

# Une approche hybride pour la détection des levées du stylo : preuve de concept pour l'analyse de l'écriture

Lauren Sismeiro<sup>1</sup>, Rémy Plastre<sup>2</sup>, Romain Sebire<sup>2</sup>, Binbin Xu<sup>1</sup>, Frédéric Puyjarinet<sup>1</sup>, Gérard Dray<sup>1</sup>

<sup>1</sup> EuroMov Digital Health in Motion, Univ Montpellier, IMT Mines Alès, Alès, France

<sup>2</sup> IMT Mines Alès, Alès, France

lauren.sismeiro@gmail.com

## Résumé

*L'évaluation de la dysgraphie repose souvent sur le test statique BHK, qui omet la dynamique motrice. Si les tablettes numériques pallient ce manque, leur signal s'interrompt dès que le stylo dépasse 10 mm de hauteur. Notre preuve de concept utilise la vision par ordinateur pour combler cette lacune. En combinant YOLOv11 et classification cinématique, nous détectons les états de contact et de levée avec un  $F_2$ -score de 0,81 sur des vidéos vues de dessus. Cette approche non invasive offre une analyse écologique et complète de l'écriture clinique. Une version étendue et plus détaillée de ce travail fait actuellement l'objet d'une soumission à une conférence internationale.*

## Mots-clés

*Analyse de l'écriture, Vision par ordinateur, Cinématique, Dysgraphie, Santé numérique.*

## 1 Introduction

En France, le diagnostic de la dysgraphie, trouble de l'apprentissage affectant la qualité et la vitesse de l'écriture, repose sur l'échelle BHK [1]. Il s'agit d'une évaluation statique effectuée a posteriori sur une tâche de copie de cinq minutes, ignorant la dynamique scripturale. Les tablettes numériques permettent d'extraire ces variables temporelles, mais leur technologie limite la détection aux levées inférieures à 10 mm [3]. Cet angle mort empêche l'analyse des levées de haute altitude.

Nous proposons d'utiliser la vision par ordinateur comme modalité complémentaire. Une acquisition vidéo vue de dessus permet de suivre la pointe du stylo sans contrainte de distance verticale. Ce pipeline multi-étapes assure une transition fluide entre contact (*Pen-Down*) et vol (*Pen-Up*) pour une évaluation clinique objective.

## 2 Matériels et Méthodes

Notre approche repose sur un pipeline hybride qui sépare la localisation de l'objet de la classification de son état (Fig. 1).

### 2.1 Collecte de données et Annotation

Le dataset comprend 13 507 images (1080p, 30 fps) issues de vidéos d'écriture variées (cursive, script, divers stylos) [2]. Trois annotateurs ont étiqueté les états *Pen-Down*, *Pen-Up* et *Transition*. L'accord inter-juges est substantiel avec un coefficient Kappa de Fleiss de 0,78. Les désaccords ont été intégrés via un étiquetage souple (*soft labeling*) pour modéliser l'incertitude des changements rapides. Après retrait des images non informatives, 28,7% des frames correspondent à des levées.

### 2.2 Pipeline en Quatre Étapes

Le système transforme le flux vidéo en événements discrets via quatre modules successifs :

**1. Tracking de la pointe.** Nous utilisons YOLOv11m pour localiser la pointe du stylo  $(x, y)$ . Le modèle a été entraîné pour être robuste aux occultations par la main du scripteur. Il atteint une erreur de localisation médiane de 3,28 px, permettant de générer une trajectoire continue même lors de mouvements aériens rapides (Fig. 2).

**2. Caractéristiques.** 147 descripteurs capturent la dynamique locale (linéarité, accélération sur fenêtres de 3 à 16 images) et globale (inclinaison, vitesse) pour normaliser les styles individuels.

**3. Classification.** Des modèles basés sur des arbres de décision sont optimisés via Optuna pour maximiser la détection et produire une probabilité  $P_{\text{Pen-Up}}$  par image.

**4. Post-traitement.** Un lissage par hystérésis, un filtrage morphologique et un alignement cinématique stabilisent les prédictions en calant les segments sur les minima de vitesse, partant du principe physique qu'une levée ou un posé de stylo s'effectue à vitesse quasi nulle.

## 3 Résultats

Les performances sont évaluées au niveau événementiel avec deux tolérances temporelles : 5 et 10 images ( $\approx 167$  et 333 ms). Une prédiction est considérée correcte si elle se situe dans cette fenêtre autour de l'annotation de référence. Ces seuils, cohérents avec la durée moyenne des pauses en écriture manuscrite ( $\approx 400$  ms) [4], permettent d'absorber les micro-décalages d'annotation humaine sans compromettre la pertinence clinique.

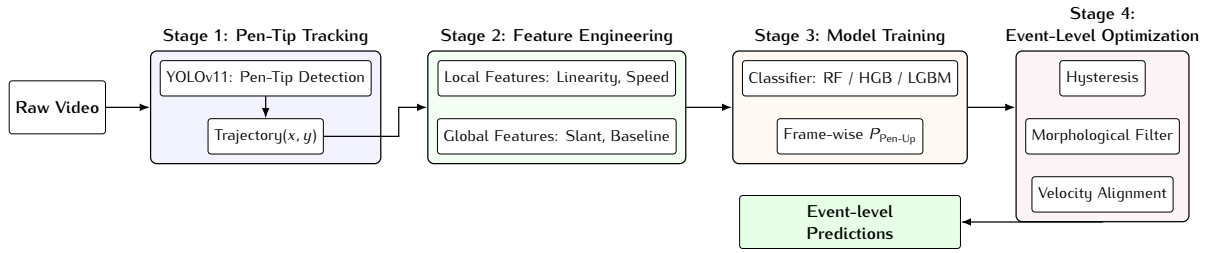


FIGURE 1 – Vue d’ensemble du pipeline multi-étapes proposé pour la détection de l’état de contact du stylo.

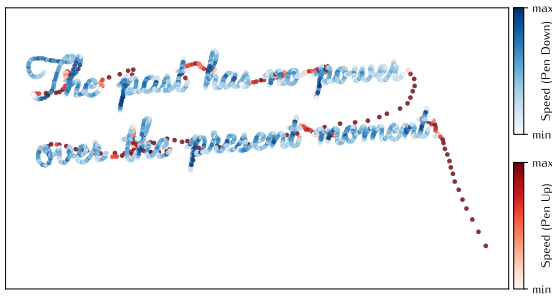


FIGURE 2 – Visualisation de la trajectoire colorée par vitesse (Bleu : contact, Rouge : levé).

**Analyse de performance.** Le post-traitement par seuils d’hystérésis s’est avéré crucial pour stabiliser le signal et supprimer les micro-oscillations (*flickering*) aux transitions d’état. À une tolérance de 10 images, le modèle Random Forest atteint un  $F_2$ -score de **0,81**. Cette métrique privilégie le rappel, choix cohérent en contexte clinique où il est essentiel de ne manquer aucun événement de levée, quitte à accepter quelques faux positifs. Ainsi, l’ensemble des pauses aériennes, potentiels marqueurs de difficultés motrices, est capturé pour l’évaluation [5]. Ces résultats confirment la capacité du système à saisir la dynamique du mouvement malgré l’hétérogénéité des supports et des éclairages.

**Interprétabilité des caractéristiques.** L’analyse de l’importance des variables indique que le modèle repose sur des critères physiques cohérents, tels que la linéarité du tracé et la régularité de la vitesse instantanée. L’usage de descripteurs cinématiques explicites constitue un avantage majeur face aux approches *boîte noire*, en permettant au clinicien de comprendre et d’interpréter la classification automatique.

## 4 Discussion et Conclusion

Cette preuve de concept démontre qu’une acquisition vidéo en vue de dessus peut compenser les limites physiques des tablettes numériques. En levant la contrainte des 10 mm, elle permet une analyse continue du mouvement, incluant les phases aériennes auparavant inaccessibles.

**Avantages opérationnels.** Le pipeline proposé facilite l’intégration clinique : peu énergivore, il fonctionne sur des processeurs standards (CPU), sans matériel coûteux. Il ga-

rantit également un haut niveau de confidentialité, les vidéos pouvant être supprimées après extraction des coordonnées  $(x, y)$  de la pointe du stylo, seules des séries temporelles anonymisées étant conservées.

**Perspectives et conclusion.** Cette approche ouvre la voie à une version *augmentée* du BHK, offrant au praticien des mesures automatiques telles que le ratio temps de vol / temps de contact, indicateur de fluidité scripturale. La vue monoculaire limite toutefois l’estimation directe de la profondeur; les travaux futurs viseront l’intégration de vues multi-angles pour quantifier la hauteur de levée (axe Z) et un déploiement en *edge computing* garantissant un traitement local conforme au RGPD. Ce travail confirme la viabilité d’un système de vision bas coût pour enrichir l’évaluation écologique de la dysgraphie.

## Usage de l’IA générative

L’IA générative a été utilisée pour reformuler certains passages et corriger l’orthographe et la grammaire.

## Remerciements

Les auteurs remercient IMT Mines Ales et la Région Occitanie pour leur soutien financier pour le projet doctoral AVIAREPTE. Ils remercient également @DorufaVSArt pour l’autorisation d’utiliser ses vidéos YouTube dans le cadre de leurs recherches sur l’écriture manuscrite.

## Références

- [1] M. Charles, R. Soppelsa, and J.-M. Albaret. *BHK - Échelle d’évaluation rapide de l’écriture chez l’enfant*. Éditions Centre de Psychologie Appliquée, 2004.
- [2] DorufaVSArt. Dorufavsart. <https://www.youtube.com/@DorufaVSArt>, 2025.
- [3] Jean-Claude Gilhodes, Elie Fabiani, Marieke Longcamp, Jean-Luc Velay, and Jérémy Danna. Chapitre 4. Traiter des données de langage écrit recueillies avec tablette graphique. In *Introduction aux statistiques en sciences du langage*, pages 117–136. Dunod, 2023.
- [4] Mariona Pascual, Olga Soler, and Naymé Salas. In a split second : Handwriting pauses in typical and struggling writers. *Frontiers in Psychology*, Volume 13 - 2022, 2023.
- [5] Sara Rosenblum, Shula Parush, and Patrice L. Weiss. The in Air Phenomenon : Temporal and Spatial Correlates of the Handwriting Process. *Perceptual and Motor Skills*, 96(3) :933–954, June 2003.