Lecture assistée de texte dans les bandes dessinées

Frédéric Rayar*, Clément Charrier**,
Rémy Leconge***, Sylvie Treuillet***, Frédéric Daubignard****

* LIFAT, Université de Tours, frederic.rayar@univ-tours.fr,
** PRISME, Université d'Orléans, clement.charrier@etu.univ-orleans.fr,
*** PRISME, Université d'Orléans, {remy.leconge, sylvie.treuillet}@univ-orleans.fr,
**** ALGONA, frederic.daubignard@algona.fr

Résumé. La reconnaissance de texte dans les documents et les images naturelles est un domaine de la recherche qui a connu des avancées spectaculaires ces dernières années. Cependant l'application de ces technologies à des domaines spécifiques, tel que l'accessibilité des livres à des publics dits "empêchés de lire", reste un défi. Dans cet article, nous proposons une approche hybride IA/Humain pour créer une application robuste d'aide à la lecture de bande dessinée via un terminal mobile, tout en maintenant le contact avec le support physique de ladite bande dessinée.

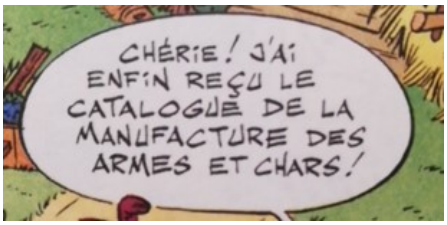
1 Introduction

L'accessibilité des livres pour les personnes dites "empêchées de lire" (dyslexiques, déficients visuels, autistes, ...) est un processus nécessaire pour donner à tous l'opportunité de bénéficier de cette activité ludique et enrichissante. Pour tirer parti de l'essor des livres numériques, des initiatives sont en cours de déploiement afin que ces derniers soient nativement accessibles : on peut citer la directive européenne relative à l'accessibilité des biens et des services [1] adoptée en 2019, mais dont la mise en application ne sera effective qu'en 2025. Mais, au-delà du format numérique, comment rendre accessible des livres imprimés à ces publics afin de maintenir le plaisir de manipuler le format papier tout en assurant une cohabitation physique/numérique pertinente et non-intrusive ?

Notre projet a pour objectif de fournir une application sur terminal mobile (smartphone ou tablette) offrant une aide à la lecture sous forme de lecture vocale de textes d'intérêt, sélectionnés par le lecteur et ce à l'aide d'algorithmes dits d'Intelligence Artificielle (analyse d'image, apprentissage machine, reconnaissance optique de caractères, ...). En particulier, nous avons opté pour un cas d'utilisation particulier et complexe : les bandes dessinées (BD). Mode d'expression culturelle transgénérationnel et répandu dans le monde entier, ce support permet de raconter des histoires en combinant des informations visuelles et textuelles. Ayant eu une reconnaissance contemporaine en recevant le qualificatif de 9e art, c'est aussi un marché en plein

1. https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:32019L0882&from=ES

(a)



(b)

FIG. 1 – Exemples de bulles de bandes dessinées

Figure	Texte reconnu	CER	WER
Figure 1a	LE DÉBUT D'UN OUR PASBLE DANS LE PETIT VILLASE QUE NOÏS CONNAISSONS BIEN...	0.08	0.31
Figure 1b	V cuérie/on % ENFIN REÇU LE CATALOGIE DE LA MANUFACTURE DES ARMES ET CHARS/	0.15	0.39

TAB. 1 – Résultats obtenus par ML-Kit sur les bulles des Figures 1

essor dans le monde. En France, le secteur a vu son taux de croissance augmenter de 50% par rapport à 2020, selon le bilan annuel dressé par GfK Market Intelligence début 2022 [2]. C'est par ailleurs un media qui est désormais accepté et reconnu, puisque salué il y a peu par le Ministère de la Culture avec l'année de la BD en 2020 [3] et plus récemment avec l'attribution de la chaire annuelle Création artistique du Collège de France [4].

Dans une première étude (Le Meur et al. (2022)), une comparaison de différents algorithmes de segmentation et de reconnaissance de texte dans des bandes dessinées à partir d'images capturées par un terminal mobile a été réalisée. Afin de pouvoir réaliser cette comparaison de manière équitable et non complaisante, une base de 50 images de bandes dessinées, issues de 14 ouvrages différents (BD franco-belges, mangas, comics, graphic novel) en français et en anglais, a été constituée. L'étude a révélé de sérieuses limitations des algorithmes d'OCR (Optical Character Recognition) les plus performants utilisant du Deep Learning, comme Tesseract [5] et ML-Kit [6]. Ce constat n'est pas forcément surprenant dans la mesure où notre cas d'usage présente des contraintes spécifiques : l'acquisition des images se fait à main levée par un terminal mobile et non à l'aide d'un scanner à plat, sous des conditions d'éclairage non maî-

2. https://www.gfk.com/fr/press/annee-2021-hors-norme-pour-les-acteurs-de-la-bd
3. https://www.culture.gouv.fr/Presse/Communiques-de-presse/BD-2020-l-Annee-de-la-bande-dessinee-prolongee-jusqu-au-30-juin-2021
4. https://www.college-de-france.fr/chaire/benoit-peeters-creation-artistique-chaire-annuelle
5. https://github.com/tesseract-ocr/tesseract
6. https://developers.google.com/ml-kit

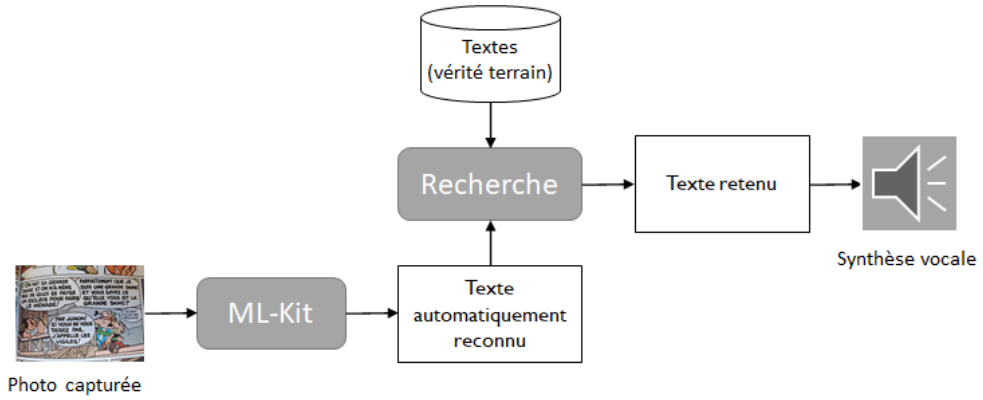


FIG. 2 – Workflow de notre système d'aide à la lecture de bande dessinée

trisées, et les bulles sont la plupart du temps manuscrites avec parfois des graphismes variés et complexes à déchiffrer (fond coloré, texte déformé, ...). La Table 1 présente des résultats de cette étude sur la reconnaissance de texte de ML-Kit sur les 2 images brutes (non segmentées) de la Figure 1. Le lecteur intéressé pourra trouver les résultats détaillés dans l'article original.

Ainsi le cas d'usage envisagé impose des contraintes fortes : luminosité de la prise de vue, angle de la prise de vue, spécificités du texte à reconnaitre, netteté du texte à reconnaitre, nécessité d'avoir une transcription exacte, rapidité du temps de réponse. Bien que les résultats soient prometteurs en terme des métriques utilisées (CER et WER, voir définition en section 2), la qualité des transcriptions obtenues n'est pas suffisante pour permettre une utilisation telle quelle. La possibilité d'utiliser des outils de correction orthographique pour post-traiter les transcriptions ou encore l'intégration de meilleurs systèmes de reconnaissance de texte (e.g. Rayar et Uchida (2019)), dédiés aux bandes dessinées est en cours d'étude et fera l'objet de futures communications. Néanmoins, le premier prototype réalisé dans le cadre de l'étude de Le Meur et al. (2022) reste non exploitable dans la mesure où la lecture vocale fait apparaitre de nombreuses erreurs, empêchant ainsi l'appropriation et l'adoption d'un tel dispositif d'aide à la lecture par nos publics cibles.

2 Approche proposée

En parallèle des nos travaux sur la reconnaissance automatique de texte dans les bandes dessinées et de l'amélioration de la qualité des transcriptions, nous proposons dans cet article une approche hybride. Cette approche permet de concevoir une solution robuste à notre cas d'usage, en exploitant des techniques d'OCR et des informations dites de vérité terrain. La Figure 2 illustre notre workflow. En amont, les informations vérité-terrain, nécessaires dans notre approche doivent être créées par un être humain : il s'agit de la transcription manuelle de l'ensemble des textes figurant dans chaque case présente dans les pages d'une bande dessinée. Bien qu'il soit possible de structurer cette vérité terrain, à l'aide de format tel que le CBML

FIG. 3 – Exemple de bulles se chevauchant, où un mauvais regroupement de blocs de textes peut s'opérer

(Comic Book Markup Language) présenté dans Walsh (2012), nous avons dans un premier temps opté pour une version brute ne faisant apparaitre que les texte bruts.

Lors de l'utilisation, le lecteur tire parti du système d'aide à la lecture en venant prendre une photo de la région de texte d'intérêt qu'il souhaite voir lue par l'application. Par la suite, ML-Kit est utilisé pour générer une transcription automatique. Cependant de manière à supprimer des écueils de regroupement des zones de texte généré par ML-Kit dans des cas de bulles complexes qui peuvent se chevaucher (voir la Figure 3), l'utilisateur doit indiquer le texte à lire dans l'image en l'englobant avec un rectangle, interaction classique sur des surfaces tactiles. Afin de pallier aux erreurs de transcriptions observées dans l'étude précédente (voir Table 1), la transcription est comparée à l'ensemble des textes vérité-terrain de notre base, et le texte de la base le plus proche est sélectionné. Les métriques utilisées pour comparer deux chaînes de texte et mesurer leurs différences sont classiquement le taux de caractères erronés (Character Error Rate / CER) et le taux de mots erronés (Word Error Rate / WER). Ces métriques se basent sur la distance d'édition ($dist$) entre deux séquences de caractères (s_{ref}, s_{rec}) normalisée par la taille de la séquence de référence (s_{ref}). On définit ainsi :

$$CER, WER = \frac{dist(s_{ref}, s_{rec})}{taille(s_{ref})}$$

La valeur du CER/WER obtenue est de 0 si les deux séquences sont identiques, de 1 si toutes les lettres ou tous les mots sont erronés, mais il est à noter qu'elle peut être supérieure à 1 en cas de nombreux faux-positifs dans la séquence reconnue (s_{rec}). Finalement, le texte de la base le plus proche est envoyé à un système de synthèse vocale. Nous avons dans le prototype développé utilisé l'API TextToSpeech de Google, intégré au système d'exploitation Android.

FIG. 4 – Image présentant multiples textes dans le décor

Texte présent dans la scène	Texte reconnu par ML-KIT	Vérité terrain associée
RESTAURANT GAULOIS	RESTAVRANT TANE GALOIS	RESTAURANT GAULOIS
VINS D'AQUITAINE ET DE NARBONNE	VINS TANE	VLAN !
SAUCISSON DE LUGDUNUM	-	-
SANGLIERS A LA BROCHE	SANGLS A LA BROCHE	SANGLIERS A LA BROCHE

TAB. 2 – Résultats obtenus par ML-Kit sur les élements de texte de la Figure 4

3 Évaluation

Afin d'exploiter notre prototype dans un cas réel d'utilisation, nous nous sommes concentré sur une bande dessiné : "Astérix gladiateur". Quatrième opus de la série Astérix scénarisée par René Goscinny et dessinée par Albert Uderzo, cet album de 44 planches a été publié en 1964. Pour réaliser une évaluation quantitative, 388 images ont été prises au smartphone, contenant 781 bulles/onomatopées. Sur ce jeu d'images, un taux de reconnaissance de 97% est atteint en ne tenant pas compte des onomatopées, difficilement géré par ML-Kit. En tenant compte des onomatopées, on descend à 92%. Le temps de traitement moyen observé est de $1407ms$. Ce temps de réponse, très souvent inférieur à 2 secondes, est occupé en grande majorité par les traitements liés à ML-Kit.

Parmi les erreurs de reconnaissance, on peut remarquer que les textes figurant sur des éléments constitutifs du décor de la scène posent souvent problème comme le montre la Figure 4. Dans cette image il y a quatre textes faisant partie intégrante du décor, deux d'entre eux sont partiellement reconnu par ML-Kit, mais la reconnaissance est suffisante pour permettre de retrouver dans la vérité terrain le texte exact à lire. Par contre, la reconnaissance des deux autres est si éloignée voire inexistante que le texte le plus approchant dans la vérité terrain (compte tenu de notre métrique) ne correspond pas du tout au texte qui devrait être lu. Ces éléments de décor textuel présentent des différences par rapport aux textes présents dans les bulles. Ils sont

FIG. 5 – Images présentant des éléments textuels de décor

Texte détecté par ML-Kit	Texte lu par notre application
NZ CALAe	ON NOUS ATTAQUE !
R	ROMA
UDERZO & GOSCINNY	-

TAB. 3 – Éléments textuels de décor de la Figure 5 mal reconnus

la plupart du temps sur un fond de couleur, et surtout sont soumis à la perspective visuelle du décor. Conséquence, avec une perspective marquée, sur une couleur de fond réduisant parfois le contraste des caractères, la séparabilité des lettres est réduite rendant la détection du texte difficile, voire impossible (Figure 5). Dès que la perspective des textes présents dans le décor est moins marquée, la reconnaissance des éléments textuels du décor ne pose plus de problème (Figure 6). Ce problème de reconnaissance des textes présentant une forte perspective, semble difficile à résoudre avec l'algorithme présenté ici, mais pourrait sans doute être réglé par un algorithme de traitement du type reconnaissance d'image associé à notre vérité terrain.

4 Conclusion et perspectives

Nous avons montré dans cet article qu'il était possible d'exploiter à bon escient une approche hybride IA/Humain pour créer une application robuste d'aide à la lecture de bande dessinée via un terminal mobile, tout en maintenant le contact avec le support physique de ladite bande dessinée.

Outre les travaux actuels que nous menons sur la reconnaissance de texte dans les bandes dessinées, plusieurs perspectives sont envisageables pour l'application présentée : (i) une évaluation et analyse de l'utilisation des utilisateurs in-situ (via des logs), (ii) une amélioration des interactions homme-machine de l'application, notamment par la prise en compte de la détection automatique des bulles (Dubray et Laubrock (2019)) pour éviter la sélection du texte d'intérêt par un rectangle englobant et (iii) une réflexion sur la structuration des textes vérité-terrain pour permettre de nouvelles fonctionnalités (prise en compte des personnages, des émotions).

FIG. 6 – Image présentant des éléments textuels de décor

Texte détecté par ML-Kit	Texte lu par notre application
ATTENTIoul SORTIER MENMIRS	CARRIÈRE OBÉLIX ATTENTION ! SORTIE DE MENHIRS
ATTENTION DALLES GHSSANTES 4	ATTENTION DALLES GLISSANTES
ROMA	ROMA
APODY TERIA	APODYTERIA

TAB. 4 – Éléments textuels de décor de la Figure 6 correctement reconnus

Références

Dubray, D. et J. Laubrock (2019). Deep cnn-based speech balloon detection and segmentation for comic books. In *2019 International Conference on Document Analysis and Recognition (ICDAR)*, pp. 1237–1243. IEEE Computer Society.

Le Meur, F., F. Rayar, S. Treuillet, et F. Daubignard (2022). Étude comparative de reconnaissance de texte dans les bandes dessinées. In *22e Conférence francophone sur l'Extraction et la Gestion des Connaissances*, Blois, France.

Rayar, F. et S. Uchida (2019). Comic text detection using neural network approach. In *Multi-Media Modeling*, pp. 672–683. Springer International Publishing.

Walsh, J. A. (2012). Comic book markup language : An introduction and rationale. *Digital Humanities Quarterly 6*(1).

Remerciements

Nous remercions le laboratoire PRISME pour le financement du stage de Clément Charrier et la société Algona, représenté par Frédéric Daubignard pour sa participation au projet.

Lecture assistée de texte dans les bandes dessinées

Summary

Text recognition in documents and natural scene images has seen great advances in the last few years, both in academic and industrial sectors. However, application of these technologies to specific fields, such as the accessibility to books for people who face reading difficulties, remain a challenge. In this paper, we propose an AI/Human hybrid approach to create a robust assistive reading system of comics using a mobile device, while maintaining the usage of the physical comic.