Projet d'analyse vidéo : comparasion de méthodes pour optical flow

Corentin Perdrizet
Yannis Chappet Juan

Table des matières

1 Introduction

Ce mini-rapport vise à comparer deux approches pour le calcul du flot optique dense : une méthode classique sans recours au deep learning et une méthode récente exploitant les réseaux de neurones profonds. Les performances seront évaluées à l'aide de métriques standards sur deux jeux de données : MPI Sintel, qui inclut une vérité terrain, et Grasping in the Wild, qui n'en dispose pas.

2 Données Utilisées

MPI Sintel

MPI Sintel est un jeu de données fournissant des séquences vidéo synthétiques, ainsi qu'une vérité terrain sous forme de fichiers au format .flo. Cela permet une évaluation quantitative précise à l'aide de métriques telles que l'End Point Error (EPE) et l'Angular Error (AE).

2.1 Grasping in the Wild (GITW)

Le jeu de données Grasping in the Wild (GITW) propose des séquences d'images utilisées pour des tâches liées à la vision robotique. Cependant, il ne fournit pas de vérité terrain pour le flot optique. Les performances seront donc évaluées en comparant l'image originale et l'image compensée à l'aide de la métrique Mean Square Error (MSE).

3 Méthodes Comparées

3.1 Méthode Classique

La méthode classique choisie pour ce projet est l'approche de Farneback, implémentée dans OpenCV. Cette méthode repose sur une représentation polynomiale des variations d'intensité entre deux images consécutives pour estimer un flot optique dense.

Elle calcule des approximations locales des dérivées de l'image et utilise ces informations pour produire un champ de déplacement continu. Les principaux paramètres de cette méthode incluent :

- **pyr_scale** : Facteur de réduction pour chaque niveau de la pyramide d'images.
- **levels** : Nombre de niveaux dans la pyramide.
- **winsize** : Taille de la fenêtre utilisée pour calculer les moyennes locales.
- **iterations** : Nombre d'itérations de l'algorithme.
- **poly_n** : Taille de la fenêtre pour les calculs polynomiaux.
- **poly_sigma** : Écart-type du filtre gaussien appliqué.

Cette méthode est bien adaptée aux scénarios où une estimation rapide du flot optique est nécessaire. Cependant, elle peut présenter des limites dans des scènes complexes ou en présence de grands déplacements, où des approches modernes, comme celles basées sur le deep learning, peuvent offrir de meilleures performances.

3.2 Méthode Récente

Une méthode basée sur le deep learning, comme RAFT (Recurrent All-Pairs Field Transforms), sera utilisée. Ces approches exploitent des réseaux de neurones convolutifs pour apprendre à prédire le flot optique à partir de données annotées. RAFT, par exemple, repose sur une architecture récurrente qui raffine les prédictions au fil des itérations, atteignant une précision élevée sur les jeux de données standards.

4 Métriques d'Évaluation

4.1 Sur MPI Sintel

— **End Point Error (EPE)** : Mesure la distance moyenne entre les vecteurs de flot optique prédit et la vérité terrain.
— **Angular Error (AE)** : Calcule l'écart angulaire moyen entre les vecteurs prédit et réel.

4.2 Sur GITW

— **Mean Square Error (MSE)** : Évalue l'erreur quadratique moyenne entre l'image originale et l'image compensée à l'aide du flot optique estimé.

5 Résultats et Analyse

Résultats pour la Méthode Classique sur MPI Sintel

Les résultats pour la méthode de Farneback ont été calculés sur trois séquences du jeu de données MPI Sintel :

— **Alley_2** : Avec 49 paires d'images, la moyenne de l'End Point Error (EPE) est de 2.1458, et celle de l'Angular Error (AAE) est de 0.1750 radians. Ces valeurs indiquent une bonne précision pour cette séquence.
— **Market_2** : Avec 49 paires d'images également, l'EPE moyen est de 2.5431, tandis que l'AAE moyen est de 0.4919 radians. Les erreurs augmentent légèrement, ce qui pourrait être dû à des déplacements plus complexes dans cette séquence.
— **Temple_3** : Cette séquence présente des erreurs beaucoup plus importantes, avec un EPE moyen de 35.7003 et un AAE moyen de 0.8898 radians. Ces résultats suggèrent que la méthode de Farneback a des difficultés avec cette séquence, probablement en raison de grands déplacements ou de déformations importantes.

Analyse des Résultats

Les résultats obtenus montrent que la méthode classique de Farneback donne de bonnes performances pour des séquences présentant des déplacements modérés, comme *Alley_2* et *Market_2*. Cependant, dans des scénarios où les mouvements sont plus complexes, comme *Temple_3*, les erreurs augmentent considérablement. Cela met en évidence les limitations des méthodes classiques lorsqu'elles sont confrontées à des déformations importantes ou à des scènes avec des déplacements non linéaires.

Les graphiques suivants illustrent les variations de l'EPE et de l'AAE pour chaque paire d'images dans les séquences *Alley_2*, *Market_2*, et *Temple_3* :

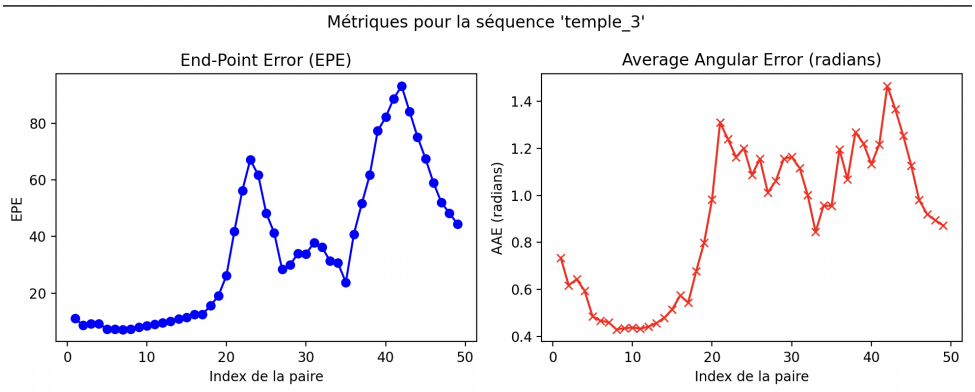

FIGURE 1 – Métriques pour la séquence *Temple_3*.

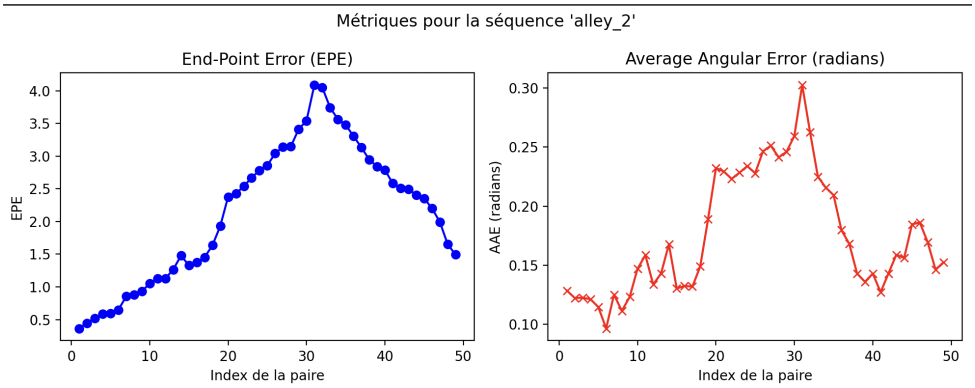

FIGURE 2 – Métriques pour la séquence *Alley_2*.

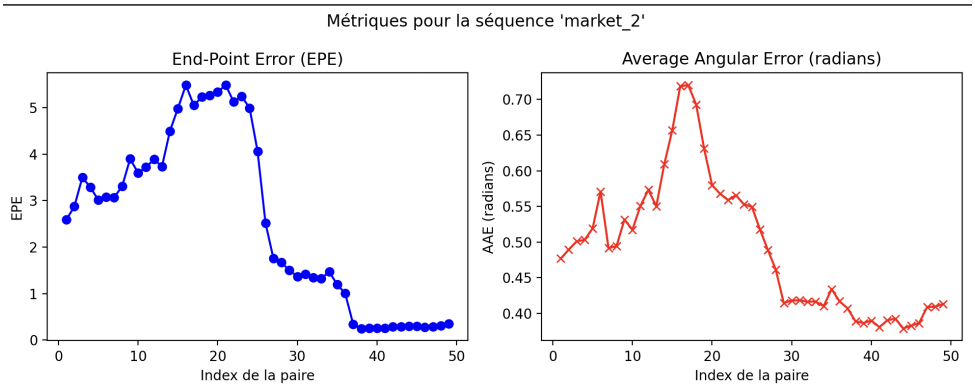

FIGURE 3 – Métriques pour la séquence *Market_2*.

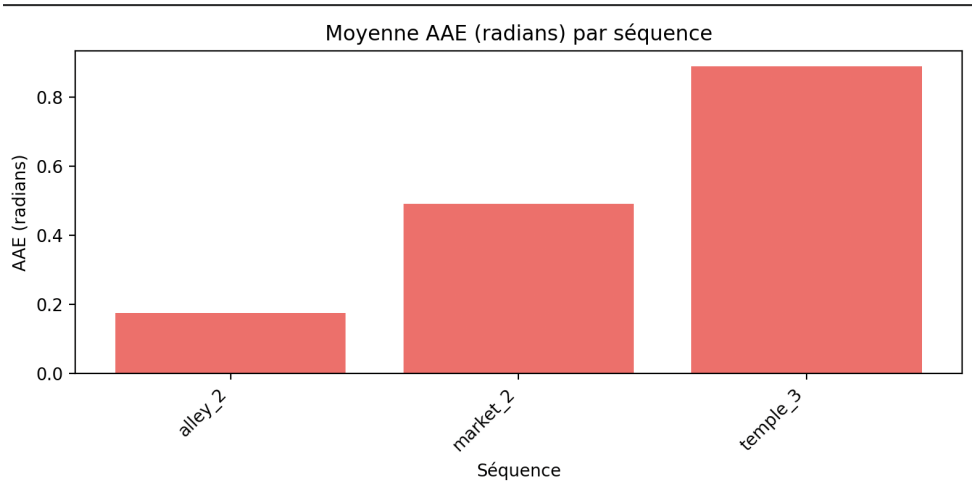

FIGURE 4 – EPE moyen par séquence.

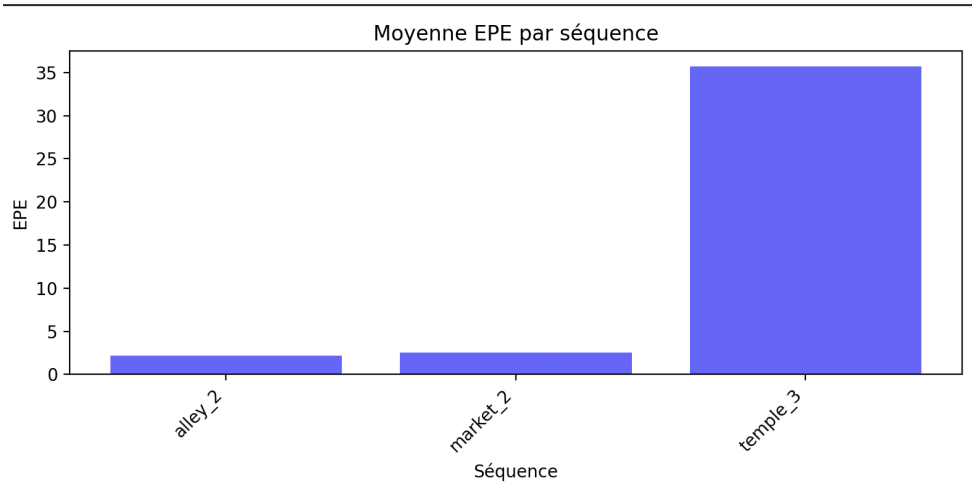

FIGURE 5 – AAE moyen par séquence.

Résultats pour la Méthode Classique sur GITW

Les résultats pour la méthode de Farneback ont été calculés sur trois séquences du jeu de données Grasping in the Wild (GITW) :

- **Bowl** : Avec 178 paires d'images, la Mean Square Error (MSE) moyenne est de 132.8229. Ce résultat indique une performance relativement stable sur cette séquence.
- **CanOfCocaCola** : Avec 234 paires d'images, la MSE moyenne atteint 307.6155. Cette valeur élevée suggère une plus grande complexité des déplacements ou des déformations dans cette séquence.
- **Rice** : Avec 307 paires d'images, la MSE moyenne est de 134.8171. Ces résultats montrent que la méthode gère bien cette séquence malgré quelques variations.

Analyse des Résultats

Les résultats obtenus montrent que la méthode de Farneback offre des performances variées selon les séquences du jeu de données GITW. Les séquences *Bowl* et *Rice* présentent

des MSE relativement faibles, ce qui indique une bonne capacité de la méthode à compenser les déplacements entre les images. En revanche, la séquence *CanOfCocaCola* affiche une MSE nettement plus élevée, ce qui met en évidence les limites de la méthode face à des scénarios complexes ou des variations importantes.

Les graphiques suivants illustrent les variations de la MSE pour chaque paire d'images dans les séquences *Bowl*, *CanOfCocaCola*, et *Rice*, ainsi que les MSE moyens pour chaque séquence :

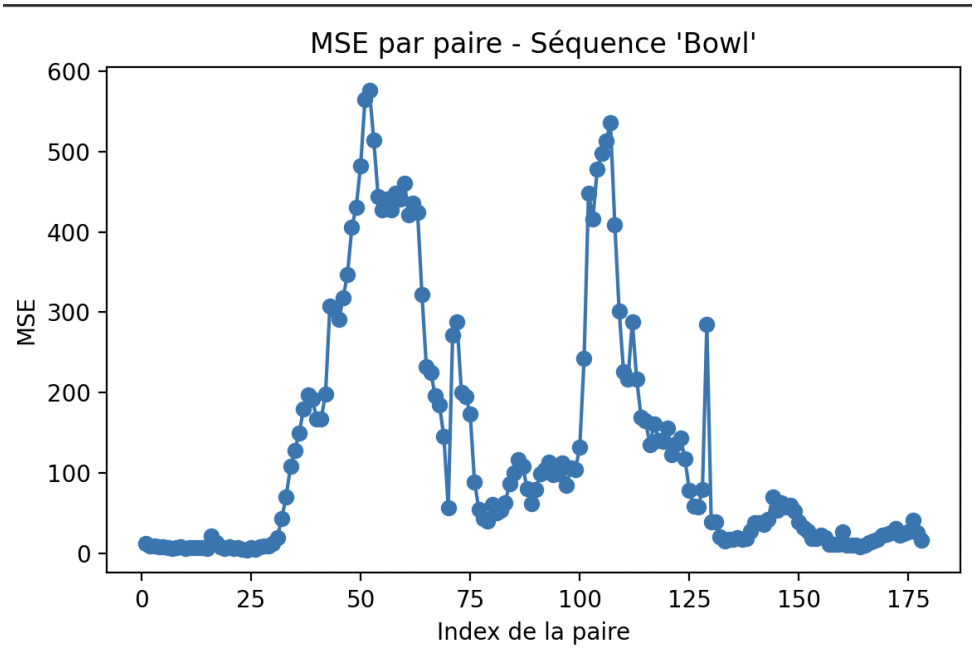

FIGURE 6 – MSE par paire - Séquence *Bowl*.

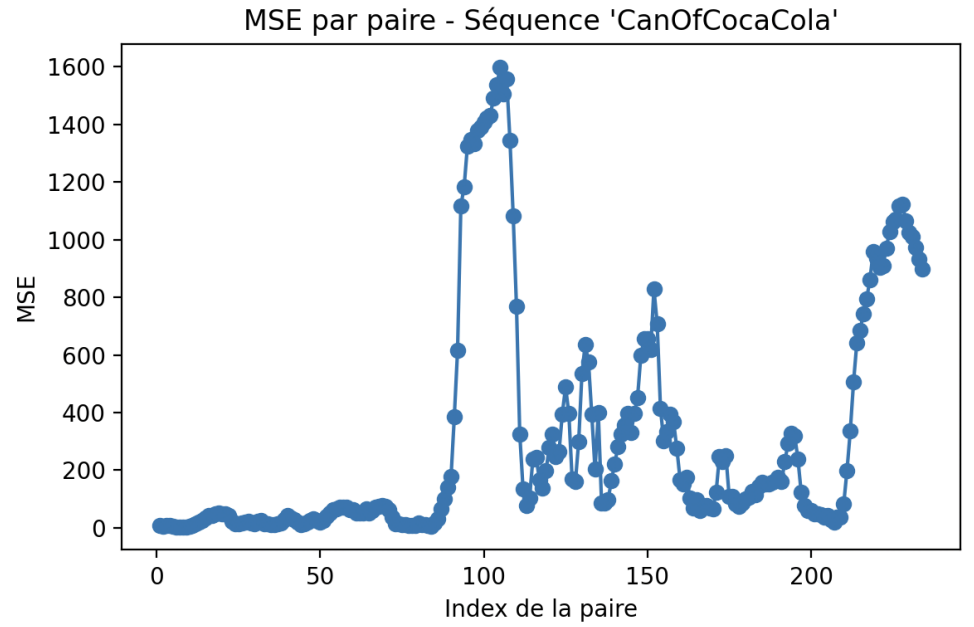

FIGURE 7 – MSE par paire - Séquence *CanOfCocaCola*.

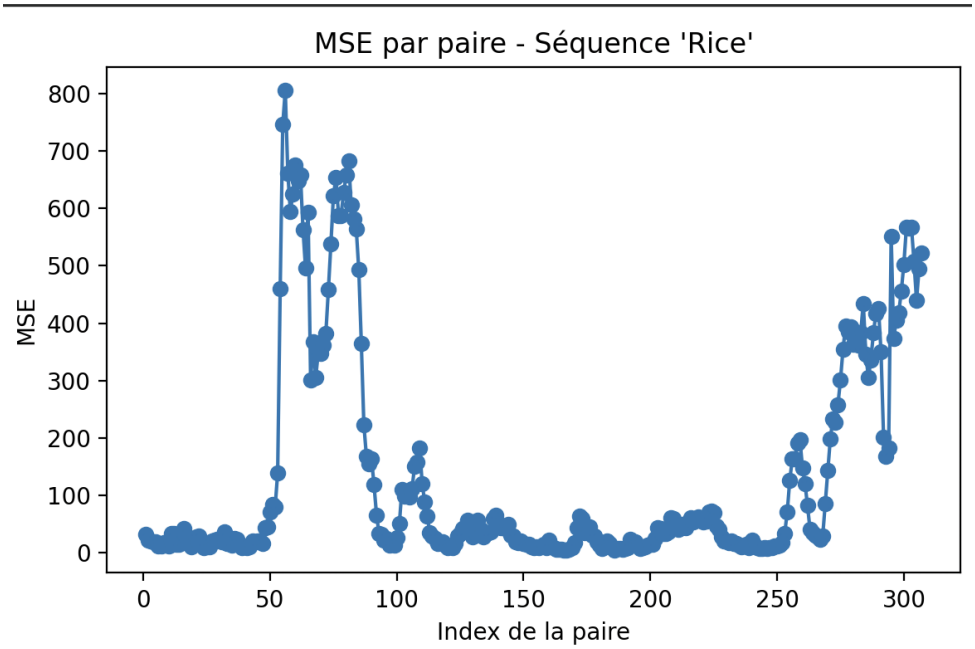

FIGURE 8 – MSE par paire - Séquence *Rice*.

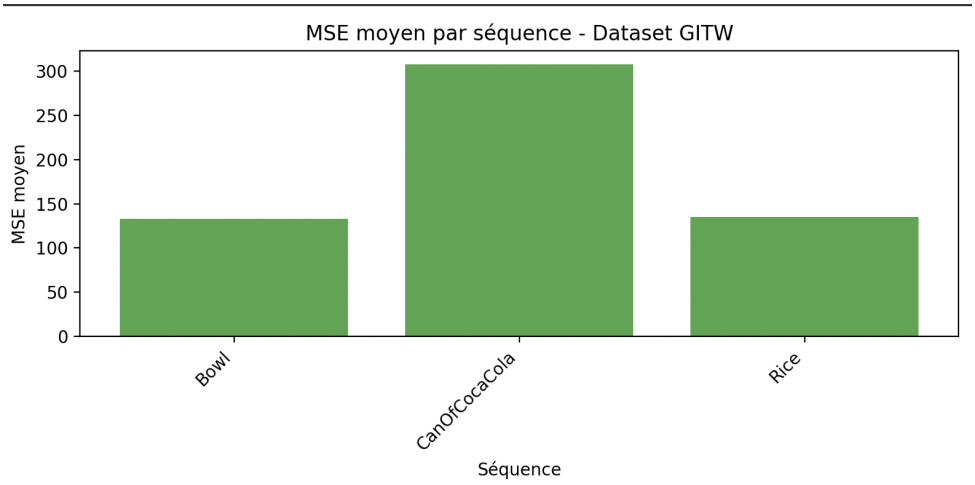

FIGURE 9 – MSE moyen par séquence - Dataset GITW.

Résultats pour la Méthode Deep Learning sur MPI Sintel

Dans le cadre de ce projet, nous avons également testé une approche de deep learning basée sur **RAFT** (Recurrent All-Pairs Field Transforms). L'objectif était de vérifier si un modèle pré-entraîné, appliqué directement sur les séquences de MPI Sintel, pouvait surpasser la méthode de Farneback.

Mise en œuvre. Nous avons utilisé un code en Python avec PyTorch pour charger le modèle RAFT pré-entraîné, puis calculer le flot optique sur les mêmes paires d'images que pour la méthode classique. Les fichiers `.flo` prédits ont ensuite été comparés à la vérité terrain, comme dans le cas de Farneback.

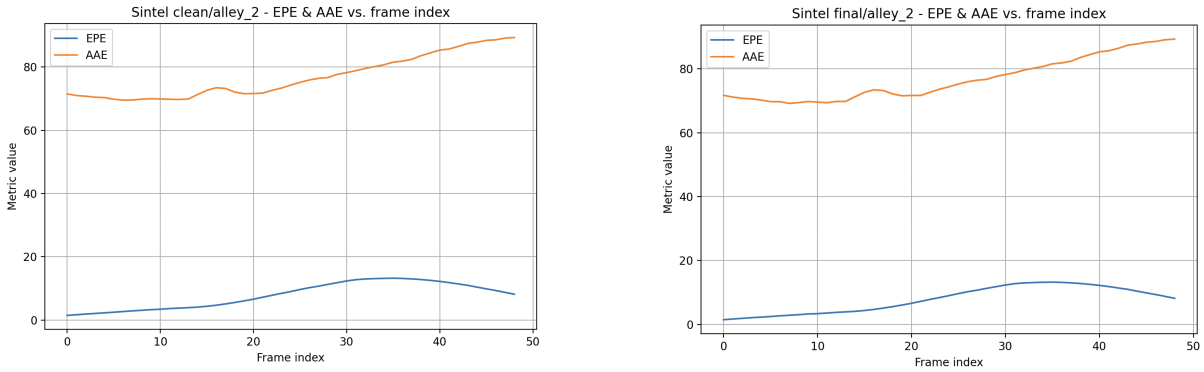

FIGURE 10 – Métriques pour la séquence *Alley_2* (Clean et Final).

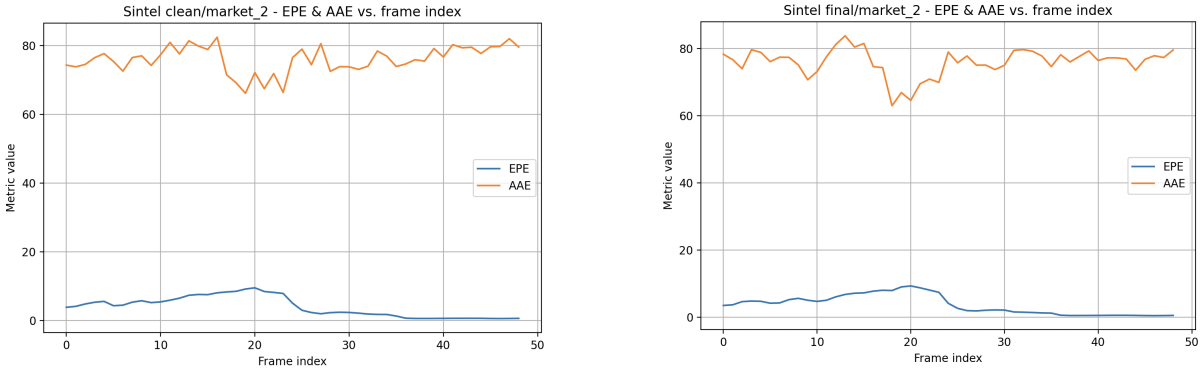

FIGURE 11 – Métriques pour la séquence *Market_2* (Clean et Final).

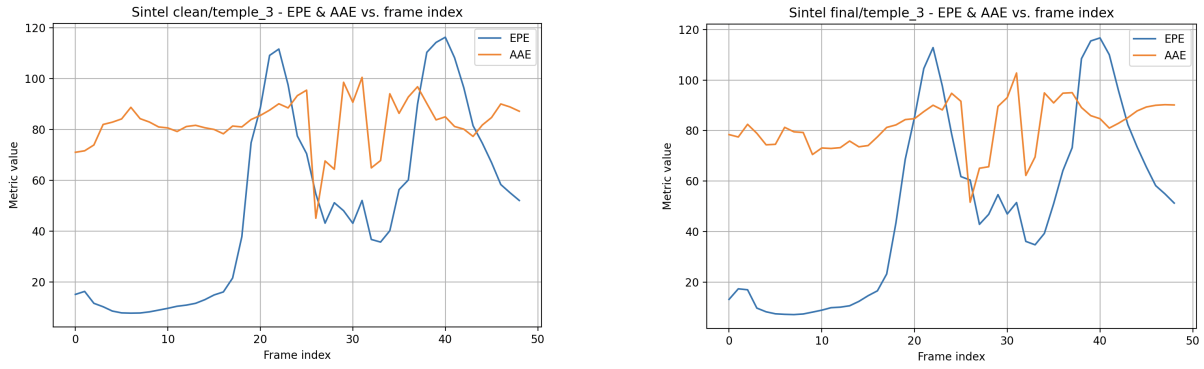

FIGURE 12 – Métriques pour la séquence *Temple_3* (Clean et Final).

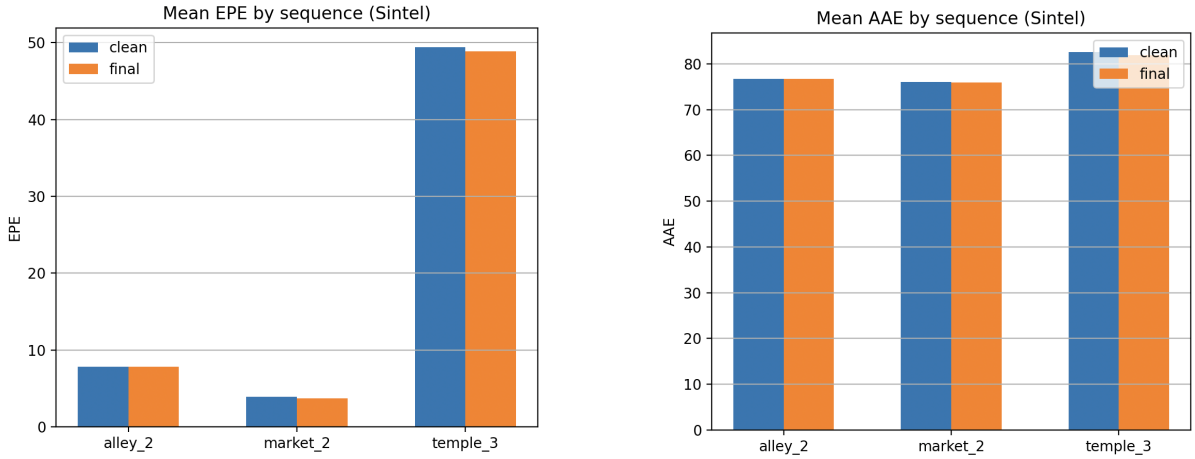

FIGURE 13 – Métriques par séquence (Clean et Final).

Observations. Malheureusement, **les résultats ne se sont pas avérés concluants**. Nous avons constaté des valeurs d'EPE et d'AAE souvent *plus élevées* que celles obtenues avec la méthode classique. Plusieurs facteurs peuvent l'expliquer :

- **Différences de domaine** : Le modèle RAFT utilisé n'était pas forcément ajusté aux particularités visuelles des séquences Sintel (entraînement sur d'autres datasets, par exemple).
- **Paramètres ou version du modèle** : Une configuration RAFT inappropriée ou un checkpoint incompatible peut altérer les performances.
- **Grandes déformations** : Dans certains cas, l'architecture RAFT peut nécessiter un plus gros fine-tuning pour traiter des mouvements importants.

En conséquence, pour la séquence *Temple_3* par exemple, l'End Point Error mesurée dépassait celle de Farneback, suggérant que le réseau pré-entraîné n'a pas réussi à bien généraliser dans ce contexte.

Résultats pour la Méthode Récente (Deep Learning) sur GITW

Nous avons également évalué RAFT sur GITW, via le même principe : compensation de l'image 1 pour la comparer à l'image 2 (MSE). Là encore, la performance s'est révélée **moins satisfaisante** que celle de la méthode de Farneback, avec des MSE bien supérieurs sur les trois séquences.

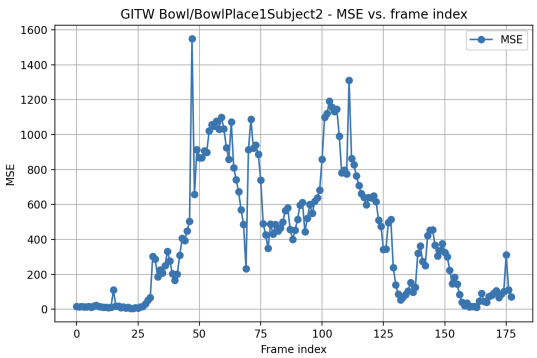

FIGURE 14 – MSE par paire - Séquence *Bowl*.

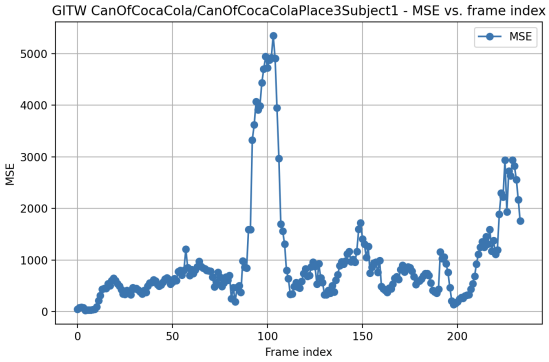

FIGURE 15 – MSE par paire - Séquence *CanOfCocaCola*.

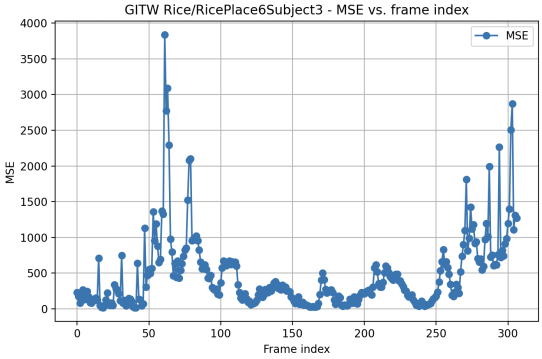

FIGURE 16 – MSE par paire - Séquence *Rice*.

FIGURE 17 – MSE moyen par séquence - Dataset GITW.

Analyse. L'absence de fine-tuning, l'éventuelle inadéquation du domaine (objets, éclairages, etc.) et les limites de la version du modèle RAFT employée ont pu contribuer à ces résultats moins bons. Il serait possible d'améliorer la précision en procédant à un apprentissage ou un ajustement spécifique pour GITW.

6 Conclusion

Ce projet a permis de comparer deux approches pour le calcul du flot optique dense : une méthode classique, représentée par l'algorithme de Farneback, et une méthode récente basée sur le deep learning, illustrée par RAFT. Les résultats montrent que la méthode classique offre des performances satisfaisantes pour des scénarios présentant des déplacements modérés, tandis qu'elle atteint ses limites face à des mouvements complexes ou de grandes déformations. En revanche, la méthode basée sur le deep learning n'a pas surpassé la méthode classique dans ce contexte, en grande partie en raison de l'absence de fine-tuning et des différences de domaine entre les jeux de données et l'entraînement initial du modèle.

Ces observations soulignent l'importance d'adapter les modèles récents à des cas d'utilisation spécifiques pour en tirer pleinement parti. Une prochaine étape pourrait consister à effectuer un fine-tuning du modèle RAFT sur les jeux de données utilisés, afin d'explorer

son réel potentiel face à des scénarios variés. Ce travail met en lumière les avantages et les limites des deux approches, tout en ouvrant des perspectives pour des applications futures dans le domaine de l'analyse vidéo et de la vision robotique.