# Extraction Multicritère de Texte Incrusté dans les Séquences Vidéo

Sébastien Lefèvre[1] – Cyril L'Orphelin[2] – Nicole Vincent[3]

[1] LSIIT – Université Louis Pasteur (Strasbourg I)
Parc d'Innovation, boulevard Brant, BP 10413, 67412 Illkirch Cedex
lefevre@lsiit.u-strasbg.fr

[2] LI – Université François Rabelais (Tours)
64, avenue Portalis, 37200 Tours

[3] CRIP5 – Université René Descartes (Paris V)
45, rue des Saints Pères, 75270 Paris Cedex 06
nicole.vincent@math-info.univ-paris5.fr

**Résumé** : *Dans cet article, nous abordons le problème de la détection des zones de texte dans les séquences vidéo. Contrairement à la plupart des approches existantes basées sur un unique détecteur suivi par un post-traitement ad hoc et coûteux, nous considérons plusieurs détecteurs et nous fusionnons leurs résultats afin de combiner les avantages de chaque détecteur. Nous débutons cet article par une étude des zones de texte incrustées dans les séquences vidéo pour déterminer comment ces zones apparaissent dans les images et identifier leurs principales caractéristiques (constance de la couleur et contraste avec l'arrière-plan, densité et régularité des contours, persistence temporelle). En se basant sur ces caractéristiques, nous sélectionnons ou définissons ensuite les détecteurs appropriés et nous comparons plusieurs stratégies de fusion qui peuvent être utilisées. Le processus logique que nous avons choisi et les résultats satisfaisants que nous avons obtenus nous permettent de valider notre contribution.*

**Mots-clés** : Caractéristiques du texte incrusté, Détection du texte incrusté, Stratégies de fusion.

## 1 Introduction

De nos jours, la quantité de données multimédia est tellement importante qu'elle nécessite des outils d'indexation afin de permettre aux utilisateurs de naviguer parmi les informations disponibles et de chercher des extraits pertinents. Parmi ces outils d'indexation, les détecteurs de changement de plan ou de changement de scène rendent possible la décomposition des séquences vidéo, tandis que les extracteurs d'images-clé peuvent être utilisés pour publier un index visuel du contenu vidéo. Une autre information cruciale est le texte incrusté présent dans les trames vidéo.

Une fois le texte extrait des séquences vidéo, il peut être utilisé pour créer des annotations textuelles des données, ensuite facilement indexables. La détection du texte incrusté dans les séquences vidéo est donc un problème majeur de l'indexation de données multimédia. Dans cet article, nous nous focalisons sur l'extraction du texte incrusté pour laquelle nous proposons une approche efficace. Nous devons préciser que la détection du texte de scène est considérée comme un autre problème et n'entre pas dans le cadre de notre étude. Avant de décrire les principaux aspects de notre contribution, nous allons rappeler brièvement les travaux effectués dans ce domaine.

Les approches décrites dans la littérature peuvent être classées selon les caractéristiques d'image utilisées : la couleur [JAI 98, JUN 01, KIM 96, ZHO 95], la texture [ZHO 00, WU 99, CHA 01], le mouvement [GAN 00], ou les contours [LIE 02, DIM 00]. Certaines méthodes utilisent également la constance temporelle des zones de texte incrusté [LI 00]. La plupart de ces méthodes partagent le même processus de traitement : un détecteur principal est tout d'abord utilisé (en se basant généralement sur l'une des caractéristiques) et un post-traitement est ensuite nécessaire afin d'améliorer la qualité des résultats. Ce post-traitement généralement *ad hoc* est souvent caractérisé par une complexité algorithmique élevée. Comme d'autres auteurs [LI 01, HUA 01], nous avons plutôt choisi d'élaborer notre méthode comme une combinaison de plusieurs détecteurs afin de prendre en compte les avantages de toutes les caractéristiques et d'éviter une étape de post-traitement. De plus nous sommes partis des caractéristiques des zones de texte incrusté pour déterminer les détecteurs optimaux.

Nous décrirons tout d'abord le but des zones de texte incrusté dans les séquences vidéo et étudierons comment ces zones sont représentées dans les trames vidéo. Cette étude nous aidera dans un second temps à déterminer les principales caractéristiques à partir desquelles nous pourrons construire les détecteurs les plus appropriés. Ces détecteurs feront l'objet de la section suivante. Nous aborderons ensuite les stratégies de fusion des différents détecteurs. Finalement nous décrirons et commenterons les résultats obtenus qui nous ont permis d'évaluer notre contribution.

## 2 Etude des caractéristiques du texte incrusté

Afin de déterminer quelles sont les caractéristiques les plus adaptées pour détecter les zones de texte incrusté dans les séquences vidéo, nous devons tout d'abord étudier ces zones. Différentes raisons peuvent être invoquées par l'équipe de

production audiovisuelle pour insérer artificiellement du texte dans les trames vidéos. Ce texte peut avoir différents objectifs : commercial pour mentionner des entreprises ou des membres participant à un film, sportif pour informer de l'évolution du jeu (temps écoulé, score) et de données associées (noms des joueurs, statistiques), informatif pour décrire le reportage en cours (nom du journaliste, lieu, résumé), légal pour indiquer les droits associés à un document, *etc*. Cependant, le texte incrusté joue dans tous les cas un rôle spécifique dans la séquence vidéo, et doit donc être facilement visible par un quelconque spectateur.

Plus précisément, à partir de l'examen d'un corpus vidéo varié, on peut remarquer que, puisque ce texte doit être facilement lu, les zones de texte sont clairement dissociées du reste du contenu de l'image. Le contraste entre les zones de texte et l'arrière-plan du texte ou de la scène est relativement élevé. De plus, ces zones sont affichées en premier-plan et ne sont jamais cachées : elles sont donc toujours complétement visibles. Nous pouvons aussi remarquer que la plupart du temps les caractères sont monochromes, et les lettres appartenant à un même mot sont généralement affichées avec des couleurs et textures uniques. La dernière idée liée à la constance de couleur de ces zones de texte est la caractéristique planaire du texte : tandis que le contenu de l'image représente généralement des données 3-D, les zones de texte sont situées sur un plan 2-D.

Une autre caractéristique importante des zones de texte incrusté est liée à la forme des caractères. Ces caractères ont la plupart du temps une taille, une forme, et une orientation constantes. La taille des caractères est généralement déterminée en suivant des règles de lisibilité : tous les caractères ont la même taille, les espaces les séparant sont constants, le nombre de mots par ligne est inférieur ou égal à 5, *etc*. Selon la région géographique, le texte peut être lu de gauche à droite, de droite à gauche, ou de haut en bas. Un autre aspect intéressant, les zones de texte comportent généralement de nombreux contours, réguliers de surcroît. C'est la concaténation des caractères issus d'une même police, et ces caractères doivent être facilement différenciables de l'arrière-plan.

Finalement, nous pouvons aussi remarquer la constance temporelle des zones de texte dans les séquences vidéo : ces zones ne se déplacent spatialement que rarement d'une trame à l'autre, et si c'est le cas le déplacement est faible. De plus, les caractères composant le texte apparaîssent dans les trames successives.

Les conclusions préliminaires de notre étude nous permettent de considérer que les caractéristiques des zones de texte incrusté peuvent être déterminées *a priori* afin de définir des détecteurs associés optimaux. Ces caractéristiques sont liées à la constance de la couleur et de la texture de ces zones et à son contraste avec les autres parties de l'image, aux formes régulières des caractères du texte et à leur densité de contours élevée, et à la persistence temporelle des zones de texte. A partir de ces conclusions, nous sommes maintenant capables de définir les détecteurs de texte incrusté appropriés. Afin d'assurer à notre méthode l'efficacité la plus grande possible, nous sélectionnerons principalement des détecteurs de faible coût de calcul.

# 3 Description des détecteurs sélectionnés

D'après les remarques de la section précédente, nous pouvons conclure que les zones recherchées sont des régions de l'image caractérisées par une couleur et une texture uniformes et un important contraste avec l'arrière-plan, des contours denses et réguliers pour délimiter les différentes lettres du texte, et une persistence temporelle sur plusieurs trames. Puisque nous tenons compte du temps de calcul, nous nous focaliserons sur des détecteurs rapides et complémentaires. Nous avons décidé de retenir trois types de détecteurs basés respectivement sur des informations de couleur, de texture, et de contours.

## 3.1 Détecteurs liés à la couleur

Le premier détecteur que nous utiliserons se base sur la constance couleur des zones de texte. Comme le contraste entre ces zones et l'arrière-plan est élevé, nous pouvons supposer que les composantes couleur du texte représentent un maximum local dans l'histogramme couleur. Nous considérons également que dans une zone de l'image, le texte est représenté par une minorité de pixels. Ces deux constats nous permettent de définir deux seuils ($s_b$ et $s_h$) utilisés dans l'analyse des histogrammes locaux des régions pour localiser les possibles zones de texte. D'autre part, les zones de texte étant de couleur uniforme, elles seront représentées par des couleurs uniques dans l'histogramme et non des plages de couleur. Pour assurer la lisibilité du texte, sa couleur n'est pas "bruitée" localement par des couleurs voisines. Cette remarque nous amène à formuler un second critère utilisé lors de l'analyse des couleurs : la répartition de la couleur des zones de texte devra être suffisamment différente des couleurs voisines. Cette différence peut être estimée par une mesure dérivative.
Nous débutons donc notre analyse par un découpage de l'image $I$ en blocs de taille $B_h \times B_l$. L'histogramme local $H$ de répartition des couleurs est ensuite calculé individuellement pour chaque bloc. Afin d'accroître la robustesse de cette analyse, nous avons décidé de réduire le nombre de composantes couleur à utiliser. En se basant sur l'intervalle de validation $[s_b, s_h]$ et la dérivée première des valeurs de l'histogramme, on détermine les couleurs des zones de texte. Cette sélection peut se formuler comme suit :
$H_c$ est une couleur de texte si :

$$\begin{cases} s_b < H_c < s_h \\ |H_c - H_{c-1}| + |H_c - H_{c+1} > S_d| \end{cases} \quad (1)$$

où $H_c$ est le nombre de pixels de couleur $c$ dans l'histogramme et où $S_d$ est un seuil prédéfini. Les pixels correspondants sont alors étiquetés.

## 3.2 Détecteurs liés à la texture

Afin de situer les zones de textes en se basant sur leur constance de texture, nous utilisons les ondelettes de Haar [LI 00]. Cependant, nous avons décidé de limiter notre analyse au niveau un pour conserver les régions de texture régulière.

Une décomposition de l'image $I$ selon les directions hotizontale, vecticale et diagonale permet d'obtenir trois images $I_{LH}$, $I_{HL}$ et $I_{HH}$ :

$$I_{LH}(i,j) = \frac{1}{4}\begin{pmatrix} I(2i,2j) - I(2i+1,2j+1) \\ +I(2i+1,2j) - I(2i,2j+1) \end{pmatrix} \quad (2)$$

$$I_{HL}(i,j) = \frac{1}{4}\begin{pmatrix} I(2i,2j) - I(2i+1,2j+1) \\ -I(2i+1,2j) + I(2i,2j+1) \end{pmatrix} \quad (3)$$

$$I_{HH}(i,j) = \frac{1}{4}\begin{pmatrix} I(2i,2j) + I(2i+1,2j+1) \\ -I(2i+1,2j) - I(2i,2j+1) \end{pmatrix} \quad (4)$$

où $I(i,j)$ représente la valeur du pixel de coordonnées $(i,j)$ dans l'image $I$. Nous calculons alors l'image de taille réduite $I'$ obtenue par seuillage de la somme des $I_{LH}$, $I_{HL}$ et $I_{HH}$ :

$$I'(x,y) = \begin{cases} 1 & \text{si } \frac{1}{3}(I_{HL} + I_{LH} + I_{HH}) > S \\ 0 & \text{sinon} \end{cases} \quad (5)$$

Le résultat est globalisé au niveau des blocs, permettant ainsi de localiser les zones à texture régulière.

### 3.3 Détecteurs liés aux contours

Deux caractéristiques principales du texte incrusté liées aux contours ont été identifiées : la densité et la régularité des pixels de contours. Nous proposons donc d'utiliser ici deux détecteurs différents, un pour chaque caractéristique.
La densité des contours est estimée selon [WOL 02]. Le but ici est d'étiqueter une région de l'image comme zone de texte si celle-ci contient de nombreux pixels de contour. Nous commençons par identifier les pixels de contours en binarisant l'image de gradient obtenue par l'opérateur de Sobel. Un traitement par blocs permet ensuite de mesurer localement la densité des contours (le nombre de pixels de contour) :

$$\delta(B_i) = \sum_{(x,y)\in B_i} E(x,y) \quad (6)$$

où :

$$E(x,y) = \begin{cases} 1 & \text{si } I_{\text{Sobel}}(x,y) > S \\ 0 & \text{sinon} \end{cases} \quad (7)$$

et $I_{\text{Sobel}}$ l'image de gradient de Sobel de l'image originale $I$. Un bloc $B_i$ de densité élevée $\delta(B_i) > S$ est supposé appartenir à une région de texte incrusté.
Comme les zones de texte incrusté sont composées de texte typographié, elles contiennent généralement des segments de lignes de directions prédéfinies (principalement verticales et horizontales). Nous utilisons donc un détecteur rapide de segments de lignes par blocs [LEF 02] particulièrement intéressant pour les directions horizontales et verticales. On obtient alors pour chaque bloc la présence ou non d'un ou plusieurs segments de droite de direction prédéfinie et la position de ce(s) segment(s). Une région caractérisée par un nombre important de segments de ligne est assimilée à une zone de texte incrusté.
Nous avons introduit ici quatre détecteurs basés sur les caractéristiques identifiées des zones de texte incrusté et un critère d'efficacité. Pour plus d'efficacité, ces détecteurs peuvent intégrer la caractéristique d'invariance temporelle du texte incrusté.

### 3.4 Prise en compte de l'invariance temporelle

La caractéristique d'invariance temporelle est intégrée dans chacun des différents détecteurs. Elle peut se traduire par deux principes alternatifs.
Le premier stipule que l'image fournie en entrée à un détecteur ne contient que les zones extraites par ce détecteur sur la trame précédente de la séquence vidéo, avec une réinitialisation à l'image complète effectuée périodiquement. Ce principe permet en outre de limiter le temps de calcul. De plus il peut se formaliser comme suit :

$$I'(t) = \begin{cases} D(I_t) & \text{si } t \mod \Delta = 0 \\ D_{I'_{t-1}}(I_t) & \text{sinon} \end{cases} \quad (8)$$

où $D(I_t)$ représente l'application du détecteur $D$ à l'image $I_t$, $D_{I'_{t-1}}$ dénote la restriction du détecteur $D$ aux zones détectées sur la trame considérée à l'instant précédent $t-1$ (et notées $I'_{t-1}$), et $\Delta$ définit le pas de réinitialisation.
Le second principe considère qu'une zone de texte ne sera conservée que si elle a été extraite par le même détecteur sur un nombre d'images successives donné. La formulation de ce principe est donc la suivante :

$$I''_t = \bigwedge_{k\in[t-\lambda,t]} I'_k \quad (9)$$

où l'image définitive $I''$ à l'instant $t$ est obtenue par conjonction des résultats $I'$ sur une plage de $\lambda$ trames successives.
En utilisant les différents détecteurs avec un principe d'invariance temporelle, on obtient des résultats de segmentation locaux. Nous allons maintenant montrer comment fusionner les différents résultats afin d'obtenir une décision de segmentation globale.

## 4 Fusion des détecteurs

Chacun de nos détecteurs opère sur des blocs de pixels. Dans chaque cas les blocs seront choisis de même taille et ils constituent une partition de l'image. Une fusion est donc nécessaire [BLO 03], et dans notre cas elle s'appliquera au niveau des blocs. Pour formaliser notre stratégie de fusion, nous représentons chaque détecteur par une fonction définie sur une image de taille quelconque et à valeur dans un ensemble binaire :

$$\begin{array}{rccl} D: & \mathcal{I} & \longrightarrow & \{0,1\} \\ & I & \longmapsto & D(I) = \begin{cases} 0 & \text{si } \mathcal{C}(\mathcal{I}) \\ 1 & \text{sinon} \end{cases} \end{array} \quad (10)$$

avec la condition :

$$\mathcal{C}(\mathcal{I}) : \text{le bloc } I \text{ n'est pas qualifié en texte} \quad (11)$$

Pour qualifier un bloc $I$ nous avons défini deux stratégies de combinaison des détecteurs.
La première considère une exécution parallèle de tous les détecteurs. Les résultats sont ensuite fusionnés en affectant des coefficients de pondération aux détecteurs. Ces coefficients peuvent être définis *a priori* à partir d'une étape d'apprentissage ou fixés en ligne. Les régions sont conservées si leur score global (la somme des scores individuels pondérés) est

supérieure à un seuil prédéfini. En utilisant les notations introduites précédemment, et pour $k$ détecteurs, on a :

$$\begin{array}{rccl} D_{\text{final}}: & \mathcal{I} & \longrightarrow & [0,1] \\ & I & \longmapsto & D_{\text{final}}(I) = \left\{ \begin{array}{ll} 1 & \text{si } \sum_{i=1}^{k} p_i D_i(I) > S \\ 0 & \text{sinon} \end{array} \right. \end{array} \quad (12)$$

où les $p_i$ représentent les différents poids associés aux détecteurs $D_i$, et $S$ le seuil global.

La seconde stratégie considère une exécution séquentielle des différents détecteurs. De plus, cette exécution peut être vue comme hierarchique. Les détecteurs sont classés et numérotés en fonction de leur rapidité et de leur tolérance quand ils sont appliqués à la trame complète. Le premier est le plus tolérant. Le dernier est de meilleure qualité mais d'efficacité plus médiocre. La stratégie peut donc se formuler pour un bloc $I$ :

– $D_{\text{final}}$ est défini à partir de cette suite d'opérateurs $D_1, \ldots, D_k$ par $D_{\text{final}} = D'_k$
– $D'_i$ est défini pour $i > 1$ par :

$$\begin{array}{rccl} D'_i: & \mathcal{I} & \longrightarrow & [0,1] \\ & I & \longmapsto & D'_i(I) = \left\{ \begin{array}{ll} D_i(I) & \text{si } D_{i-1}(I) = 1 \\ 0 & \text{sinon} \end{array} \right. \end{array} \quad (13)$$

Les deux stratégies introduites ici ont avantages et inconvénients. Tandis que la première stratégie peut se baser sur une étape d'apprentissage pour déterminer les coefficients optimaux, elle nécessite le traitement de tous les détecteurs sur les trames complètes. Au contraire, la seconde stratégie est plus rapide surtout sur un système mono-processeur mais certaines zones de texte peuvent demeurer non détectées.

# 5 Résultats et discussion

La méthode introduite dans cet article a été testée sur un corpus de séquences vidéo couleur. Les détecteurs décrits dans la section 3 ont été évalués indépendamment en terme d'efficacité et de qualité, respectivement en mesurant le temps de calcul (normalisé) et en estimant manuellement les taux de rappel $T_r$ et de précision $T_p$. L'intérêt des deux stratégies de fusion a aussi été quantifié avec ces mesures. Le tableau 1 résume les mesures moyennes d'efficacité et de qualité pour tous les détecteurs (individuels ou globaux). La valeur de référence $1.000$ utilisée pour comparer les temps moyens est celle obtenue par le détecteur basé sur la couleur, qui traite une image couleur de taille $192 \times 144$ pixels en 60 millisecondes sur une architecture PC à base de processeur Celeron 600 MHz. Nous rappellons que les taux de rappel et de précision peuvent être mesurés de la manière suivante :

$$T_r = \frac{N_c}{N_c + N_m} \quad (14)$$

$$T_p = \frac{N_c}{N_c + N_f} \quad (15)$$

avec $N_c$, $N_m$, et $N_f$ représentant respectivement le nombre de détections correctes, le nombre de détections manquées, et le nombre de fausses détections.

La figure 1 illustre les résultats obtenus avec la stratégie parallèle en considérant tous les détecteurs individuels, tandis que la figure 2 donne la même illustration en considérant la stratégie hiérarchique. Cette dernière, quoique nécessitant un temps de calcul plus réduit, peut engendrer plus facilement des zones non détectées.


FIG. 1 – Résultats obtenus avec la stratégie parallèle (de gauche à droite et de haut en bas) : image originale, détections par la couleur, la texture, la densité des contours, la régularité des contours, et la fusion.

# 6 Conclusion

Dans cet article nous avons introduit une nouvelle méthode pour la détection des zones incrustées dans les séquences vidéo. Contrairement à la plupart des autres approches, nous ne nous basons pas sur un unique détecteur suivi par un post-traitement *ad hoc* et coûteux mais nous considérons plusieurs détecteurs en même temps. Afin de déterminer les détecteurs à utiliser, nous avons tout d'abord mené une étude des zones de texte incrusté dans les séquences vidéo et avons identifié leurs principales caractéristiques, liées à la couleur, la texture, les contours, et l'invariance temporelle. A partir de ces caractéristiques nous avons défini ou sélectionné les détecteurs appropriés. Nous avons ensuite introduit deux stratégies différentes pour fusionner les résultats obtenus par chacun de ces détecteurs en une décision globale, soit d'une manière parallèle, soit d'une manière hiérarchique. Nous avons finalement comparé les détecteurs et les stratégies sur un corpus vidéo, ce qui nous a permis de valider notre contribution.

Parmi les perspectives considérées, nous pouvons mentionner l'utilisation de détecteurs robustes comme les opérateurs morphologiques plats pour détecter des régions de couleur uniforme. Nous souhaitons également adapter notre méthode à des données vidéo compressées afin de traiter les trames vidéo directement dans le domaine compressé.

| Détecteur | Temps moyen normalisé | Taux de rappel $T_r$ et de précision $T_p$ |
|---|---|---|
| Couleur | $1,000$ | $T_r$ élevé, $T_p$ faible |
| Texture | $3,196$ | $T_r$ faible, $T_p$ élevé |
| Densité des contours | $1,981$ | $T_r$ moyen, $T_p$ moyen |
| Régularité des contours | $1,299$ | $T_r$ moyen, $T_p$ moyen |
| Stratégie parallèle | $7,576$ | $T_r$ élevé, $T_p$ moyen |
| Stratégie hiérarchique | $5,837$ | $T_r$ moyen, $T_p$ élevé |

TAB. 1 – Bilan des temps de calcul et des mesures de qualité.


FIG. 2 – Résultats obtenus avec la stratégie hiérarchique (de gauche à droite et de haut en bas) : image originale, détections par la couleur, la régularité des contours, la texture, la densité des contours, et le résultat final.

# Références


[BLO 03] BLOCH I., *Fusion d'informations en traitement du signal et des images*, Hermès, 2003.

[CHA 01] CHAN W., COGHILL G., Text analysis using local energy, *Pattern Recognition*, vol. 34, 2001, pp. 2523–2532.

[DIM 00] DIMITROVA N., AGNIHOTRI L., DORAI C., BOOLE R., MPEG-7 Videotext description scheme for superimposed text in images and video, *Signal Processing : Image Communication*, vol. 16, 2000, pp. 137–155.

[GAN 00] GANDHI T., KASTURI R., ANTANI S., Application of planar motion segmentation for scene text extraction, *IAPR International Conference on Pattern Recognition*, vol. 1, Barcelone, Espagne, Septembre 2000, pp. 445–449.

[HUA 01] HUA X., WENYIN L., ZHANG H., Automatic performance evaluation for video text detection, *International Conference on Document Analysis and Recognition*, Seattle, USA, Septembre 2001, pp. 545–550.

[JAI 98] JAIN A., YU B., Automatic text location in images and video frames, *Pattern Recognition*, vol. 31, n° 12, 1998, pp. 2055–2076.

[JUN 01] JUNG K., Neural network-based text location in color images, *Pattern Recognition Letters*, vol. 22, 2001, pp. 1503–1515.

[KIM 96] KIM H., Efficient automatic text location method and content-based indexing and structuring of video database, *Journal of Visual Communication and Image Representation*, , n° 4, 1996, pp. 336–344.

[LEF 02] LEFÈVRE S., DIXON C., JEUSSE C., VINCENT N., A Local Approach for Fast Line Detection, *IEEE International Conference on Digital Signal Processing*, vol. 2, Santorin, Grèce, Août 2002, pp. 1109–1112.

[LI 00] LI H., DOERMAN D., KIA O., Automatic Text Detection and Tracking in Digital Video, *IEEE Transactions on Image Processing*, vol. 9, n° 1, 2000, pp. 147–156.

[LI 01] LI C., DING X., WU Y., Automatic text location in natural scene images, *International Conference on Document Analysis and Recognition*, Seattle, USA, Septembre 2001, pp. 1069–1074.

[LIE 02] LIENHART R., WERNICKE A., Localizing and Segmenting Text in Images and Videos, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, n° 4, 2002, pp. 256-268.

[WOL 02] WOLF C., JOLION J., Vidéo OCR - Détection et extraction du texte, *Colloque International Francophone sur l'Ecrit et le Document*, Hammamet, Tunisie, Octobre 2002, pp. 215–224.

[WU 99] WU V., MANMATHA R., RISEMAN E., TextFinder : an automatic system to detect and recognize text in images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, n° 11, 1999, pp. 1224–1229.

[ZHO 95] ZHONG Y., KANT K., JAIN A., Locating text in complex color image, *Pattern Recognition*, vol. 28, n° 10, 1995, pp. 1528–1535.

[ZHO 00] ZHONG Y., ZHANG H., JAIN A., Automatic caption localization in compressed video, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, n° 4, 2000, pp. 385–392.