

Fusion multimodale centrale RGB-polarimétrique pour l'analyse de scènes routières dans des conditions météorologiques dégradées *

L. Deregnacourt¹

H. Laghmar¹

A. Lechervy²

S. Ainouz¹

¹ Normandie Univ, INSA Rouen, UNIROUEN, UNIHAVRE, LITIS, 76000 Rouen, France

² Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, 14000 Caen, France

lucas.deregnacourt@insa-rouen.fr

Résumé

Les modèles de détection d'objets entraînés uniquement sur des images RGB souffrent d'un manque de robustesse face aux conditions météorologiques défavorables comme la pluie ou encore le brouillard. C'est pourquoi il est nécessaire d'ajouter d'autres capteurs non conventionnels. Nous proposons ainsi un schéma de fusion centrale entre les images RGB et polarimétriques basé sur l'architecture RetinaNet. Pour cela, nous utiliserons le jeu de données multimodal PolarLITIS.

Mots Clef

Détection d'objets, fusion multimodale, conduite autonome

Abstract

Object detection models trained only on RGB images suffer from a lack of robustness against adverse weather conditions such as rain or fog. This is why it is necessary to add other non-conventional sensors. We propose a central fusion scheme between RGB and polarimetric images based on the RetinaNet architecture. For this purpose, we will use the PolarLITIS multimodal dataset.

Keywords

Object detection, multimodal fusion, autonomous driving

1 Introduction

L'analyse de scènes routières est une étape importante dans le développement du véhicule autonome et des ADAS. La détection d'objets doit donc être fiable et robuste.

La fusion multimodale est ainsi devenue un centre d'intérêt majeur pour la communauté scientifique[3], les principales modalités explorées étant le LiDAR et l'infrarouge. Le problème du LiDAR, outre son coût, est qu'il peut être facilement bruité lorsque les conditions météorologiques se dégradent tandis que la caméra infrarouge est fortement impactée par le brouillard et la pluie.

Nous proposons de fusionner les images issues d'une caméra polarimétrique et celles issues d'une caméra RGB via un schéma de fusion centrale. Les caméras polarimétriques sont en effet très efficaces pour décrire les propriétés physiques des objets même lorsque la luminosité ou la visibilité est faible.

*Le premier auteur est financé par l'ANR et la Région Normandie dans le cadre du projet HAISCoDe.

Cet article s'inscrit dans la continuité des travaux de R. Blin dans lesquels les fusions précoce (*early fusion*) et tardive (*late fusion*) ont été explorées[1].

2 Méthode

2.1 Formalisme de l'imagerie polarimétrique

L'imagerie polarimétrique consiste à caractériser l'état de polarisation de la lumière réfléchie, c'est-à-dire la direction dans laquelle elle se déplace, pour chaque pixel de l'image. Une image polarimétrique s'obtient en orientant un polariseur à un certain angle, ne laissant ainsi passer qu'une partie de l'onde partiellement polarisée.

Les polariseurs sont communément orientés à 0°, 45°, 90° et 135° afin d'obtenir une image d'intensité composée de superpixels (Figure 1). Nous pouvons alors construire une image à 4 canaux $I = [I_0 I_{45} I_{90} I_{135}]^T$.

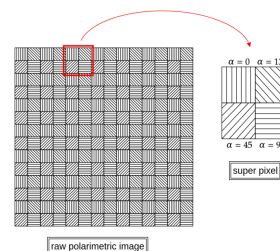


FIGURE 1 – Image d'intensité polarimétrique

Il en découle un deuxième encodage, nommé encodage de Stokes, défini comme suit :

$$S = \begin{bmatrix} I_0 + I_{90} \\ I_0 - I_{90} \\ I_{45} - I_{135} \end{bmatrix} = \begin{bmatrix} S_0 \\ S_1 \\ S_2 \end{bmatrix} \quad (1)$$

qui permet de caractériser l'énergie totale de l'onde (S_0) ainsi que les énergies partielles (S_1, S_2).

2.2 Architecture proposée

Nous nous basons sur l'architecture de RetinaNet[4], réputée pour sa capacité à détecter des objets de toute taille ce qui est un atout non négligeable lorsque l'on souhaite analyser des scènes routières.

Dans l'architecture proposée (Figure 2), nous extrayons dans un premier temps les cartes de caractéristiques pour

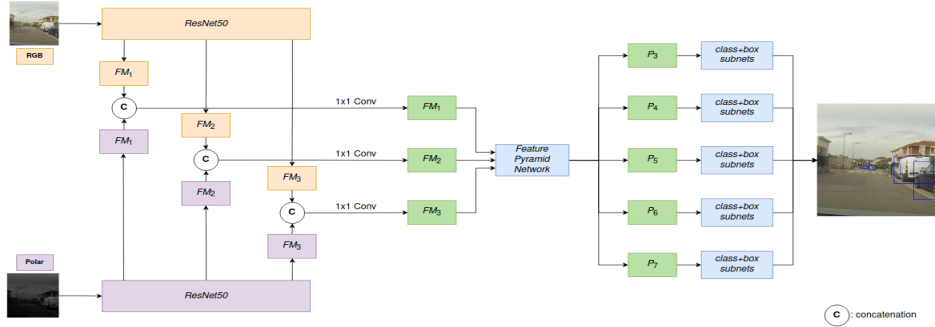


FIGURE 2 – Schéma de la fusion multimodale centrale proposée

chacune des modalités avec un backbone *ResNet50*. C'est ici que se déroule la fusion multimodale que nous proposons. Pour ce faire, ces cartes de caractéristiques de taille $H_i \times W_i \times C_i$ ($i \in \{1, 2, 3\}$) sont concaténées 2 à 2 pour former des cartes de caractéristiques de taille $W_i \times H_i \times 2C_i$. Une convolution de noyau 1×1 est ensuite appliquée pour retrouver des cartes de caractéristiques de taille $W_i \times H_i \times C_i$ contenant ainsi un mélange des cartes de caractéristiques de chacune des modalités. La fin du réseau proposé reprend la même architecture que RetinaNet à partir du *Feature Pyramid Network*.

3 Expérimentations

3.1 Présentation du jeu de données

Le jeu de données PolarLITIS[2] contient des images de scènes routières RGB et polarimétriques appariées. Étant donné que les classes *bike* et *motorbike* sont sous-représentées, nous ne conservons que les objets des classes *car* et *person*. Les propriétés de ce jeu de données sont résumées dans le Tableau 1.

	Train	Val	Test
<i>images</i>	1640	420	509
<i>car</i>	6061	2102	9265
<i>person</i>	527	134	442
<i>weather</i>	Sun+Cloud	Sun+Cloud	Fog

TABLE 1 – Caractéristiques de PolarLITIS

3.2 Protocole expérimental

Pour toutes les expériences, les modèles sont entraînés sur des données où les conditions météorologiques sont normales puis testés sur des données présentant du brouillard. Cela nous permet d'évaluer la robustesse des modèles face à la dégradation des conditions météorologiques. Nous testons notre méthode sur les images d'intensité polarimétrique (*I*), leur encodage de Stokes (*Stokes*) et les images RGB (*RGB*) que nous fusionnons 2 à 2.

3.3 Résultats

Comme attendu, les performances du modèle *RGB* se dégradent fortement lorsque nous l'évaluons sur des données où la visibilité est fortement réduite. Bien que les fusions *I+RGB* et *Stokes+RGB* atténuent cette dégradation, le

modèle unimodal *I* semble être le moins impacté. L'ensemble des résultats est résumé dans le Tableau 2.

		<i>I</i>	<i>Stokes</i>	<i>RGB</i>
MAP_{50}	val	80.44%	78.15%	84.87%
	test	76.03%	70.27%	45.31%
		<i>I+RGB</i>	<i>I+Stokes</i>	<i>Stokes+RGB</i>
MAP_{50}	val	80.32%	79.54%	77.90%
	test	66.72%	74.23%	65.88%

TABLE 2 – Résultats des modèles unimodaux et multimodaux

4 Conclusion et perspectives

Dans cet article, nous avons proposé une méthode de fusion multimodale centrale en tirant profit de la robustesse de l'imagerie polarimétrique face à la dégradation des conditions météorologiques. Bien que le meilleur modèle soit le modèle unimodal *I*, les fusions multimodales montrent des résultats prometteurs. Un des problème majeur de l'architecture proposée est que les poids de la fusion sont fixés à la fin de l'entraînement. La fusion s'effectue donc de manière identique, quelles que soient les conditions météorologiques rencontrées. Une piste d'amélioration serait donc de rendre cette fusion plus adaptative en fonction de la visibilité ou de la qualité des images.

Références

- [1] R. Blin, How polarimetry may contribute to deep road scene analysis in adverse weather conditions, Automatic Control Engineering, Normandie Université, 2021. English. NNT : 2021NORMIR20. tel-03588761
- [2] R. Blin, S. Ainouz, S. Canu, and F. Meriaudeau, A new multimodal rgb and polarimetric image dataset for road scenes analysis, *Conference on Computer Vision and Pattern Recognition Workshops*, pp. 867-876, 2020.
- [3] A. Gupta, , A. Anpalagan, L. Guan, A.S. Khwaja, Deep learning for object detection and scene perception in self-driving cars : Survey, challenges, and open issues, *Array*, 10 :100057, 2021.
- [4] T.-Y. Lin, , P. Goyal, R. Girshick, K. He, P. Dollár, Focal Loss for Dense Object Detection, *IEEE International Conference on Computer Vision*, pp. 2980–2988, 2017.