

UNE METHODE D'ACQUISITION ET DE MODELISATION D'IMAGES COULEUR

Pierre COURTELLEMONT, Christian OLIVIER, Philippe SUZONNI

La3i - Université de ROUEN, UFR des Sciences - 76821 MONT-SAINT-AIGNAN CEDEX

Tel.: 35.14.68.74 Fax: 35.14.66.18 e-mail: @courtellemont-la3i-univ.rouen.fr

RÉSUMÉ

Nous proposons une méthode de prétraitement d'images de type document dans les cas où une acquisition classique en niveaux de gris n'est pas suffisante, afin d'en extraire les blocs de texte ou en comprendre la structure physique. A partir d'une acquisition RVB du document, une transformation de Karhunen-Loeve permet par apprentissage d'obtenir les niveaux les plus informatifs pour la segmentation, pour un type de document donné. Le traitement du document s'effectue par un filtrage du fond et une extraction des zones informatives, à l'aide d'une modélisation AR-2D. La segmentation est assurée par des mesures de dissemblance entre les histogrammes des erreurs de modélisation, histogrammes optimaux, au sens d'un critère d'information spécifique, par le nombre et la forme des classes.

INTRODUCTION

Nous proposons une méthode de pré-traitement d'images de type document, qui peuvent présenter des fonds non homogènes (texturés, colorés, ou avec graphiques...). Le but est d'obtenir l'étiquetage en blocs de texte, ou bien d'obtenir la structure physique du document. L'une des difficultés de l'analyse du document est en effet, la localisation de l'information manuscrite. A titre d'exemple, cette information peut être l'adresse postale sur des enveloppes ou paquets, ou le montant manuscrit de chèques bancaires. Même dans ce dernier cas, car les documents peuvent provenir de banques différentes et chacune d'entre elle est susceptible de modifier la présentation de ses chéquiers, il n'est pas possible d'utiliser une quelconque information *a priori* précise sur la localisation du montant.

Dans une première partie, nous montrons que dans de nombreux documents, l'acquisition binaire ou même en niveaux de gris n'est pas suffisante lorsque les différentes informations ne peuvent se différencier que si l'on dispose de toutes les composantes de la couleur. Nous nous intéressons ainsi à la détermination des primitives codant la couleur qui permettent la meilleure segmentation.

La segmentation des niveaux traités ne peut se faire directement quand les informations se superposent et que le fond est variable sur l'image. La seconde partie traite alors de l'analyse par un modèle AR causal quart de plan de l'image. De cette

ABSTRACT

We propose a method of pre-processing of document images when a classical grey-levels acquiring is not sufficient, in the aim to extract text blocks, or to understand the physical structure of the document. From a RVB acquisition of the document, a Karhunen Loeve transform allows, by apprenticeship, to obtain the more informative color features for the segmentation, for a given kind of documents. The processing is performed by a background filtering and an extraction of the informative areas, thanks to an AR-2D modeling. The segmentation is made by the use of dissimilarity measures between the histograms of prediction errors. These histograms are optimal, in the sense of a particular criterion, by the number and the width of their bins.

modélisation, sont extraites les lois d'innovations des erreurs de prédiction sur chaque plan. La recherche des changements dans l'image, par exemple, présence de texte, se pose alors comme un problème de détection de ruptures sur ces lois d'innovation.

Les erreurs de prédiction constituent des histogrammes qui vont permettre la segmentation. Par apprentissage, un critère d'information de type Akaike permet l'obtention d'un histogramme au nombre minimal de classes. La dernière partie expose la méthode adoptée pour classifier différents types de ruptures, par exemple, l'apparition d'une zone de caractères ou un nouveau champ de formulaire, ou encore, un changement de couleur ou de texture: disposant des histogrammes des lois d'innovation, leur dissemblance peut être calculée par des mesures de type Kullback, puis seuillée.

1. ACQUISITION ET CODAGE

De nombreux documents ne peuvent être traités actuellement avec une acquisition en niveaux de gris, quand les informations ne peuvent être séparées sans les autres composantes de la couleur. Introduisons les notions de teinte, intensité et saturation dont les expressions, fonctions des plans RVB (Rouge, Vert, Bleu) seront données plus loin. Prenons l'exemple de la figure 1 où le texte et le fond ont la même intensité I, ce qui se rencontre souvent dans des documents



imprimés en couleur. Dans cet exemple, le texte est séparable aisément du fond en utilisant l'information teinte H. La figure 2 montre une partie d'image en dégradés d'une même teinte. Alors que l'analyse en niveaux de gris ne permet pas d'identifier de discontinuité entre les parties les plus claires et le fond, l'analyse de la saturation S met en évidence ce changement. Ainsi, la figure 3 montre une partie d'un document dans lequel il est facile d'extraire le texte, noir, du fond. Mais l'application, exigeant de trouver la structure physique du document (localisation des cellules de la facture), il est nécessaire d'étudier la composition, toute en dégradé, de ce document.

Nous traitons des documents acquis à l'aide d'un scanner permettant une haute résolution (jusqu'à 2400x800 dpi, sur 30 bits), mais, outre la taille mémoire requise (plus de 500 Moctets pour une image A4), une trop haute résolution peut être néfaste à la phase de segmentation en atténuant les ruptures. Une résolution standard de 300 dpi est donc choisie pour une numérisation donnant de chaque pixel, les 3 composantes RVB sur 8 bits chacune. Cette représentation semble intuitivement moins adaptée à la segmentation de documents - images artificielles créées pour être visibles! - qu'un ensemble de composantes plus proches de la perception visuelle humaine, comme les niveaux HIS (Hue, Intensity, Saturation). Cette hypothèse fut vérifiée à l'aide d'une transformation de Karhunen-Loeve sur un ensemble d'images de différents documents, ainsi que sur quelques images de scènes naturelles pour comparaison. Nous sommes partis des images initiales RVB ainsi que des images en composantes HIS obtenues par les transformations suivantes (s'inspirant de [1], [2]), donnant les résultats les plus proches de ceux obtenus à l'aide de logiciels classiques de traitement d'images:

$$I = \frac{R + V + B}{3}$$

$$S = 255 \left(1 - \frac{\min(R, V, B)}{I} \right) \text{ si } R + B + V \neq 0,$$

$$255 \text{ sinon,}$$

$$H = \frac{255}{2\pi} \arctg \frac{\sqrt{3}(V - B)}{(2R - V - B)} \text{ si } 2R - V - B \neq 0,$$

$$255 \left(\frac{1}{2} + \frac{1}{2\pi} \arctg \frac{\sqrt{3}(V - B)}{2R - V - B} \right) \text{ si } 2R - V - B = 0,$$

$$0 \text{ si } 2R - V - B < 0.$$

Nous appliquons ainsi le principe qu'un paramètre est dit posséder un pouvoir de discrimination important si sa variance est grande. Soit λ_1 , λ_2 et λ_3 les valeurs propres de la matrice de variance-covariance des distributions RVB ou HIS, telles que $\lambda_1 > \lambda_2 > \lambda_3$. Les vecteurs propres W_i respectivement associés aux λ_i permettent d'obtenir de nouvelles représentations, non corrélées, de chacun des pixels de l'image, X_1 , X_2 et X_3 . Si nous rejoignons les conclusions de Ohta ([3]) sur les scènes naturelles, les résultats sur les documents sont complètement différents et confirment en particulier l'importance de la teinte H.

Les résultats obtenus présentent une stabilité sur l'ensemble des images traitées d'un même type de documents, par exemple des enveloppes colorées. Suivant le type de document, une ou plusieurs composantes peuvent être retenues, ou encore, la modélisation proposée au paragraphe 2 peut porter sur une image, combinaison linéaire pondérée par les valeurs propres.



Figure 1: Niveaux I et H de l'exemple 1; puis niveau H segmenté et binarisé.



Figure 2: Niveaux I, H et S du dégradé de l'exemple 2, puis contour détecté sur S.

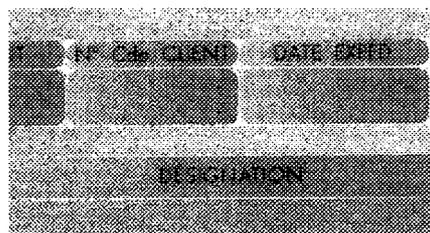


Figure 3: Exemple 3: structure en dégradé d'une facture.

A titre d'exemple, nous obtenons, pour l'image d'une enveloppe dont une partie est visible en figure 4, à partir des niveaux HIS, les valeurs suivantes pour \mathcal{R} , matrice de variance-covariance initiale, \mathcal{R}' , la nouvelle matrice et \mathcal{P} , matrice de passage:

$$\mathcal{R} = \begin{pmatrix} 2895 & -189 & -748 \\ -189 & 211 & -45 \\ -748 & -45 & 609 \end{pmatrix}$$

$$\mathcal{R}' = \begin{pmatrix} 1855 & & \\ & 185 & \\ & & 76 \end{pmatrix}$$

$$\begin{pmatrix} W_1 \\ W_2 \\ W_3 \end{pmatrix} = \begin{pmatrix} 0.957 & 0.239 & 0.163 \\ -0.057 & -0.392 & 0.912 \\ -0.283 & 0.088 & 0.362 \end{pmatrix}^T \begin{pmatrix} H \\ I \\ S \end{pmatrix}$$

Dans ce cas précis, seul W_1 est conservé, avec des résultats de segmentation semblables à ceux obtenus avec les 2 niveaux H et S, comme le laissaient présager leur prédominance dans les matrices ci dessus.

2. MODELISATION

Une analyse directe des niveaux retenus ne permet pas une segmentation de l'image lorsque celle-ci comprend des informations superposées à un fond variable. Nous proposons alors un filtrage du fond par une modélisation autorégressive bidimensionnelle d'ordre $(pq-1)$. Il s'agit d'une modélisation causale-quart de plan ([4]).

Dans ce modèle, la valeur de chaque composante du pixel d'abscisse m et d'ordonnée n s'écrit:

$$y(m,n) = \sum_{i=0}^{p-1} \sum_{j=0}^{q-1} a(i,j) y(m-i,n-j) + e(m,n)$$

avec $i+j \neq 0$,

où: $a(i,j)$ sont les paramètres du modèle, $e(m,n)$ un bruit blanc d'entrée, et $(pq-1)$ l'ordre du modèle.

Cette expression peut être mise sous forme matricielle:

$$y(m,n) = \begin{bmatrix} A_0^T & \dots & A_{p-1}^T \end{bmatrix} \begin{bmatrix} Y_0 \\ \dots \\ Y_{p-1} \end{bmatrix} + e(m,n)$$

avec:

$$A_i^T = [a(i,0), \dots, a(i,q-1)],$$

$$Y_i^T = [y(m-i,n), \dots, y(m-i, n-q+1)] \quad \text{si } i \neq 0,$$

$$A_0^T = [a(0,1), \dots, a(0,q-1)],$$

$$Y_0^T = [y(m,n-1), \dots, y(m, n-q+1)],$$

ainsi:

$$y(m,n) = A^T Y_{m,n} + e(m,n).$$

La condition $i+j \neq 0$ permet d'exclure la valeur $y(m,n)$ qui est la valeur à prédire.

Les équations de récurrence du modèle 1-D sur le gain paramétrique $\sum_m \sum_n Y_{m,n} Y_{m,n}^T$ et sur le vecteur des paramètres A conviennent toujours pour ce modèle. Ils permettent d'obtenir une prédiction $z(m,n)$ du pixel:

$$z(m,n) = \begin{bmatrix} A_0^T & \dots & A_{p-1}^T \end{bmatrix} \begin{bmatrix} Y_0 \\ \dots \\ Y_{p-1} \end{bmatrix}$$

A chaque pixel, l'erreur de prédiction est calculée:

$$\epsilon(m,n) = y(m,n) - z(m,n).$$

Pour ce travail, les ordres optimaux sont étudiés à l'aide du critère BIC ([5]) (voir figure 5, où figure également AIC surparamétrisant le modèle) mais on peut se contenter pour ce type d'images, d'ordres inférieurs pour la détection de ruptures. En particulier l'ordre q sur la verticale est pris égal à 2 seulement pour ne pas noyer les ruptures rencontrées dans un balayage ligne à ligne, donc horizontal, de l'image.

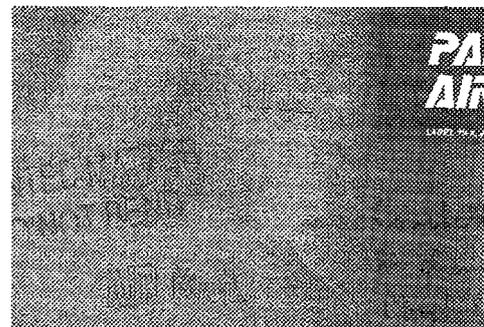


Figure 4: Partie d'une image test d'enveloppe.

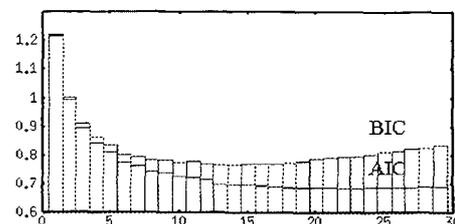


Figure 5: Ordre p du modèle.

Sur le fond de l'image, les erreurs de prédiction sont connues pour leur gaussiennité. Elles augmentent lors des changements de modèles. L'image constituée de ces erreurs forme le support de la partie suivante de notre étude. En effet, grâce à l'hypothèse de gaussiennité du bruit, un seuil adaptatif peut être choisi. Ce seuil α est tel que la probabilité pour que $|e(m,n)|$ dépasse ce seuil soit par exemple de 1%. Cette valeur correspondant à $2,57\sigma$, où σ est l'écart type, est récursivement calculée sur l'image. L'image



constituée des seuls pixels tels que $|\varepsilon(m,n)| > \alpha$ contient les informations des ruptures de l'image, où le fond a été supprimé. Les figures 6a et 6b montrent le comportement de $\varepsilon(m,n)$ sur une ligne du plan H, ainsi que sur le plan I, ce qui confirme les résultats du paragraphe précédent, dans leur importance relative dans W_1 . La figure 7 montre l'image des erreurs binarisées.

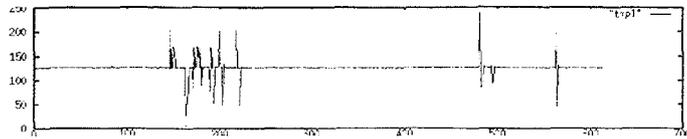


Figure 6a: Comportement des erreurs de prédiction sur la teinte H de l'image de la figure 4 (ligne 252).

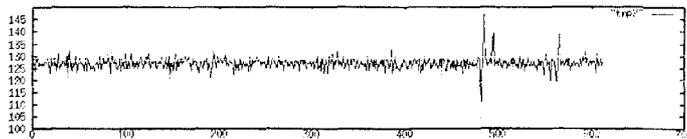


Figure 6b: Comportement des erreurs de prédiction sur l'intensité I de l'image de la figure 4 (ligne 252).

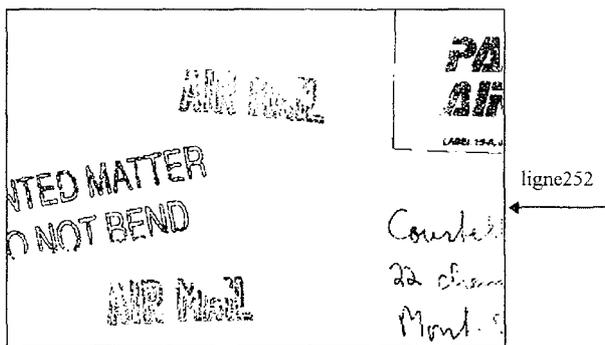


Figure 7: Image des erreurs de prédiction, après binarisation à l'aide du seuil α .

L'image obtenue est à utiliser comme un « masque » sur l'image initiale pour extraire le texte avec une qualité nécessaire au module de reconnaissance.

3. SEGMENTATION

Dans des images plus complexes, la simple détection des ruptures ne suffit pas à la segmentation. Après franchissement du premier seuil, on doit décider de l'appartenance ou non du pixel à un nouveau modèle. Le changement de modèle se caractérise par un changement de la loi suivie par les erreurs de prédiction. Pour comparer ces lois, les erreurs sont réparties en histogrammes construits à l'aide du critère AIC* ([6]):

$$AIC^*(c) = \frac{(c-1)(1+\log N)}{N} - 2 \sum_{B \in C} \theta_c(B) \log \left(\frac{\theta_c(B)}{\mu(B)} \right)$$

où c est le nombre de classes de la sous partition C construite à partir de l'histogramme initial des N échantillons, fusionnés récursivement en classes B jusqu'au minimum de AIC*. La loi θ_c est l'estimateur optimal au sens d'un maximum de vraisemblance et d'un coût de Hellinger, de la vraie loi λ supposée absolument continue à une loi μ donnée a priori.

Nous utilisons alors la divergence de Kullback D , avec respect des précautions d'usage en cas de classes vides ([6]). D s'écrit, pour comparer deux lois θ_c et v :

$$D = \sum_{i=1}^c (\theta_c(B_i) - v(B_i)) \log \frac{\theta_c(B_i)}{v(B_i)}$$

permettant de mesurer ainsi la dissemblance entre les lois suivies par les erreurs, sur le fond et après rupture.

CONCLUSION

Nous avons montré que l'analyse de documents au très faible contraste de niveaux de gris, rejetés habituellement dans les applications de lecture automatique, peut s'effectuer par une modélisation de composantes convenables de la couleur. La méthode de segmentation testée, tout en s'affranchissant des hypothèses de gaussianité, nécessaires par exemple dans [7] en segmentation de textures, permet une classification des différents modèles d'une image. Cette méthode permet de distinguer des ruptures correspondant à des changements dans des lois de moyenne et d'énergie identiques.

REFERENCES

- [1]: Kender J.R., « Saturation, hue and normalized colors », *technical report*, Depart¹ of Computer Science, CMU, 1976.
- [2]: Chapron M., Cocquerez J.P., « La segmentation d'images couleur à partir d'un nouveau détecteur de contours », *13^{ème} colloque GRETSI*, Juan-les-Pins, 16-20 sept. 1991, pp 1089-1092.
- [3]: Ohta Y., Kanade T., Sakai T., « Color information for region segmentation. », *Computer Graphics Image Process.*, 13, pp 224-241, 1980.
- [4]: Olivier C., Avila M., Courtellemont P., & al, « Handwritten Word Recognition by Image Segmentation and Hidden Markov Models », *Proc. of IEEE-IECON'93*, Maui, Hawaii, Nov 11-15, pp 2093-2097, 1993.
- [5]: Schwartz G., « Estimating the dimension of a model », *the Annals of Stat.*, vol.6, n°2, pp 461-464, 1978.
- [6]: Olivier C., Courtellemont & al, « Comparison of histograms: a tool for detection », *European Journal of Diagnosis and Safety in Automation*, vol. 4, 3, pp 335-355, 1994.
- [7]: Oe S., « Texture segmentation method by using two-dimensional AR model and Kullback Information », *Pattern Recognition*, vol. 26, 2, pp 237-244, 1993.