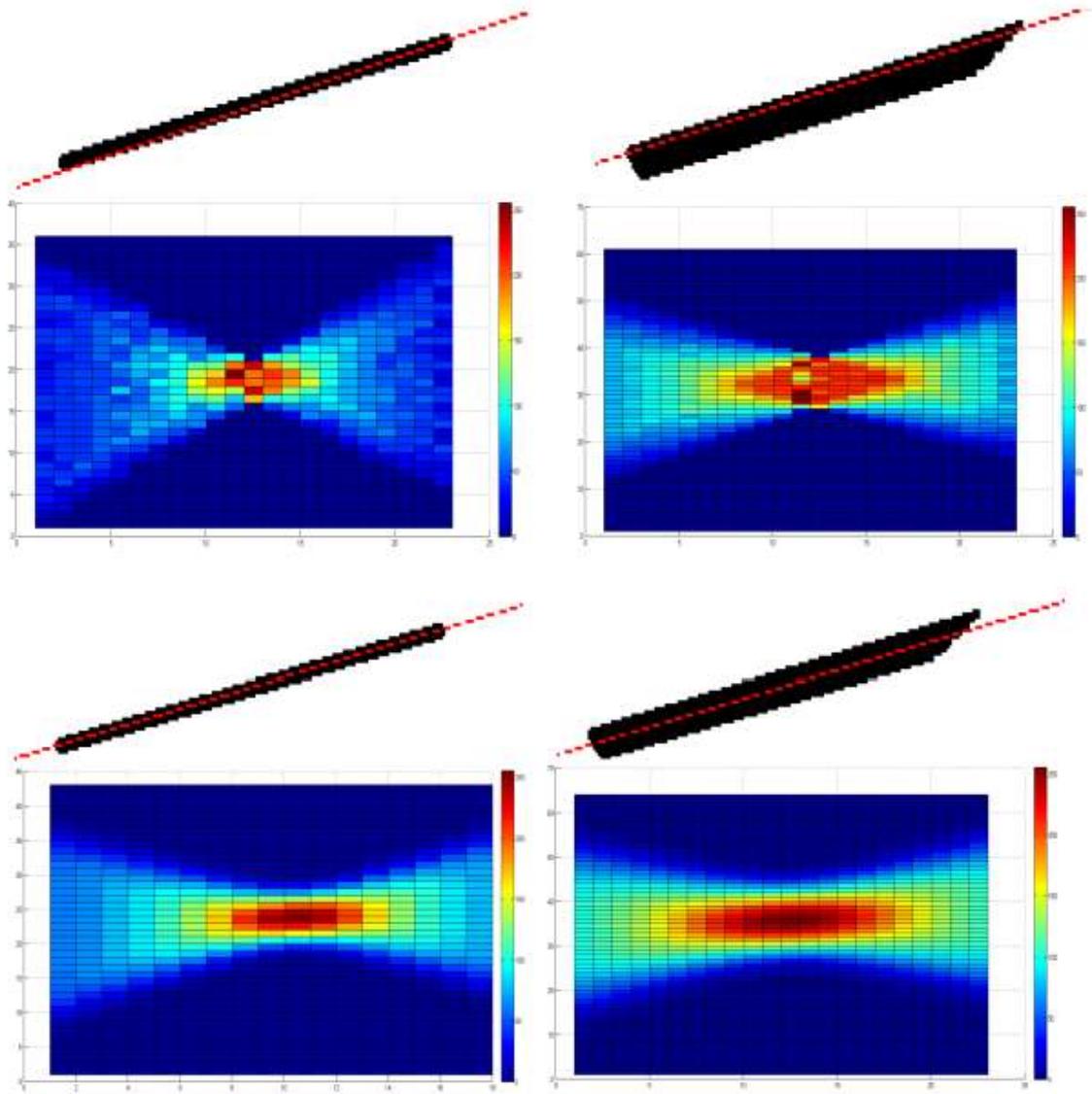


Figure 5-4: Détection de lignes. (Haut) : Transformée de Radon. (Bas) : Transformée de Radon Gaussienne.

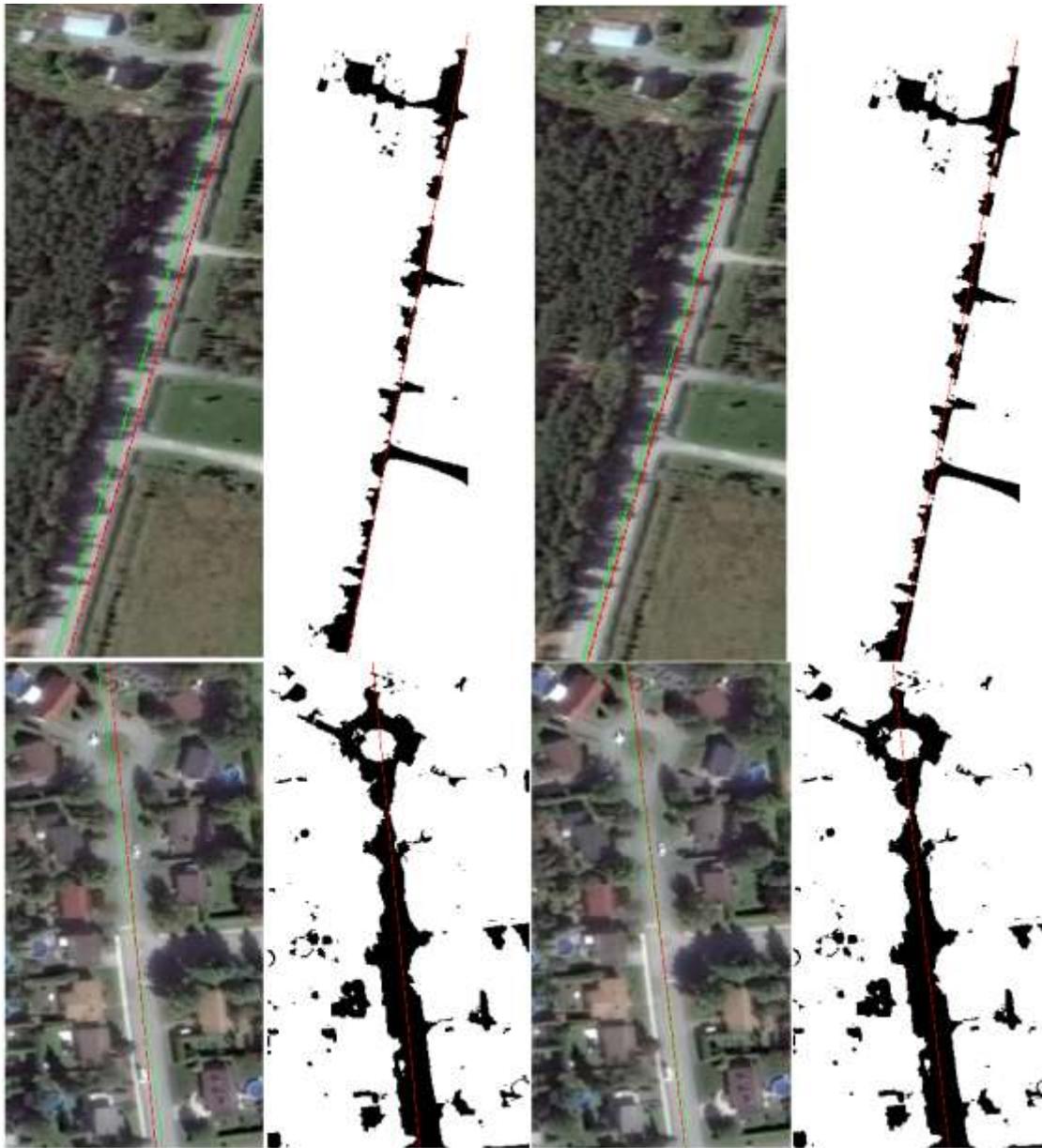


Figure 5-5: Extraction de réseau routier. (Gauche) : Transformée de Radon. (Droite) : Transformée de Radon Gaussienne. Les lignes vertes sont la vérité terrain, en rouge les détection des méthodes.

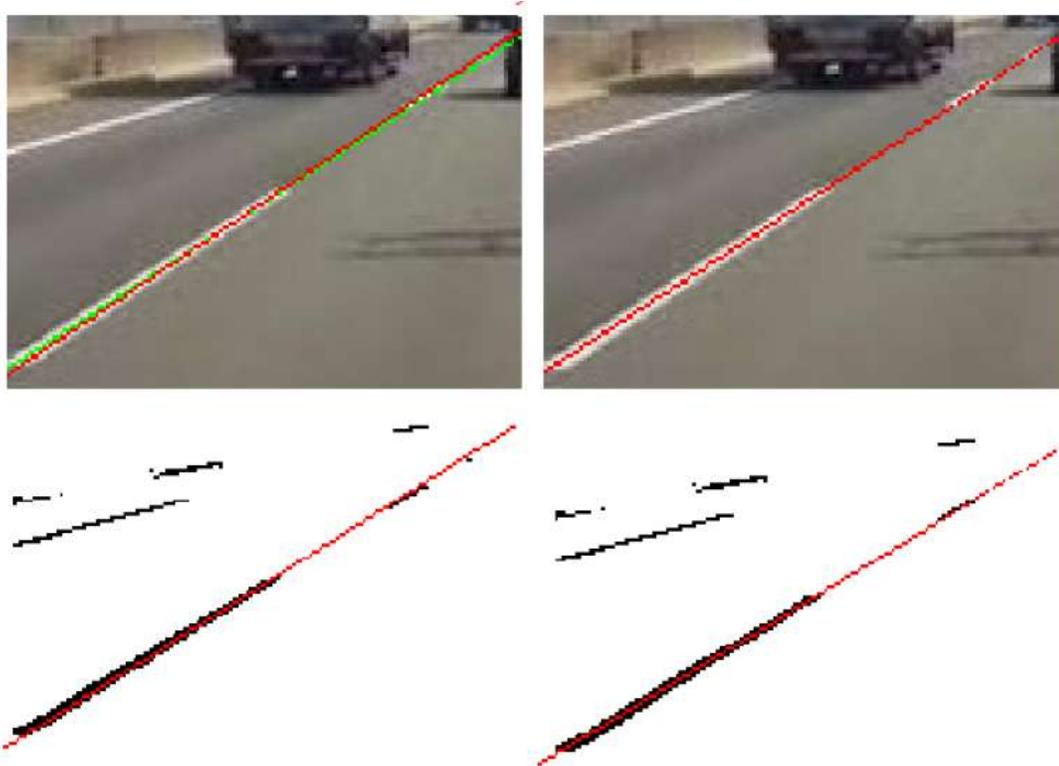


Figure 5-6: Détection de lignes de route (Gauche) : Transformée de Radon. (Droite) : Transformée de Radon lisse. Les lignes vertes sont la vérité terrain, en rouge les détections des méthodes.

5.5 Conclusion

Dans ce chapitre nous avons introduit une transformée de Radon Gaussienne pour palier au problème de l'imprécision face aux lignes épaisses ainsi qu'au bruit. Des expérimentations sur des images synthétiques et réelles ont montrées que les lignes retournées par la méthode proposée sont précises. L'amélioration en précision est significative pour plusieurs applications de vision par ordinateur tel que l'extraction de réseau routier d'images satellitaires, ou bien la détection des marquages routiers pour la création de systèmes d'assistance à la conduite.

	Largeur de la ligne	Vérité terrain		Transformée de Radon			Radon Gaussienne		
		Theta	Rho	Théta	Rho	Erreur	Théta	Rho	Erreur
Extraction du réseau routier	16	9	174	8	173	0,83	8	174	0,2
	18	-7	100	-14	100	7	-10	100	0,63
	24	5	173	3	175	5,04	4	174	1,2
Détection de marquages routiers	2	-14	76	-14	76	0	-14	76	0
	5	-17	59	-17	58	0,47	-17	59	0
	17	-4	55	-6	57	2,84	-4	55	0

Tableau 8 : Comparaison entre la transformée de Radon et sa version Gaussienne qu'on propose pour la détection des lignes.

IV - Conclusion générale et perspectives

L'aide à la conduite est un domaine qui a attiré énormément d'efforts de recherche durant les deux dernières décennies. L'enjeu est très important car il s'agit non seulement d'offrir une aide et un confort durant la conduite, mais surtout de diminuer le nombre et la gravité des accidents. De nombreux systèmes d'aide à la conduite sont déjà présents dans les voitures d'aujourd'hui mais peu d'entre eux sont complètement fiables. Ceux qui souffrent le plus de robustesse sont les systèmes basés sur des capteurs passifs, notamment la caméra. Dans cette thèse nous avons proposé des techniques de traitement d'image et de vision par ordinateur visant à améliorer la robustesse des systèmes d'aide à la conduite basés sur les caméras. Dans la première partie de la thèse nous avons développé des méthodes de fusion d'expositions très rapides qui permettent de capturer des images correctes face aux scènes à grande plage dynamique. Cet outil permettra à un système d'aide à la conduite basé sur une caméra d'obtenir les informations nécessaires à son fonctionnement face à des situations particulièrement difficiles à la capture, tel que la conduite face au soleil, la sortie d'un tunnel, des scènes rendues compliquées à cause de la distribution de la lumière (ombres, zones ensoleillées)...etc.

Durant la deuxième partie de la thèse, nous avons démontré que l'utilisation d'une transformée de Radon Gaussienne apporte beaucoup plus de précision que la transformée de Radon classique durant la détection des marquages routiers. Ceci est vrai aussi pour la détection des lignes qui apparaissent épaisses sur les images telles que des vaisseaux sanguins dans des images médicales ou bien un réseau routier dans des images satellitaires.

La transformée de Radon produit un espace de paramètres où les lignes à forte intensité sont pointés par des grandes valeurs d'accumulateur. Durant la phase de pondération des paramètres, les propriétés géométriques entre les lignes ne sont pas prises en considération. Pour la détection des marquages routiers par exemple, on sait que les lignes recherchées s'intersectent au point de fuite à cause de l'effet de perspective. Nous travaillons donc sur la recherche d'une généralisation de la transformée de Radon qui permet d'utiliser les propriétés géométriques des lignes à détecter. Nous voulons que la configuration géométrique des lignes puisse influencer la phase de pondération des paramètres.

V - Table des figures

Figure 2-1: Scène à grande plage dynamique. Impossibilité de capturer tous les détails de la scène dans une seule photographie.....	22
Figure 2-2 Acquisition face à une scène à grande plage dynamique. (Haut) : Capture automatique de la caméra. (Bas) Capture avec la méthode de fusion d'expositions proposée dans le chapitre 3.	22
Figure 1-1: Mesurer la lumière avec un luminance mètre.....	29
Figure 1-2: Les différentes intensités lumineuses retrouvées dans une scène réelle.	29
Figure 1-3: Formation du signal image.	30
Figure 1-4: Photographies capturées face à une scène à grande plage dynamique.	31
Figure 1-5: Imagerie à grande plage dynamique et fusion d'exposition directe.	33
Figure 1-6: Un exemple de capture d'expositions et obtention d'une image GPD. Les expositions sont prises à différents temps d'exposition. (a) 1/250 sec. (b) 1/30 sec. (c) ¼ sec. (d) 2 sec. (e) 8sec. (f) Est un rendu de la luminance de l'image GPD obtenue. [Debevec , 1997]	33
Figure 1-7: Diagramme montrant les différentes sources de bruit ainsi que les près traitement typiques [Richard Szeliski, 2010].....	35
Figure 1-8: Fusion d'images différemment exposées à l'aide d'une décomposition laplacienne et d'une pyramide gaussienne des cartes de poids, qui représentent les mesures de contraste, saturation et exposition [Mertens , 2007].	48
Figure 3-1: Segmentation d'images en régions avec l'algorithme d'Otsu.	61
Figure 3-2: Les trois types de régions qu'on retrouve dans un jeu d'expositions.	62
Figure 3-3: Histogrammes des expositions.	63
Figure 3-4: Représentation des zones recherchées par des Gaussiennes.....	65
Figure 3-5: Segmentation d'histogrammes. (Bleu) : Zones sombres, (Rouge) : zones claires, (Vert) : zones intermédiaires.	66
Figure 3-6: Reconstruction de l'image de fusion avec les régions segmentées. (a) sans aucun ajustement. (b) : Ajustement proposé. (c) Egalisation d'histogrammes.	68
Figure 3-7: Capture automatique d'exposition pour la fusion. (a,b) : Zones sombres et claires acquises avec notre algorithme. (c,d) : Fcam acquisition. (temps d'exposition, gain) : a(1666,1), (b)(40000,1.58),c(2397,1),d(40000,6.68).	69
Figure 3-8: Résultats obtenus par les trois méthodes : (Gauche) Algorithme de Mertens, (Milieu) notre méthode, (Droite) : Photomatix.....	70
Figure 3-9: Résultats obtenus par les trois méthodes : (Gauche) Algorithme de Mertens, (Milieu) notre méthode, (Droite) : Photomatix.....	70
Figure 3-10: Organigramme de la méthode fusion d'expositions basée sur la segmentatin d'histogrammes.	73
Figure 3-11: L'apport de la fusion d'exposition face à une scène de conduite à grande plage dynamique. (Haut) Algorithme original de la caméra. (Bas) Image résultante de la fusion d'expositions.	74
Figure 4-1: Estimation des poids à partir de collection d'images.	80

Figure 4-2: Estimation des poids de luminance à partir de la scène « Eiffel ». (Gauche) collection d'images basée sur l'algorithme de Mertens. (Droite) : collection d'image basée sur la fusion d'exposition de Photomatix.	81
Figure 4-3: (a) Poids estimés et leur utilisation pour la fusion de différentes scènes. A : les poids estimés à partir des scènes « Eiffel » et « landscape ». (b) Scènes fournies par [Photomatix] (c) : notre acquisition (d) : Sebastien Negraszus.....	83
Figure 4-4: Comparaison entre les quatre méthodes. (a) : Parties des entrées. (b): fusion d'exposition de Mertens (c): Méthode proposée dans le chapitre 3. (d) Fusion d'exposition avec la méthode proposée dans 3.1. (e) Fusion d'exposition de Photomatix.	84
Figure 4-5 Organigramme de la fusion d'exposition basée sur les poids généralisés	87
Figure 4-6: Estimation des modèles de régression. (Gauche) Méthodes originales, (Droite) Fusion utilisant les modèles estimés. (a) Photomatix software. (b) Algorithme de Mertens (c) HDR + reproduction de tons.	92
Figure 4-7: Fusion d'expositions à l'aide du modèle estimé de la collection d'images Photomatix.	93
Figure 4-8: les résultats des trois méthodes : Mertens, notre méthode, Photomatix.	95
Figure 4-9: Organigramme de la fusion d'exposition basée sur un modèle de regression.	96
Figure 5-1 : Représentation des lignes pour la transformée de Radon.	99
Figure 5-2: L'échantillonnage de la transformée de Radon en utilisant différentes tolérances à l'expression $f(X,a)=0$. $ f(X,a) <0.1$ pour la ligne la plus mince, $ f(X,a) <1$ pour la ligne épaisse.	101
Figure 5-3:Le comportement d'une ligne quand l'image est convoluée avec un filtre Gaussien g et sa deuxième dérivée g''	102
Figure 5-4: Détection de lignes. (Haut) : Transformée de Radon. (Bas) : Transformée de Radon Gaussienne.....	104
Figure 5-5: Extraction de réseau routier. (Gauche) : Transformée de Radon. (Droite) : Transformée de Radon Gaussienne. Les lignes vertes sont la vérité terrain, en rouge les détection des méthodes.	105
Figure 5-6: Détection de lignes de route (Gauche) : Transformée de Radon. (Droite) : Transformée de Radon lisse. Les lignes vertes sont la vérité terrain, en rouge les détection des méthodes.....	106

VI - Liste des tableaux

Tableau 1 : Comparaison objective : Contraste (différence en magnitude de gradient), Couleur (CIEDE2000 [Sharma , 2005]) et l'information mutuelle [Guihong , 2002]	72
Tableau 2 : Temps de fusion d'exposition en secondes (Implémentation Matlab) et pourcentage des zones segmentées.....	73
Tableau 3 : Comparaison objective. Différence de contraste (magnitude de gradient), Différence de couleur (CIEDE2000) et l'information mutuelle.....	86
Tableau 4 : Temps d'exécution en secondes des méthodes de fusion d'expositions.....	87
Tableau 5 : Estimation des modèles de régression à partir de trois collections d'images.....	92
Tableau 6 : Comparaison entre les différentes méthodes à l'aide du contraste, la couleur et l'information mutuelle.....	95
Tableau 7 : Temps d'exécution des différentes méthodes de fusion d'expositions (Implémentation Matlab).....	96
Tableau 8 : Comparaison entre la transformée de Radon et sa version Gaussienne qu'on propose pour la détection des lignes.....	108

VII– Bibliographie

[Adam, 2010] Andrew Adams, Eino-ville Talvala, Sung Hee Park, David E. Jacobs, Boris Ajudin, Natasha Gelfand, Jennifer Dolson, Daniel Vaquero, Jongmin Baek, Marius Tico, Hendrik P. A. Lensch, Wojciech Matusik, Kari Pulli, Mark Horowitz, Marc Levoy: “The frankencamera: An experimental platform for computational photography”, In: SIGGRAPH, 2010.

[Alexandre, 2009] Alexandre BARGETON, “Fusion multi sources pour l’interprétation de l’environnement routier”, 2009

[Alon , 2006] Yaniv Alon, A Ferencz, A Shashua: “Off-road path following using region classification and geometric projection constraints”. In: IEEE Conference on Computer Vision and Pattern Recognition. (pp. 689–696), 2006.

[Alvarez , 2007] José Manuel Alvarez, Antonio Lopez, Ramon Baldrich: “Shadow resistant road segmentation from a mobile monocular system”. In: Iberian conference on Pattern Recognition and Image Analysis, (pp. 9–16), 2007.

[Ashikhmin , 2002] Michael Ashikhmin : “A Tone Mapping Algorithm for High Contrast Images”, 13th Eurographics workshop on Rendering, (pp. 145-156), 2002.

[Bilcu , 2008] Radu Ciprian Bilcu, Adrian Burian, A Knuutila, M Vehvilainen: “High dynamic range imaging on mobile devices”. In: 15th IEEE International Conference on Electronics, Circuits and Systems, ICECS 2008, (pp. 1312–1315), August 2008.

[Bilcu , 2010] Alain Hore, Djemel Ziou: “Image quality metrics: Psnr vs. ssim”. Pattern Recognition (ICPR), 20th International Conference on, pp. (pp. 2366 – 2369), 2010.

[Bloch, 2003] Bloch I : “Fusion d’informations en traitement du signal et des images”. Lavoisier, Hermes Science Publication, 2003.

[Borkar , 2009] Amol Borkar, Monson Hayes, Mark Smith: “Robust lane detection and tracking with RANSAC and Kalman filter”. In: International Conference on Image Processing. (pp. 3261–3264), 2009.

[Borkar , 2012] Amol Borkar, Monson Hayes: “A novel lane detection system with efficient ground truth generation”. IEEE transactions on intelligent transport systems, vol 13, (pp. 365–374), 2012.

[Broggi , 2006] Alberto Broggi and Stefano Cattani: “An agent based evolutionary approach to path detection for off-road vehicle guidance”. Pattern Recognition Letters (pp. 1164–1173), 2006.

[Chapiro , 2011] Alexandre Chapiro, Marcelo Cicconet, and Luiz Velho: “Filter based deghosting for exposure fusion video,” in SIGGRAPH, 2011.

[Cheng , 2006] Hsu Yung Cheng, Bor Jeng, Pei Tseng, KC Fan: “Lane detection with moving vehicles in the traffic scenes”. IEEE Transactions on Intelligent Transportation Systems, (pp. 571-582), 2006.

[Chiu , 1993] Ken Chiu, Michael Herf, Peter Shirley, S Swamy, Changyaw Wang, Kurt Zimmerman: “Spatially nonuniform scaling functions for high contrast images”. In Proceedings of Graphics Interface 93, Canadian Information Processing Society, (pp. 245–245), 1993.

[Cualain , 2011] Diarmaid O Cualain, C Hughes, Martin Glavin, Edward Jones: “Automotive standards-grade lane departure warning system”. IEEE transactions on intelligent transport systems, vol 6, (pp. 4–57), 2011.

[Dean, 1995] Pomerleau Dean: “RALPH: Rapidly adapting lateral position handler”. In: IEEE Intelligent Vehicles Symposium, 1995.

[Debevec , 1997] Paul Debevec and Jitendra Malik : “Recovering High Dynamic Range Radiance Maps from Photographs.”. In SIGGRAPH '97 Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, (pp. 369–378), 1997.

[Fattal , 2002] Raanan Fattal, Dani Lischinski, Michael Werman: “Gradient Domain High Dynamic Range Compression”. 29th annual conference on Computer graphics and interactive techniques, 2002.

[Gelfand , 2010] Natasha Gelfand, N Adams, A Park, S Pulli: “Multi-exposure imaging on mobile devices”. In: MM 2010 Proceedings of the International Conference on Multimedia, (pp. 823–826), 2010.

[Goshtasby., 2005] Ardeshir Goshtasby: “Fusion of multi-exposure images”. Image Vision Comput., 23 :6, (pp. 611–618), June 2005.

[Guihong , 2002] Qu Guihong, Dali Zhang, Yan Pingfan: “Information measure for performance of image fusion”. Electronics Letters 38:7, (pp. 313–315), 2002.

[Hernandez, 2009] Jorge Hernandez, Beatriz Marcotegui: “Filtering of artifacts and pavement segmentation from mobile lidar data”. Paris Tech, 2009.

[Huang, 2009] Albert S Huang, D Moore, M Antone, E Olson, S Teller: “Finding multiple lanes in urban road networks with vision and LIDAR”. Autonomous Robots, (pp. 1011-1020), 2009.

[Jiang , 2010] Yan Jiang, F Gao,G Xu: “Computer vision-based multiple lane detection on straight road and in a curve”. In: ImageAnalysis and Signal Processing., (pp. 114–117), 2010.

[Jinno , 2012] Takao Jinno and M Okuda: “Multiple exposure fusion for high dynamic range image acquisition”. IEEE TRANSACTIONS ON IMAGE PROCESSING, 21:1, (pp. 358–365), January 2012.

[Kang , 2003] Sing Bing Kang, Matt Uyttendaele, Simon Winder, and Richard Szeliski: “High dynamic range video,” ACM Transactions on Graphics, 22:3, (pp. 319–325), July 2003.

[Kao, 2007] Wen Chung Kao: “Real-time image fusion and adaptive exposure control for smart surveillance systems”. Electronics Letters 43:18, (pp. 975–976), 2007.

[Kao., 2007] W Kao, C Hsu, C Kao: “Real-time image fusion and adaptive exposure control for smart surveillance systems,”. Electronics Letters, 43:18, (pp. 975 – 976), August 2007.

[Katramados , 2009] Ioannis Katramados, Steve Crumpler and Toby P. Breckon: “Real-time traversable surface detection by colour space fusion and temporal analysis”. In: Computer Vision Systems. (pp. 265–274), 2009.

[Kim, 2008] Zuwhan Kim: “Robust lane detection and tracking in challenging scenarios”. IEEE Transactions on Intelligent Transportation Systems 9, (pp. 16–26), 2008.

[Kong , 2009] Hui Kong, Jean-Yves Audibert, Jean Ponce: “Vanishing point detection for road detection”. In: IEEE Conference on Computer Vision and Pattern Recognition., (pp. 96–103), 2009.

[Kornhauser , 2007] Alain L Kornhauser: “DARPA Urban Challenge Princeton University technical paper”. <http://www.stanford.edu/~jmayer/papers/darpa07.pdf> , 2007.

[Labayrade, 2006] Raphalel Labayrade, J Douret, J Laneurit, R Chapuis: “A reliable and robust lane detection system based on the parallel use of three algorithms for driving safety assistance”. IEICE - Transactions on Information and Systems E89-D (pp. 2092–2100), 2006.

[Lauffenburger , 2007] Lauffenburger Jean Philippe : “Navigation as a virtual sensor for enhanced lighting preview control”. In IEEE Intelligent Vehicles Symposium, (pp. 2069-2074), 2007.

[Ledda , 2004] Patrick Ledda, Luis Paulo Santos, Alan Chalmers: “A Local Model of Eye Adaptation for High Dynamic Range Images”. AFRIGRAPH '04 Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualization and interaction in Africa, (pp. 754-769), 2004.

[Li , 2012] Shutao Li, Xudong Kang: “Fast multi-exposure image fusion with median filter and recursive filter”. IEEE Transactions on Consumer Electronics 58:2, (pp. 626–632), 2012.

[Mann , 1995] Steve Mann and Rosalind W. Picard: “Being ‘Undigital’ with Digital Cameras: Extending Dynamic Range by Combining Differently Exposed Pictures.” In Proceedings of IS&T 48th Annual Conference, (pp. 422–428), 1995.

[Mathworks] “Mathworks radon transform.” [Online]. Available: <http://www.mathworks.com/help/images/ref/radon.html>

[McCall , 2006] JC McCall, M Trivedi: “Video-based lane estimation and tracking for driver assistance”: Survey, system, and evaluation. IEEE Transactions on Intelligent Transportation Systems 7 (pp. 20–37), 2006.

[Mertens , 2007] Tom Mertens, Juan Kautz, and Frank Van Reeth : “Exposure fusion”. Computer Graphics and Applications, Pacific Conference on, (pp. 382–390), 2007.

[Michon, 1985] Michon, John.: “A critical view of driver behavior models. What do we know, what should we do?” In L. Evans & R. Schwing (Eds.), Human behavior and traffic safety, (pp. 485-521), 1985.

[Mitsunaga , 1999] Tomoo Mitsunaga and Shree K. Nayar: “Radiometric Self Calibration.”. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (168–210), 1999.

[Mobileye] Mobileye homepage. (<http://www.mobileye.com/manufacturer-products/brochures>)

[Moumene , 2014a] Moumene Mohammed Elamine, Nourine Rachid , Ziou, Djemel: “Fast Exposure Fusion Based on Histograms Segmentation”. ICISP 2014, (pp. 367–374), 2014.

[Moumene , 2014b] Moumene Mohammed Elamine, Nourine Rachid , Ziou, Djemel : “Generalized Exposure Fusion Weights Estimation”. Canadian Conference on Computer and Robot Vision, (pp. 71-76), 2014.

[Moumene , 2017] Moumene Mohammed Elamine, Nourine Rachid. “Fast Exposure Fusion Based on Regression Analysis”. International Journal of Imaging and Robotics, 17:4, (pp. 89-96), 2017.

[Negraszus , 2010] S Negraszus and D Pirch: “High dynamic range imaging and tonemapping”. [Online]. Available: <http://cybertron.cg.tu-berlin.de/pdci09>

[Nieto , 2008] Marcos Nieto, L Salgado, F Jaureguizar, J Arrospide: “Robust multiple lane road modeling based on perspective analysis. In: International Conference on Image Processing”. (pp. 2396–2399), 2008.

[Nieto , 2008] Marcos Nieto, L Salgado, F Jaureguizar, J Arrospide: “Robust multiple lane road modeling based on perspective analysis”. In: International Conference on Image Processing, (pp. 2396–2399), 2008.

[Nooij , 2011] Margriet van Schijndel-de Nooij, Bastiaan Krosse, Thijs van den Broek, Sander Maas, Ellen van Nunen, Han Zwijnenberg: “Definition of necessary vehicle and infrastructure systems for automated driving”. Smart 2010/0064 Study Report, 2011.

[Nourine, 2004] R. Nourine, M. Elarbi Boudihir, S. Khelifi: “Application of Radon Transform to Lane Boundaries Tracking”, Lecture Notes in Computer Science, Springer, ISBN: 3-540-23223-0, (pp. 563-571), 2004.

[Nourine, 2006a] Rachid Nourine, M Elarbi. Boudihir, N. Benamrane, S. F. Khelifi: “Lane boundaries following to driver assistance system”. 8eme Colloque Africain sur la Recherche en Informatique, (pp. 350-360), 2006.

[Nourine, 2006b] Rachid Nourine, M Elarbi. Boudihir, N. Benamrane, S. F. Khelifi: “Vision-based Lane Boundaries Following for Autonomous Vehicle Guidance”, 18th National Computer Conference, 2006.

[Nourine, 2013] Rachid NOURINE and Mohamed Elamine MOUMENE: “Genetic Algorithm Based Road Marking Detection” International Conference on signal, Image, Vision and Their Application, (pp. 21-25), 2013.

[Ogawa , 2006] Takashi Ogawa and Kiyokazu Takagi: “Lane recognition using on-vehicle lidar”. IEEE Intelligent Vehicles Symposium (pp. 540–545), 2006.

[Paris , 2009] Sylvain Paris and Frédo Durand : “A fast approximation of the bilateral filter using a signal processing approach”. Int. J. Comput. Vision, 81:1, (pp. 24–52), Jan 2009.

[Photomatix] hdrsoft, <http://www.hdrsoft.com>

[Polychronopoulos et al, 2005] Polychronopoulos Aris: “Extended path prediction using camera and map data for lane keeping support”. In IEEE Intelligent Transportation System ITS, (PP. 166-171), 2005.

[Rasmussen , 2005] Christopher Rasmussen, Thommem Korah: “On-vehicle and aerial texture analysis for vision-based desert road following”. In: CVPR workshop on machine vision for intelligent vehicles, (pp. 66–66), 2005.

[Reinhard , 2002] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda: “Photographic tone reproduction for digital images”. In ACM Transactions on Graphics (TOG), volume 21, (pp. 267–276). 2002.

[Revue et al, 2002] Revue A., Wybo S. et Nashashibi F.: “Multisensor based collision risk assessment: a case of crossroads approach”. In IEEE Intelligent Vehicles Conference, (pp. 448-452), 2002.

[Rezaei , 2014] Mahdi Rezaei and Mutsuhiro Terauchi: “Dataset intercity roads and adverse driving scenarios.”, 2014.

[Richard Szeliski, 2010] Richard Szeliski. Computer Vision: “Algorithms and Applications”. 2010.

[Robertson , 1999] Mark A. Robertson, Sean Borman, and Robert L. Stevenson: “Dynamic Range Improvement Through Multiple Exposures.” In Proceedings of the 1999 International Conference on Image Processing (ICIP-99). Los Alamitos, CA, USA: IEEE, (pp. 159–163), 1999.

[Robertson , 2003] Mark A. Robertson, Sean Borman, and Robert L. Stevenson: “Estimation Theoretic Approach to Dynamic Range Enhancement Using Multiple Exposures.” Journal of Electronic Imaging, 12:2, (pp. 219–228), 2003.

[Samadzadegan , 2006] Farhad Samadzadegan, A Sarafraz, M Tabibi: “Automatic lane detection in image sequences for vision based navigation purposes”. In: ISPRS Image Engineering and Vision Metrology, (pp. 462-475), 2006.

[Sawano , 2006] Hiroaki Sawano, M Okada: “A road extraction method by an active contour model with inertia and differential features”. IEICE - Transactions on Information and Systems E89-D (pp. 2257–2267), 2006.

[Schlick , 1995] Christophe Schlick: “Quantization techniques for visualization of high dynamic range pictures”. In Georgios Sakas, Stefan Mller, and Peter Shirley, editors,

Photorealistic Rendering Techniques, Focus on Computer Graphics, pages (pp. 7–20), 1995.

[Sharma , 2005] Gaurav Sharma, Wencheng Wu, Dalal, Edul: “The color-difference formula: Implementation notes, supplementary test data, and mathematical observations”. *Color Research & Application*, 30:1, (pp. 21–30), 2005.

[Steger, 2013] Carsten Steger: “Unbiased extraction of lines with parabolic and gaussian profiles” *Computer Vision and Image Understanding*, vol 117, (pp. 97–112), February 2013.

[Tardif , 2009] JP Tardif: “Non-iterative approach for fast and accurate vanishing point detection,” in 2009 IEEE 12th International Conference on Computer Vision, (pp. 1250-1257), 2009.

[Tarel , 2002] Jean Philippe Tarel, Seng, P Charbonnier: “Using robust estimation algorithms for tracking explicit curves”. In: *ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part I*, London, UK, Springer-Verlag (pp. 492–507), 2002.

[Tico , 2010] Marius Tico, Natasha Gelfand, Pulli: “Motion blur free exposure fusion”. In: *IEEE International Conference on Image Processing (ICIP)*, (pp. 26–29), September 2010.

[Tumblin , 1993] Jack Tumblin and Holly Rushmeier: “Tone reproduction for realistic images”. *Computer Graphics and Applications*, IEEE, 13:6 , (pp. 42–48), 1993.

[Wang , 2007] Tsun-Hsien Wang, Wei-Su Wong, Fang-Chu Chen, Fang-Chu Chen: “Design and Implementation of a Real-Time Global Tone Mapping Processor for High Dynamic Range Video”. *International Conference on Image Processing - ICIP* , (pp. 209-212), 2007.

[Ward , 1997] Gregory Ward Larson, Holly Rushmeier, and Christine Piatko: “A visibility matching tone reproduction operator for high dynamic range scenes”. *Visualization and Computer Graphics*, IEEE Transactions on, 3:4, (pp. 291–306), 1997.

[Wu, S , 2008] S Wu, H Chiang, J Perng, C Chen, B Wu, T Lee: “The heterogeneous systems integration design and implementation for lane keeping on a vehicle”. IEEE Transactions on Intelligent Transportation Systems 9, (pp. 246–263), 2008.

[Yamaguchi, K , 2008] Koichiro Yamaguchi, Akihiro Watanabe, T Naito, Y Ninomiya: “Road region estimation using a sequence of monocular images”. In: International Conference on Pattern Recognition, (pp. 487-503), 2008.

[Zhang , 2012] Wei Zhang and Wai Kuen Cham : “Gradient-directed composition of multi-exposure images”. IEEE Conference on Computer Vision and Pattern Recognition, (pp. 530–536), 2010

[Zhang, G , 2009] Geng Zhang, Nanning Zheng, Chao Cui, Yuzhen Yan and Zejian Yuan: “An efficient road detection method in noisy urban environment”. In: IEEE Intelligent Vehicles Symposium, (pp. 654-673), 2009.

Publications scientifiques réalisées au cours de la thèse:

[Moumene , 2014a] Moumene Mohammed Elamine, Nourine Rachid , Ziou, Djemel: “Fast Exposure Fusion Based on Histograms Segmentation”. ICISP 2014, (pp. 367–374), 2014.

[Moumene , 2014b] Moumene Mohammed Elamine, Nourine Rachid , Ziou, Djemel : “Generalized Exposure Fusion Weights Estimation”. Canadian Conference on Computer and Robot Vision, (pp. 71-76), 2014.

[Moumene , 2017] Moumene Mohammed Elamine, Nourine Rachid. “Fast Exposure Fusion Based on Regression Analysis”. International Journal of Imaging and Robotics, 17:4, (pp. 89-96), 2017.

Résumé

L'aide à la conduite est devenue un domaine de recherche très actif et pluridisciplinaire ayant pour objectif d'apporter au conducteur le maximum de confort et de sécurité sur les routes. Les systèmes d'aide à la conduite les plus prometteurs sont souvent basés sur la vision. Parmi eux, on retrouve le système d'anti franchissement involontaire de la voie routière. A l'aide d'une caméra orientée vers l'avant, les marquages routiers sont détectés en appliquant la transformée de Radon sur les images. Cet algorithme de vision permet la localisation et le maintien du véhicule sur sa voie. Malgré qu'ils soient déjà commercialisés, les systèmes d'aide à la conduite basés sur la caméra restent peu fiables. Leur robustesse qui n'a pas arrêté de croître au fil des dernières années se heurte aux limites de la camera. Des limites dues à certaines conditions particulières de capture telles que les scènes à grande plage dynamique. Un exemple de ce genre de scènes problématiques serait de capturer des images face à un soleil éblouissant. La caméra étant incapable de restituer toute la grande plage dynamique de la scène en une seule photographie, elle fournit des images avec une perte considérable de détails. Afin de capturer tous les détails d'une scène à grande plage dynamique, les photographes appliquent des algorithmes de photographie computationnelle tels que la fusion d'expositions. Cependant, ce genre de processus est lourd pour la réalisation d'un système d'aide à la conduite nécessitant de fonctionner en temps réel. Nous proposons dans cette thèse trois méthodes originales de fusion d'expositions qui produisent des images riches en détails face aux scènes à grande plage dynamique, fonctionnant en temps réel. Ces méthodes sont non seulement rapides, mais elles produisent aussi une qualité d'images comparable à celle issue de méthodes pertinentes de la littérature. La première méthode est basée sur la segmentation d'histogrammes. Les deux autres sont basées sur un apprentissage supervisé. La transformée de Radon reste l'une des méthodes les plus utilisées pour la détection des marquages routiers sur les images. Cette technique est utilisée à la base pour la détection des structures linéaires dans la scène naturelle qui se transforment en lignes dans les images. Cependant, les marquages routiers sont présents dans les images sous forme de bandes (des lignes considérablement épaisses). Cette épaisseur des lignes à détecter engendre une certaine ambiguïté dans l'espace de recherche de la transformée de Radon classique et génère souvent une perte considérable de précision durant la détection. Nous avons proposé dans cette thèse l'utilisation d'une généralisation de la transformée de Radon afin d'achever une meilleure précision durant la détection des lignes routières. Cette généralisation consiste à exploiter les propriétés d'une fonction Gaussienne au lieu d'utiliser l'impulsion de Dirac pour la sélection des pixels, dans le but de prendre en considération l'épaisseur des marquages routiers à détecter.

Mots clés :

Imagerie à grande plage dynamique; Fusion d'expositions; Segmentation d'Histogrammes; Apprentissage supervisé; Détection de lignes; Transformée de Radon; Fonction Gaussienne.