Extraction robuste des cases et du texte de bandes dessinées

Christophe Rigaud, Norbert Tsopze, Jean-Christophe Burie, Jean-Marc Ogier

Extraction robuste des cases et du texte de bandes dessinées

Christophe Rigaud, Norbert Tsopze, Jean-Christophe Burie, Jean-Marc Ogier

Laboratoire L3i - Université de La Rochelle
Avenue Michel Crépeau 17042 La Rochelle Cedex 1
{christophe.rigaud, norbert.tsopze, jcburie, jmogier}@univ-lr.fr

RÉSUMÉ. *Les bandes dessinées représentent un patrimoine culturel important dans de nombreux pays. La numérisation en masse offre l'opportunité d'effectuer des recherches sur le contenu des albums et pas uniquement sur des métadonnées associées (e.g. nom de l'auteur ou de la collection). Peu de travaux ont été menés à ce jour. Seule l'extraction des cases et des bulles de dialogues a été étudiée et ce, pour des structures de pages relativement simples. En effet, la structure des pages est propre à chaque auteur, ce qui engendre une très grande diversité de dessins. Malgré cette diversité, les dessins ont une particularité commune de part leurs méthodes de conception : ils sont constitués ou entourés d'un trait noir. Dans cet article, nous proposons de nous appuyer sur cette particularité des bandes dessinées pour extraire automatiquement les cases et le texte avec une méthode basée sur la classification de composantes connexes. Nous comparerons notre méthode avec des outils de la littérature et discuterons des résultats.*

ABSTRACT. *Comic books represents an important heritage in many countries. Nowadays, digitisation allows to search directly from content instead of metadata only (e.g. album title). Few studies has been done in this direction. Only frame and speech balloon extraction have been experimented in the case of a simple page structure. In fact, the page structure depends on the author which is why many different structures and drawings exist. Despite of the differences, drawings have a common characteristic because of design process: they are all surrounded by a black line. In this paper we propose to rely on this particularity of comic books to automatically extract frame and text using a connected-component labeling analysis. The approach is compared with some existing methods found in the literature and results are presented.*

MOTS-CLÉS : *bandes dessinées, segmentation de case, segmentation de texte, composantes connexes, k-means*

KEYWORDS: *comic books, frame extraction, text extraction, connected-component, k-means*

1. Introduction

La bande dessinée représente un patrimoine culturel important pour beaucoup de pays, elle est confrontée aujourd'hui à la numérisation en masse par besoin d'archivage et de valorisation du contenu. Cette dématérialisation à échelle industrielle permet actuellement l'indexation des pages mais pas de leurs contenus. Si cette "limite à la page" peut être franchie, alors de nouvelles utilisations du 9ème art sont envisageables comme la lecture sur appareil mobile case par case (Arai *et al.,* 2011, In *et al.,* 2011) ou encore la recherche d'éléments spécifiques dans un grand nombre d'albums pour, par exemple, interpréter le contenu par le texte. Ces applications sont aujourd'hui réservées uniquement aux bandes dessinées dites "électroniques" (e-comics) puisque elles sont créées directement avec l'outil informatique et donc potentiellement indexables tout au long de la conception. Pour permettre la valorisation du contenu des bandes dessinées, des traitements postérieurs sont à l'étude depuis quelques années mais ils ne sont pas encore assez robustes pour être utilisés par l'industrie de la dématérialisation. Ces travaux concernent l'extraction des cases, des bulles et du texte à l'intérieur des bulles. Ce document propose une méthode pour extraire automatiquement les cases et tout le texte contenu dans les bandes dessinées (pas uniquement celui contenu par les bulles). La méthode proposée s'appuie sur l'algorithme des composantes connexes suivi d'une classification par l'algorithme des k-Means (Tou *et al.,* 1974).

Ce document est organisé comme suit. La section 2 présente le vocabulaire et le contenu des bandes dessinées. Un état de l'art sur l'extraction de cases et de textes est détaillé en section 3. Les sections 4 et 5 présentent respectivement la méthode proposée et les résultats obtenus lors de l'expérimentation. Et enfin, une conclusion et des perspectives sont données en section 6.

2. Les bandes dessinées

Les bandes dessinées sont pour la plupart une succession de dessins organisés pour raconter une histoire selon un scénario écrit par l'auteur. La seule contrainte à laquelle doit se soumettre le dessinateur est le support qui contiendra la bande dessinées. C'était, jusqu'à l'émergence des supports numériques et d'internet, une page de papier souvent au format A4. Autant dire que le mode de conception laisse une quasi totale liberté d'expression à l'auteur ce qui fait la richesse de son album mais aussi la diversité des éléments que l'on peux rencontrer d'un point de vue traitement d'images.

2.1. *Le vocabulaire*

Une bande dessinée raconte une histoire, cette histoire est contenue dans un album. Les albums sont constitués de planches (assimilables aux pages) qui contiennent plusieurs cases. Une case est une image ou une vignette contenant un dessin est généralement encadrée. Notons qu'une bande dessinée ne comporte pas nécessairement

de case. Dans ce cas, la case se confond avec la planche. Une succession de cases alignées est appelée une bande d'où le nom de bandes dessinées.

2.2. *Le contenu des planches*

On distingue trois types de bandes dessinées (Ponsard *et al.,* 2008) reconnaissables par leurs origines culturelles : États-Unis, Asie (manga) et Europe. Dans notre étude nous nous concentrerons sur les bandes dessinées Européennes et Américaines car, pour les mangas, de trop grandes différences existent avec les bandes dessinées dites "classiques" tant en terme de traits, de cases (Yamada *et al.,* 2004) et de textes (Arai *et al.,* 2011).

2.2.1. *Les contours*

Une observation attentive du contenu des planches permet de dire que la principale caractéristique des dessins de bande dessinée est le trait noir qui entoure chaque élément ou presque contenu dans une planche. C'est à partir de cette observation qu'il nous est apparu pertinent de baser notre méthode sur l'algorithme des composantes connexes (CC) pour extraire le contenu des cases à partir de ses contours. Cet algorithme à un double intérêt dans notre étude car il est, de part l'observation ci-dessus, adapté à l'extraction des cases mais aussi à l'extraction du texte (Fletcher *et al.,* 1988). L'utilisation d'un même algorithme pour extraire tous les éléments d'une planche améliore nettement le temps de traitement.

2.2.2. *Les cases*

Les cases représentent différentes scènes de l'histoire comme des arrêts sur image pour le cinéma. Elles contiennent toutes un dessin mais ne sont pas nécessairement encadrées. C'est d'ailleurs ce qui peut rendre la lecture (ou la segmentation) plus difficile. Il peut y avoir aussi des recouvrements partiels entre les cases ou encore l'apposition d'éléments sur plusieurs cases (Ho *et al.,* 2011), autant de particularités qui peuvent, ponctuellement, venir perturber le traitement d'un algorithme comme CC.

2.2.3. *Les zones de texte*

La bande dessinée utilise différents types de textes, manuscrits ou dactylographiés, selon la nature du message à transmettre au lecteur. On trouve majoritairement des dialogues entre les personnages très souvent dans des bulles à fond blanc, puis du texte récitatif (la voix "off") et enfin des onomatopées qui représentent des bruits sous forme textuelle ou d'une suite de symboles.

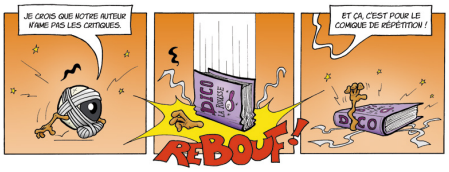
3. Méthodes existantes

3.1. *Extraction des cases*

L'extraction automatique des cases a principalement été étudiée pour permettre la lecture de pages numérisées sur des écrans d'appareils mobiles case par case. Notre étude concerne plutôt l'indexation du contenu de banques d'albums ce qui soulève de nouveaux problèmes en terme de diversité de format, de résolution et de contenu.

Plusieurs techniques de séparation du fond et des cases ont étaient proposées depuis 2007 comme le montre (In *et al.,* 2011), la plupart sont basées sur un découpage par des lignes tracées selon la transformée de Hough (Duda *et al.,* 1972), l'algorithme X-Y récursif (Han *et al.,* 2007) ou encore à partir de gradient (Tanaka *et al.,* 2007) ; ce qui ne permet pas de prendre en compte les zones vides qui peuvent apparaître entre les cases (In *et al.,* 2011) ou les éléments sans bordure (figure 1a). Cet aspect a été corrigé par des approches basées sur les composantes connexes (Arai *et al.,* 2010) mais lorsque des éléments chevauchent plusieurs cases (figure 1b) il devient difficile de séparer les cases. Les régions d'intérêts (ROI) sont souvent classifiées selon un filtrage heuristique (Arai *et al.,* 2011, Ho *et al.,* 2011) relatif à la taille de la page ce qui est très sensible au changement de format de page. Une succession de N érosions suivies de N dilatations a été proposée par (Ho *et al.,* 2011) pour "couper" les éléments qui chevauchent plusieurs cases mais la détermination de la valeur de N reste floue et le temps de calcul qui en découle important. Ce dernier propose aussi une segmentation du fond par croissance de région au lieu de la binarisation utilisée majoritairement par les méthodes citées précédemment.

(a) Encadrement partiel (Cyb, 2008)



(b) Chevauchement d'éléments sur plusieurs cases (Lamisseb, 2011)

Figure 1. *Exemple de cases particulières.*

3.2. *Extraction du texte*

Dans une bande dessinée, la plus grande partie du texte est contenue dans les bulles de dialogue. C'est sans doute pour cette raison que c'est le seul type de texte qui a été étudié à ce jour. Ces travaux consistent en l'extraction du texte à partir des bulles

(Yamada *et al.,* 2004, Arai *et al.,* 2010) ou inversement des bulles à partir du texte (Ho *et al.,* 2011) ; Ces approches donnent de très bons résultats mais elles font l'hypothèse que le texte est écrit en noir dans une bulle blanche. Dans notre étude nous élargirons cette hypothèse de manière à ce que le fond des zones de textes soit d'une intensité proche de celle du fond de la page.

4. Contribution

Nous présentons une nouvelle méthode pour extraire simultanément les cases et les zones de textes d'une planche de bandes dessinées afin de permettre une indexation du contenu. Notre méthode s'applique page par page et consiste à effectuer un prétraitement pour obtenir une image binarisée à partir de laquelle on définit l'ensemble des ROI comme étant l'ensemble des boîtes englobantes (rectangulaires) des composantes connexes. Ensuite, les ROI sont classifiées comme étant du bruit, du texte ou une case en fonction de leurs tailles, relations topologiques et relations spatiales (pour le texte uniquement). Notons que nous ne limiterons pas notre étude au texte contenu dans les bulles mais à tout le texte contenu dans une planche à l'exception des onomatopées qui feront l'objet d'un travail ultérieur. L'originalité de notre travail est l'extraction des cases, encadrées ou non, et du texte récitatif (la voix "off") puisqu'il peut être détecté par l'algorithme des CC au même titre que celui contenu dans les bulles.

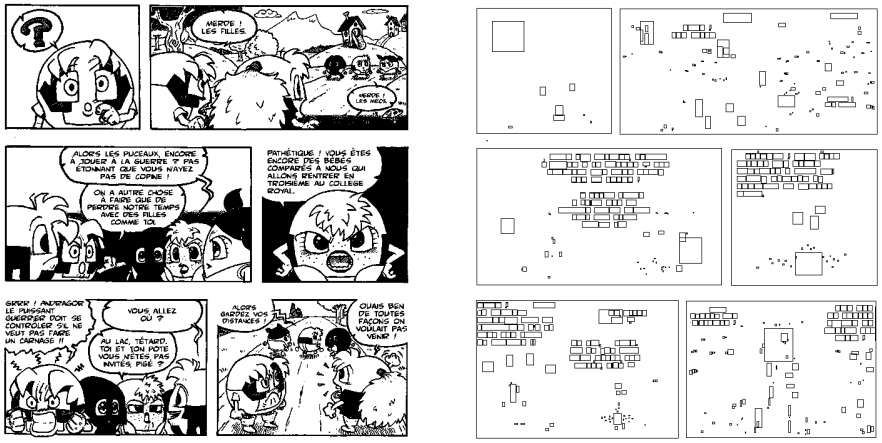
4.1. *Pré-traitement*

Le pré-traitement a pour but de séparer le fond et le contenu de la page afin de pouvoir se concentrer sur le contenu par la suite. Pour cela, plusieurs traitements sont appliqués à l'image pour pouvoir appliquer l'algorithme des CC et extraire les régions d'intérêts par la suite. Ces traitements suivent la séquence suivante :

1) Conversion en niveaux de gris
2) Calcul du seuil de binarisation
3) Inversion de l'image selon le seuil trouvé à l'étape précédente
4) Binarisation
5) Extraction des composantes connexes

La première étape consiste à convertir l'image en niveaux de gris selon (Pratt *et al.,* 1991) puis on procède à une binarisation (figure 2a) dont le seuil est calculé à partir de la médiane d'un échantillon de pixels répartis sur le pourtour de la page. Pour les expérimentations, nous avons choisi de prendre cinq pixels par coté. Si la médiane est plus proche du noir que du blanc alors une inversion de l'image est effectuée et le calcul relancé de manière à toujours avoir un fond blanc à l'issue du prétraitement. Ce prétraitement permet à la méthode d'être plus robuste que (Arai *et al.,* 2011) qui utilise un seuil constant pour la binarisation. Cette dernière est une opération clef pour la suite des traitements. Notons qu'il est possible que certaines zones de texte soient

"effacées" lors de la binarisation si l'intensité du fond est supérieure au seuil. Ensuite, on détermine l'ensemble des composantes connexes pour pouvoir extraire, à partir de celles-ci, les boîtes englobantes de tous les éléments (suite continue de pixels noirs) de l'image (figure 2b).



(a) Page binarisée (Cyb, 2009) (b) Ensemble des boîtes englobantes

Figure 2. *Étapes de pré-traitements*

4.2. *Classification des régions d'intérêts*

Les régions d'intérêts sont définies comme étant les boîtes englobantes des composantes connexes de l'image binarisée. On définit l'ensemble des régions d'intérêts $R = \{R_1, R_2, ..., R_n\}$. La classification des régions d'intérêts est basée sur les hauteurs des ROI (figure 3). On distingue alors trois classes que l'on interprète comme des "cases" pour (les plus grandes), du "texte" (les plus nombreuses) et du "bruit" (généralement des éléments de moins de cinq pixels). Cette classification est effectuée dynamiquement à partir de l'algorithme des K-Means ce qui rend notre méthode invariante au changement de résolution et de format de la page. En effet, la méthode des K-Means ne dépend pas des dimensions de la page contrairement aux méthodes proposées par (Ho *et al.,* 2011, Arai *et al.,* 2011), et la taille en pixel des éléments de la page est proportionnelle à la résolution.

A l'issue de la classification, on calcule la variance de chaque classe pour vérifier l'homogénéité des ROI. Si elle est élevée, on peut appliquer des algorithmes spécifiques pour d'améliorer la binarisation et/ou la classification. Prenons l'exemple de la figure 4 qui montre le résultat de deux cases reliées par une flèche noire (figure 4a). Celles-ci sont identifiées à tort comme une seule et même case par l'algorithme CC (figure 4b). On peut voir sur l'histogramme 4c (échelle logarithmique) que la première région est beaucoup plus grande que les autres, la variance en sera de même. Dans ce

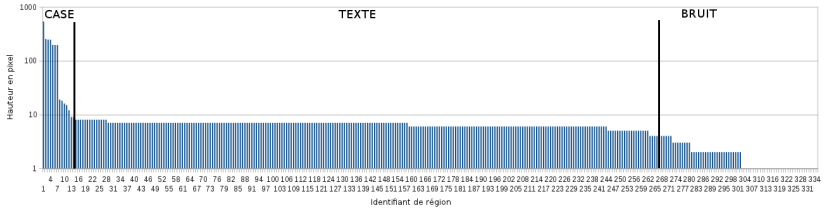


Figure 3. *Histogramme décroissant des hauteurs des ROI et classification*

cas nous appliquons l'algorithme de (Ho *et al.,* 2011) qui consiste en une croissance de région suivie d'une succession d'érosions et dilatations pour supprimer les éléments chevauchants (ici une flèche), puis on réitère le prétraitement et la classification des cases.

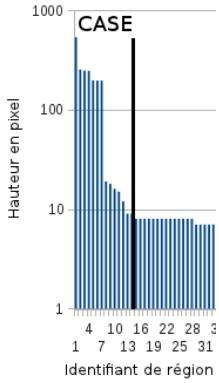
4.3. *Filtrage*

A l'issue de la classification, deux filtrages sont effectués pour supprimer les faux positif (régions identifiées à tort). Le premier est de type topologique et consiste à ne conserver que les "cases" qui ne sont pas incluses dans une autre ($R_i \notin R_j \forall j, i \neq j$) (figure 5a et 5b).

Le second forme des zones de texte à partir des ROI identifiées comme "texte" (correspondant souvent à une ou plusieurs lettres) qui sont suffisamment proches les unes des autres. Le seuil utilisé est calculé à partir de la valeur médiane de la classe "texte" (figure 6).

Il arrive que les zones de textes ainsi crées ne contiennent pas de texte lisible mais un ensemble de petits traits de hauteur similaire au texte (voir figure 7). Dans ce cas nous utilisons une méthode de séparation texte/graphique pour éliminer les zones qui ne contiennent pas de texte. Cette méthode compare les histogrammes projetés (horizontal et vertical) du nombre de pixels blancs des zones identifiées comme "texte". Expérimentalement on vérifie, pour chaque zone de texte identifiée, que la variance du nombre de pixels blancs par ligne est supérieure à la variance du nombre de pixels blancs par colonne. Ceci s'explique par le fait que dans l'histogramme projeté horizontal d'une zone de texte il y a de fortes variations dues à l'alternance de lignes de texte et d'interlignes. Ce phénomène ne se produit pas sur l'histogramme projeté vertical (figure 8).

(a) Élément chevauchant (Jolivet, 2010) (b) Case issues de CC



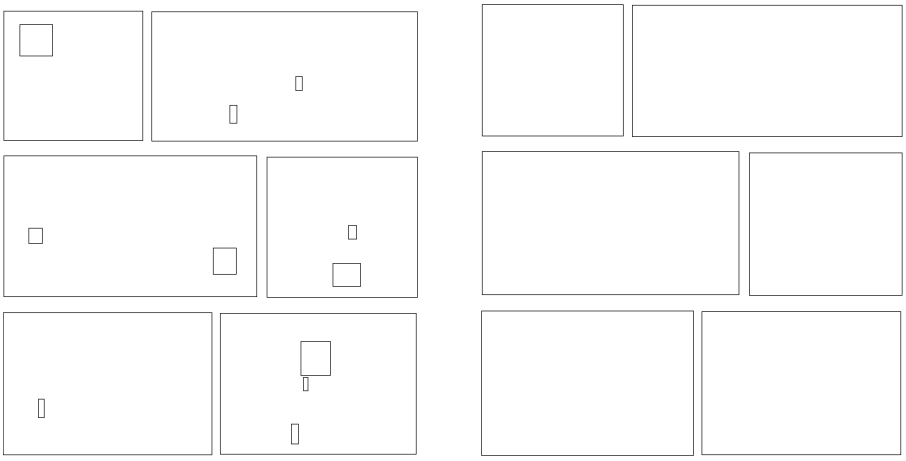
(c) Histogramme partiel

Figure 4. *Détection des défauts de détection des cases*

5. Expérimentation et résultats

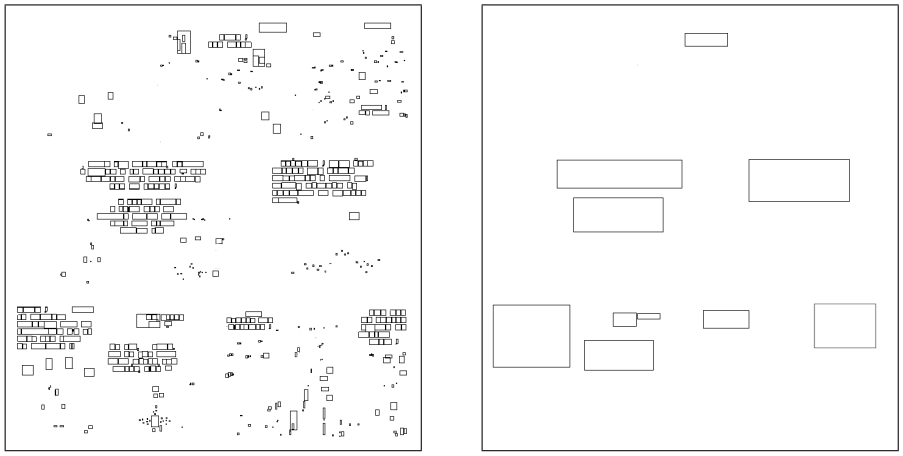
5.1. *Extraction des cases*

Nous avons réalisé l'expérimentation avec la même base d'albums que (Ho *et al.,* 2011) afin de pouvoir comparer nos résultats. A savoir, une base composée de 42 pages au format A4 issues de 7 auteurs différents de bandes dessinées européennes et américaines soit 355 cases au total. Chacune des méthodes testées ont été implémentées et appliquées sur la même base d'images. Pour l'évaluation nous considèrerons plusieurs critères. Le premier est le taux de segmentation des pages. Une page est correctement segmentée si toutes ses cases sont correctement segmentées. Le second s'applique



(a) Classification des cases par K-Means (b) Résultat du filtrage topologique des cases

Figure 5. *Filtrage topologique des cases*

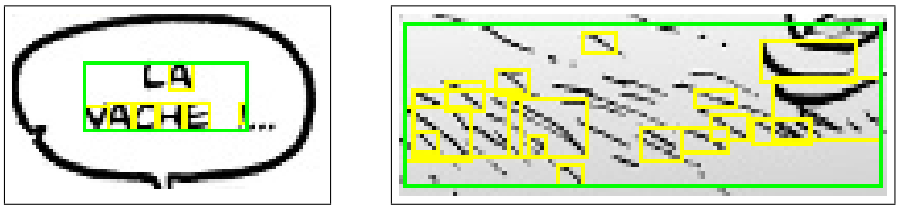


(a) Classification du texte par K-Means (b) Zones de texte identifiées

Figure 6. *Filtrage spatial du texte*

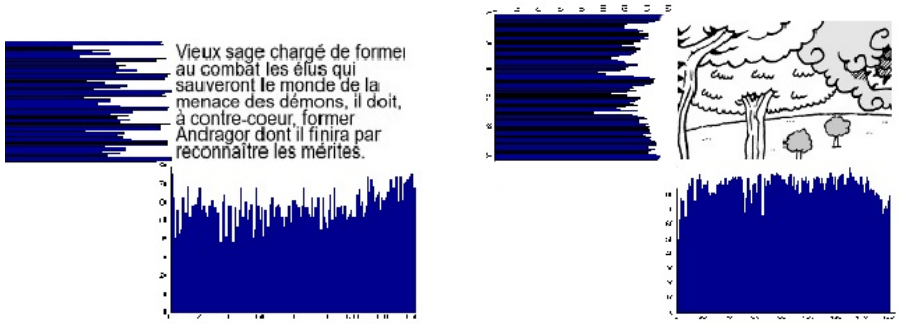
aux cases elles-même et indique la proportion de cases correctement segmentées sur l'ensemble de la base.

Notre méthode est plus efficace que (Tanaka *et al.,* 2007, Arai *et al.,* 2010, Ho *et al.,* 2011) pour l'extraction automatique des cases notamment parce qu'elle permet d'extraire les cases sans contour. Lors des tests, nous avons pu constater un gain de temps global de plus de 60% par rapport à (Ho *et al.,* 2011). Notre méthode est



(a) Zone de texte bien identifiée (b) Zone de texte mal identifiée

Figure 7. *Zone de texte (ensemble des rectangles encadrés)*



(a) Histogrammes d'une zone de texte (b) Histogrammes d'une zone de dessin

Figure 8. *Exemple d'histogrammes projetés (nombre de pixels blanc)*

Méthode	Tanaka	Arai	Khoi	Proposée
Page (%)	42.8	47.6	64.3	66.7
Case (%)	63.9	75.6	87.3	88.2

Figure 9. *Taux de réussite des méthodes testées.*

donc beaucoup plus rapide, notamment quand il n'y a pas d'objet à cheval sur plusieurs cases. Néanmoins, cela n'influence pas l'extraction du texte puisque qu'elle est effectuée à partir de la page et non de la case.

5.2. *Extraction du texte*

Nous avons effectué l'extraction des zones de textes à partir de la même base d'images que précédemment. Pour plus de précision, nous avons distingué les zones de texte de dialogues (bulles) et les zones de textes narratifs soit respectivement 435 et 79 zones dans notre base. On définit :

– VP : les zones identifiées comme “texte” qui contiennent uniquement du texte (“Vrai positif”)

– FN : les zones qui contiennent du texte qui n’ont pas été identifiées (“Faux négatif”)

Les zones dont le texte a été détecté partiellement ou en plusieurs parties sont considérées comme appartenant à la catégorie “Faux négatif”.

Type de texte	VP	FN
Dialogue (%)	78	21
Narratif (%)	53	47

Figure 10. *Taux de détection des zones de textes*

Nos résultats préliminaires sont encourageants pour la catégorie des dialogues mais nécessitent encore des améliorations pour le texte narratif. Il est difficile de comparer avec d’autres travaux puisque, à notre connaissance, il semblerait que l’extraction du texte en dehors des bulles (zone à fond blanc) n’ait jamais été étudiée en analyse d’image de bandes dessinées.

6. Conclusion et perspectives

Nous avons proposé et évalué une méthode d’extraction simultanée des cases et du texte de bandes dessinées. Notre approche est rapide et particulièrement robuste au changement de format de page. Elle est également capable d’extraire le texte situé en dehors des bulles.

L’évaluation montre qu’il y a encore des améliorations à apporter au niveau des éléments à cheval sur plusieurs cases et du filtrage des zones de texte. Les travaux futurs porterons sur l’amélioration de la classification du texte (e.g. dialogue, narratif) et l’extraction du contenu des cases.

7. Remerciements

Ce travail est réalisé dans le cadre de l’action e-BDthèque du contrat de plan État-Région avec le soutien financier de la région Poitou-Charentes, le conseil général de Charente Maritime et la communauté d’agglomération de La Rochelle.

8. Bibliographie

Arai K., Tolle H., « Method for Automatic E-Comic Scene Frame Extraction for Reading Comic on Mobile Devices », *Seventh International Conference on Information Technology : New Generations*, ITNG '10, IEEE Computer Society, Washington, DC, USA, p. 370-375, 2010.

Arai K., Tolle H., « Method for Real Time Text Extraction of Digital Manga Comic », *International Journal of Image Processing (IJIP)*, vol. 4, n° 6, p. 669-676, 2011.

Cyb, *La légende des Yaouanks*, Studio Cyborga, Goven, France, 2008.

Cyb, *Bubblegôm*, Studio Cyborga, Goven, France, 2009.

Duda R. O., Hart P. E., « Use of the Hough transformation to detect lines and curves in pictures », *Commun. ACM*, vol. 15, p. 11-15, January, 1972.

Fletcher L., Kasturi R., « A robust algorithm for text string separation from mixed text/graphics images », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, n° 6, p. 910-918, November, 1988.

Han E., Kim K., Yang H., Jung K., « Frame segmentation used MLP-based X-Y recursive for mobile cartoon content », *Proceedings of the 12th international conference on Human-computer interaction : intelligent multimodal interaction environments*, HCI'07, Springer-Verlag, Berlin, Heidelberg, p. 872-881, 2007.

Ho A. K. N., Burie J.-C., Ogier J.-M., « Comics page structure analysis based on automatic panel extraction », *GREC 2011, Nineth IAPR International Workshop on Graphics Recognition*, Seoul, Korea, September, 15-16, 2011.

In Y., Oie T., Higuchi M., Kawasaki S., Koike A., Murakami H., « Fast frame decomposition and sorting by contour tracing for mobile phone comic images », *Internatinal journal of systems applications, engineering and development*, vol. 5, n° 2, p. 216-223, 2011.

Jolivet O., *BostonPolice*, Clair de Lune, Allauch, France, 2010.

Lamisseb, *Les noeils Tome 1*, Bac@BD, Valence, France, 2011.

Ponsard C., Fries V., « An Accessible Viewer for Digital Comic Books », *ICCHP, LNCS 5105*, Springer-Verlag Berlin Heidelberg, p. 569-577, 2008.

Pratt, K. W., *Digital image processing (2nd ed.)*, John Wiley & Sons, Inc., New York, NY, USA, 1991.

Tanaka T., Shoji K., Toyama F., Miyamichi J., « Layout Analysis of Tree-Structured Scene Frames in Comic Images. », *IJCAI'07*, p. 2885-2890, 2007.

Tou J., Gonzalez R., *Pattern Recognition Principles*, Addison-Wesley, USA, 1974.

Yamada M., Budiarto R., Endo M., Miyazaki S., « Comic Image Decomposition for Reading Comics on Cellular Phones », *IEICE Transactions*, vol. 87-D, n° 6, p. 1370-1376, 2004.