



# Thèse Projet ANR "Numalyse"

## Compte Rendu 4

Benjamin Serva  
GitHub : [https://github.com/bservaa34/Numalyse\\_thesis](https://github.com/bservaa34/Numalyse_thesis)

24 novembre 2025 - 28 novembre 2025

Encadrants :

Olivier Strauss & William Puech & Frédéric Comby

# 1 Tâches effectuées cette semaine

## 1.1 Récolte Dataset

Le Tab. 1 présente un état des différents datasets existants pour l'évaluation.

	Nb vidéos	Nb plans	Moy durée vidéo (s)	Moy durée plan (s)	Type de vidéos	Dispo
BBC	11	4900 (est)	2945	6.57	docs	✓✓
RAI	10	987	591	5.65	docs/talk shows	✗
AutoShot	853	11606	39.5	2.59	tiktok	✓✓
ClipShot	5917	91 200 (est)	237	15.34	tiktok	✓✓
TRECVID-IACC.3	4600	-	468	-	40% movie / 21% web / 39% divers	✓
TRECVID-V3C1	7475	1 082 659	482	-	aucune informations	▲

TABLE 1 – Comparatif Dataset (✓ : disponible, ✗ : non trouvée, ✓✓ : que j'ai sur ma machine, ▲ : potentiellement disponible

## 1.2 Présentation Loig LE BIHAN dans le cadre de l'UE conférence du master 2 Imagine

Voici quelques notes que j'ai prises lors de cette présentation. **Analyse de film au fil du temps :**

À l'origine, l'analyse se faisait directement à partir de la pellicule. L'arrêt sur image était difficile et risqué (risque de brûlure de la pellicule), ce qui faisait que l'analyse portait souvent sur le début du film, plus facilement accessible.

L'apparition de la table de montage a permis un arrêt sur image sécurisé : un geste fondateur dans l'histoire de l'analyse filmique.

Années 1980–2000 :

L'arrivée des cassettes VHS bouleverse l'analyse : accès facilité, possibilité de revoir et manipuler les images.

Dans les concours des écoles de cinéma, l'analyse de séquence devient une épreuve importante. L'utilisation du ralenti se généralise : il altère légèrement la perception mais permet une meilleure observation.

La lecture à vitesse variable devient un nouveau geste analytique.

Années 1990–2020 :

Le DVD introduit un ralenti saccadé et peu agréable, ce qui entraîne l'abandon progressif de cette technique.

L'accès par chapitres ou par timecode facilite une analyse globale du film, et non plus seulement d'extraits isolés.

L'apparition des bonus (making-of, commentaires) encourage une analyse intégrant le contexte de production.

Années 2010–2020 :

L'analyse se fait désormais à partir de fichiers numériques et de logiciels de lecture multimédia.

### 1.3 Réunion Encadrant 27/11/2025 9h30-10h30

Lors de cette réunion, j'ai présenté le postulat de départ de mon travail : la méthode Adaptive Detector utilisée dans le SLV. Celle-ci présente plusieurs limites, notamment :

- la génération de faux positifs lors de mouvements rapides de caméra,
- l'absence de détection des transitions progressives (exemple dans la Fig. 1).



FIGURE 1 – Exemple de transition progressive

Pour rappel l'évaluation de cette méthode sur le dataset BBC et sur un sous-ensemble de AutoShot avait donné : Précision : 71.42%, Rappel : 64.60% et F1-score : 67.84%.

J'ai lu de nombreux articles et en ai retenu huit, représentant les meilleures contributions pour chaque type d'approche.

#### 1. Deep Learning

- *Large-scale, Fast and Accurate Shot Boundary Detection through Spatio-temporal Convolutional Neural Networks* (Hassanien et al. [1], 2017)
- *AutoShot : A Short Video Dataset and State-of-the-Art Shot Boundary Detection* (Zhu et al. [2], 2023)
- *Attention-Driven LSTM with Multi-Attribute Feature Learning for Shot Boundary Detection* (Hadke et al. [3], 2025)

#### 2. Multi-Features

##### (a) Avec Seuil

- Adaptive Detector
- *The Detection of Video Shot Transitions Based on Primary Segments Using the Adaptive Threshold of Colour-Based Histogram Differences and Candidate Segments Using the SURF Feature Descriptor* (Raja Suguna et al. [4], 2022)

##### (b) Avec classifieur

- *Shot Boundary Detection Based on Global Features and the Target Features* (Li et al. [5], 2023)
- *Efficient Shot Boundary Detection with Multiple Visual Representations* (Jose et al. [6], 2022)

#### 3. Hybride (DL + Features)

- *Hybridized Neural Network Based Approaches Used for Video Shot Boundary Detection* (Chavate et al. [7], 2023)

#### 4. Méthodes géométriques / topologiques

- *Dynamic Mode Decomposition Based Video Shot Detection* (Bi et al. [8], 2018)

J'ai aussi fais un tableau récapitulatif des performances de ces méthodes (Tab. 2).

Référence	Approche	Caractéristiques utilisées	Architecture / Modèle	P	R	F1	T	Dataset
AdaptiveDetector	Seuil adaptatif	HSV, histogrammes différenciés	Aucun (non supervisé)	71.42%	64.60%	67.84%	C	BBC / AutoShot
1) Li et al. [5], 2023	Fusion de caractéristiques	RGB global + SIFT local	Fusion pondérée	91.75%	90.82%	91.28%	C+G	RAI
2) Bi et al. [8], 2018	Méthode spectrale	DMD (décomposition modale dynamique), intensité RGB	Analyse spectrale non supervisée	91.6%	85.6%	88.6%	C+G	TRECVID (?)
3.a) Chavate et al.-1 [7], 2023	Fusion multi-domaines via réseaux hybrides	DTCWT + WHT (ondelettes)	DBN optimisé (SSDOA)	93.92%	93.08%	93.08%	C+G	TRECVID 2007, 2016–2019
3.b) Chavate et al.-2 [7], 2023	Deep learning supervisé	Images RGB, features convolutionnels	DCNN (CNN profond classique)	92.62%	92.32%	92.46%	C+G	TRECVID 2007, 2016–2019
4) Hassanien et al. [1], 2017	Deep learning spatio-temporel	Séquences vidéo 16–32 frames, RGB	3D CNN (spatio-temporelle)	/	/	94.00%	C+G	RAI
5) Zhu et al. [2], 2023	Deep learning avancé moderne	RGB + motion + contexte temporel	3D-CNN + Transformer	/	/	95.50%	C	BBC / RAI / ClipShots
6) Suguna et al. [4], 2022	Méthode hybride non supervisée + features locales	HSV histogrammes + seuils adaptatifs + SURF	Pipeline heuristique + SURF matcher	94.60%	95.10%	94.80%	C+G	TRECVID 2001
7) Jose et al. [6], 2022	Fusion multi-features + SVM supervisé	CLD (Color Layout Descriptor) + ECR + SIFT	SVM avec multi-représentations	97.6%	96.9%	97.2%	C	TRECVID 2007
8) Hadke et al. [3], 2025	Deep learning avancé hybride	apparence + mouvement + temporel	Réseaux hybrides : SSN-CAP + MSN + AtDLSTM	97.27%	97.81%	98.11%	C+G	TRECVID (2018-2021) / Autoshot++

TABLE 2 – Tableau récapitulatif des performances et types de méthodes pour les articles étudiés et la méthode expérimentée (AdaptiveDetector).

Cependant, on constate que chaque méthode est évaluée sur des datasets différents, parfois peu cohérents avec les besoins du projet Numalyse.

La totalité du travail présenté lors de la réunion a permis de définir plusieurs objectifs (présentés dans la Section 2). Elle a aussi soulevé une question importante : où placer la frontière de plan dans le cas des transitions progressives ? J'ai contacté M. Le Bihan par mail pour obtenir son avis sur ce point.

## 2 Objectifs pour la semaine prochaine

- Rédiger une description détaillée des méthodes de segmentation en plans, notamment pour les approches deep learning, en précisant clairement les modèles utilisés.
- Implémenter et comparer, sur un même dataset (complet et pertinent pour le projet), au moins une méthode représentative de chaque approche.
- Démarrer une recherche d'articles sur l'analyse sémantique du contenu vidéo
- Réunion Encadrants le 04/12/2025 à 9h30 en salle B5.02.156.

## 3 Tâches à effectuer durant la thèse

- La rentrée du Collège Doctoral de l'Université se tiendra le vendredi 23 janvier de 09h00 à 12h30.
- Installer un double boot Linux.
- Réaliser les 100 heures de formations complémentaires obligatoires.

## Références

- [1] A. Hassanien, M. Elgharib, A. Selim, S.-H. Bae, M. Hefeeda, and W. Matusik, “Large-scale, fast and accurate shot boundary detection through spatio-temporal convolutional neural networks,” 2017. [Online]. Available : <https://arxiv.org/abs/1705.03281>
- [2] W. Zhu, Y. Huang, X. Xie, W. Liu, J. Deng, D. Zhang, Z. Wang, and J. Liu, “Auto-shot : A short video dataset and state-of-the-art shot boundary detection,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023.
- [3] S. ChaitandasHadke, R. Mishra, R. T. Bankar, S. A. Chhabria, S. P. Chavate, and L. S. Pinjarkar, “An attention driven long short term memory based multi-attribute feature learning for shot boundary detection,” *Knowledge-Based Systems*, vol. 317, p. 113379, May 2025. [Online]. Available : <https://linkinghub.elsevier.com/retrieve/pii/S0950705125004265>
- [4] R. S. M., K. A., and A. S., “The detection of video shot transitions based on primary segments using the adaptive threshold of colour-based histogram differences and candidate segments using the surf feature descriptor,” *Symmetry*, vol. 14, no. 10, 2022. [Online]. Available : <https://www.mdpi.com/2073-8994/14/10/2041>
- [5] Q. Li, X. Chen, B. Wang, J. Liu, G. Zhang, and B. Feng, “Shot boundary detection based on global features and the target features,” *Symmetry*, vol. 15, no. 3, 2023. [Online]. Available : <https://www.mdpi.com/2073-8994/15/3/565>
- [6] J. Jose, R. Soundrapandian, D. M. Ghalib, A. Shankar, P. Sharma, and M. Khosravi, “Efficient shot boundary detection with multiple visual representations,” *Mobile Information Systems*, vol. 2022, 10 2022.
- [7] S. Chavate, R. Mishra, S. K. Singh, and D. Sharma, “Hybridized neural network based approaches used for video shot boundary detection,” *Indian Journal of Science and Technology*, vol. 16, no. 33, pp. 2670–2680, 2023.

- [8] C. Bi, Y. Yuan, J. Zhang, Y. Shi, Y. Xiang, Y. Wang, and R. Zhang, “Dynamic mode decomposition based video shot detection,” *IEEE Access*, vol. 6, pp. 21 397–21 407, 2018.