Normalisation et validation d'images de documents capturées en mobilité

Marçal Rusiñol[†*] — **Joseph Chazalon**[†] — **Jean-Marc Ogier**[†]

[†] *Laboratoire L3i, Université de La Rochelle*
Avenue Michel Crépeau, 17042 La Rochelle Cédex 1, France
[*] *Computer Vision Center, Dept. Ciències de la Computació*
Edifici O, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain

RÉSUMÉ. *La numérisation de documents à l'aide des smartphones introduit un nombre important de dégradations qui doivent être corrigées ou détectées sur le mobile, avant l'envoi de données sur un réseau payant ou la perte de disponibilité du document. Dans cet article, nous proposons un système permettant de corriger les problèmes de perspective et d'illumination avant d'estimer la netteté de l'image pour un traitement OCR. L'étape corrective repose sur une détection des contours, suivie d'une normalisation de l'illumination. Son évaluation sur un jeu de données privé montre une amélioration franche des résultats OCR. L'étape de contrôle repose sur une combinaison de mesures de focus. Son évaluation sur un jeu de données public montre que cette approche simple donne des performances comparables à celles des meilleures méthodes basées sur des traitements lourds, et surpasse les méthodes basées sur des métriques.*

ABSTRACT. *Mobile document image acquisition integrates many distortions which must be corrected or detected on the device, before the document becomes unavailable or paying data transmission fees. In this paper, we propose a system to correct perspective and illumination issues, and estimate the sharpness of the image for OCR recognition. The correction step relies on fast and accurate border detection followed by illumination normalization. Its evaluation on a private dataset shows a clear improvement on OCR accuracy. The quality assessment step relies on a combination of focus measures. Its evaluation on a public dataset shows that this simple method compares well to state of the art, learning-based methods which cannot be embedded on a mobile, and outperforms metric-based methods.*

MOTS-CLÉS : *acquisition d'image de document avec mobile ; correction de perspective ; correction d'illumination ; estimation de qualité ; mesure de focus ; prédiction de la fiabilité OCR*

KEYWORDS: *mobile document image acquisition; perspective correction; illumination correction; quality assessment; focus measure; OCR accuracy prediction*

1. Introduction

L'acquisition d'images de documents avec des périphériques mobiles, les *smartphones* en particulier, est devenu un point d'entrée essentiel dans les chaînes de traitement documentaires industrielles. Malgré l'attrait évident de la numérisation à la volée et le transfert quasi instantané des images de documents, trois défis majeurs restent à surmonter pour libérer le potentiel de la capture mobile d'images de documents.

Les distorsions de numérisation sont, dans le cas d'images de documents capturées avec un dispositif mobile, la principale cause de perturbation du signal de l'image. Parmi les distorsions les plus courantes, le flou provoqué par un mauvais focus est particulièrement délicat. Il modifie sérieusement la lisibilité, à la fois pour les humains et les systèmes OCR, et l'utilisateur a peu de contrôle sur lui car il est lié au comportement interne de l'appareil, contrairement au flou de mouvement, à la distorsion de perspective ou aux conditions d'éclairage. Pour éviter de mauvaises performances dans le traitement des documents, la lisibilité d'une image de document capturée avec un mobile doit donc être optimisée et contrôlée le plus tôt possible dans le processus.

Les frais de communication des réseaux mobiles imposent une sélection rigoureuse sur les images à transférer. Par conséquent, le contrôle de la lisibilité doit être effectué sur le dispositif mobile, avant l'envoi de données.

La mobilité change la façon dont les utilisateurs capturent les documents : l'opportunité de numériser un document peut être temporaire. Tout contrôle sur les images capturées doit donc être effectué pendant ou juste après la capture, pour éviter de manquer l'occasion d'une autre capture.

Il est alors nécessaire de viser un contrôle aussi rapide que possible, en temps réel idéalement, pour débloquer progressivement différents niveaux d'assistance à la capture tels que : 1) la notification à l'utilisateur d'une mauvaise capture, ainsi que l'explication de cette évaluation ; 2) le déclenchement automatique de la capture lorsque les conditions de capture sont optimales ; 3) le guidage de l'utilisateur lors de la capture à l'aide d'indications précises telles que *« rapprochez l'appareil du document »*, ou *« l'éclairage est trop faible »*, comme suggéré dans (Chen *et al.*, 2013a).

Dans cet article, nous présentons un prototype de système d'acquisition d'images de documents destinés à un traitement OCR. Ce système permet la productions d'images où les distorsions liées à la perspective, à l'éclairage et au flou sont maîtrisées. Une première étape corrective, présentée en section 2, permet de corriger la perspective de l'image et son illumination afin d'optimiser de la qualité de l'image avant l'évaluation de cette dernière. Les résultats en terme de performance OCR montrent le gain apporté par cette étape. Une seconde étape de contrôle, présentée en section 3, permet prédire la fiabilité attendue d'un système OCR à l'aide d'une analyse du focus de l'image et donc, indirectement, des conditions de capture (on fait l'hypothèse que la qualité intrinsèque du document est suffisante). La technique proposée dans cette étape se base sur évaluation de mesures de focus jamais testées (à notre connaissance) sur des images de documents. Elle présente des performances comparables aux

meilleures techniques basées sur des algorithmes lourds, tout en étant rapide et simple à calculer.

2. Correction de la perspective et de l'illumination

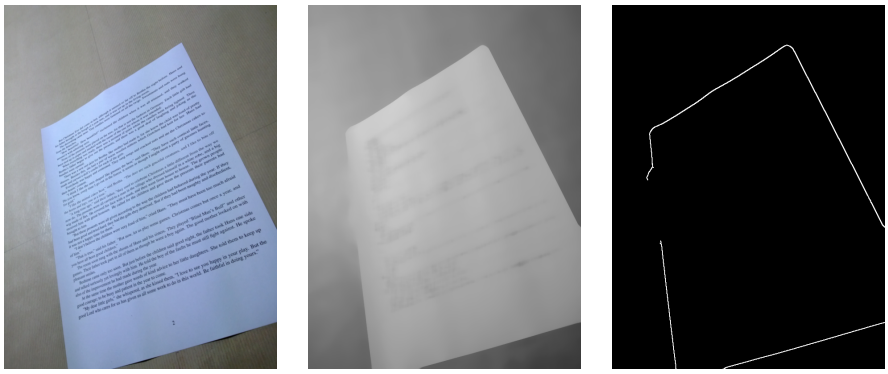
Les images de documents capturées avec un dispositif mobile contiennent, en plus d'une région représentant la page, une portion du fonds. De plus, si le dispositif de capture n'est pas orthogonal au plan de la page, une déformation liée à la perspective sera visible. Ces deux perturbations peuvent dégrader très sérieusement la performance d'un système OCR. Par ailleurs, l'éclairage de la scène peut être affecté par l'utilisation d'un illuminant « non standard » qui provoquera, après une capture avec un mobile, des teintes bleuâtres ou verdâtres à la place de la teinte blanche attendue du papier. Ce changement dans l'espace de couleur peut, lui aussi, avoir un effet négatif sur la performance des systèmes OCR conçus pour traiter des images numérisées à l'aide de scanners avec lesquels ces artéfacts sont négligeables. Cette section présente les deux solutions proposées afin de faire face à ces défauts de perspective et d'illumination.

2.1. *Correction de la perspective*

Depuis quelques années, on peut trouver dans la littérature un bon nombre de publications qui traitent de la segmentation des pages acquises avec des dispositifs mobiles et de la correction des effets de perspective, comme par exemple (Clark et Mirmehdi, 2001 ; Lu *et al.*, 2005 ; Jagannathan et Jawahar, 2005 ; Rodríguez-Piñeiro *et al.*, 2011). La plupart de ces travaux visent soit à détecter les points de fuite horizontaux et verticaux en se basant sur les lignes horizontales du texte et les éléments verticaux des glyphes du texte, soit à détecter les quatre coins de la page pour essayer ensuite de transformer le quadrilatère trouvé en un rectangle parfait. Puisque la numérisation de pages A4 complètes est un scénario courant, nous avons décidé d'attaquer le problème en essayant d'extraire la page (supposée complète) en détectant ses bords et ses quatre coins.

Pour détecter les bords de la page, on utilise un détecteur de contours classique comme la méthode de Canny. Au préalable, on traite l'image avec un filtre médian, avec un élément structurant assez large afin d'éliminer le texte imprimé et ne conserver que le fond. La figure 1 illustre ce processus.

Une fois la détection des contours de la page réalisée, on procède à la détection de ses quatre coins. Au lieu d'utiliser directement un détecteur de coins comme celui de Harris, nous cherchons d'abord à identifier les bords de la page à l'aide la transformée de Hough (Ballard, 1981). Ceci a l'avantage de permettre une détection des coins même lorsque ceux-ci ne sont pas visibles dans l'image, dès lors qu'une partie de chacun des quatre bords est présente dans l'image.

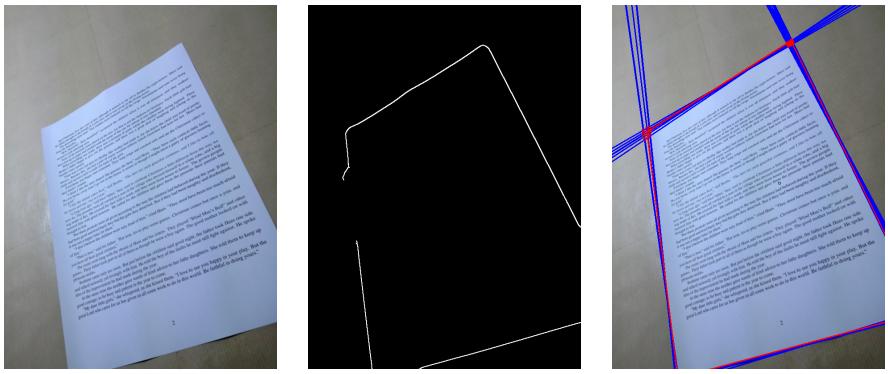


a) b) c)

Figure 1. *Exemple d'extraction des contours. a) Image originale, b) image traitée avec le filtre médian, c) image des contours détectés par la méthode de Canny.*

Lors de la sélection des paramètres ρ et θ de la transformée de Hough, il faut tenir compte de l'influence du facteur de quantification dans l'espace de vote : une quantification grossière aura tendance à provoquer davantage d'erreurs dans la détection des lignes, tandis qu'en diminuant la taille des *bins* on fait progressivement apparaître plusieurs lignes candidates pour une même zone de l'image. Dans notre cas, on favorise la sur-segmentation des lignes. Une fois ces lignes détectées, on les analyse deux à deux pour calculer l'angle qu'elles forment entre elles. Si l'angle entre deux lignes est extrêmement aigu ou grave, on écarte cette paire qui correspond à des lignes presque parallèles qui se croisent. Dans tous les autres cas, on génère l'ensemble des points d'intersection pour former les candidats à la représentation de chacun des quatre coins de la page. Le résultat de ce processus d'extraction de lignes est illustré par la figure 2.

Une fois l'ensemble des points candidats identifiés, il faut commencer par les affecter à un quadrant de la page, afin de déterminer s'il s'agit d'un candidat pour le coin haut gauche, haut droit, etc. Il suffit pour cela de calculer les coordonnées du centre de gravité du nuage de points candidats pour ensuite changer de repère et identifier le cadran de chaque point. Pour chacun de ces points, suivant leur cadran, on construit des couples associant leurs coordonnées dans l'image d'origine et celles dans l'image cible : par exemple, si on cherche à numériser des pages A4 à 300 dpi, tous les points correspondant au quadrant haut gauche auraient comme point modèle le `(0,0)` ; ceux du quadrant haut droit auraient le point `(2480,0)` ; et ainsi de suite. Finalement, on applique la méthode de RANSAC (Fischler et Bolles, 1981) aux couples ainsi formés afin de trouver la transformation à appliquer à l'image acquise pour rectifier l'effet de perspective, tout en éliminant les aberrations liées aux points détectés par erreur et ne correspondant à aucun coin. La figure 3 illustre quelques exemples de corrections de perspective à laquelle notre méthode aboutit.



a) b) c)

Figure 2. *Détection des bords et de coins de la page avec la transformée de Hough. a) Image originale, b) image des contours, c) lignes et croisements détectés.*

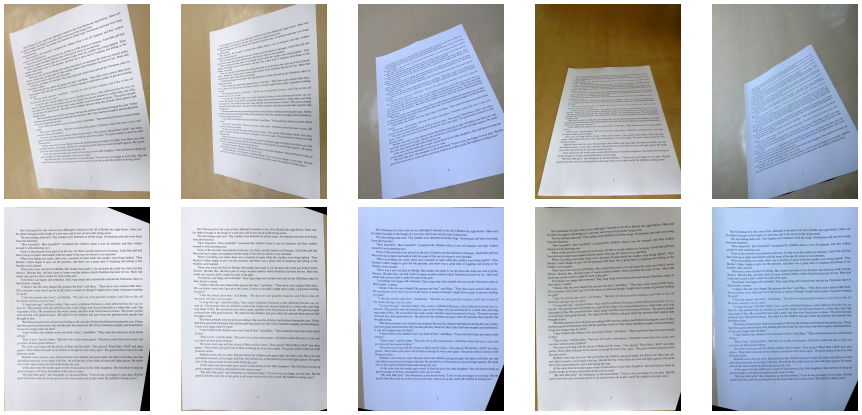


Figure 3. *Exemples d'images acquises avec un dispositif mobile avec des effets de perspective en haut, et les corrections correspondantes en bas.*

2.2. *Correction de l'illumination*

Afin de corriger l'illumination de l'image, on adapte la méthode classique dite du *monde gris (gray world assumption)* au domaine de la numérisation de documents. Notre hypothèse de base est que le document à numériser consiste en une page idéalement blanche sur laquelle est imprimé un texte noir.

Dans l'algorithme original du monde gris, la couleur grise moyenne est calculée en prenant la moyenne de la valeur moyenne de chaque composante couleur. La valeur de chaque composante de l'image est ensuite multipliée par la valeur de couleur grise

	Image scannée	Image acquise avec mobile	Correction perspective	Correction illumination
Taux OCR	96.2%	66.29%	94.62%	95.23%

Tableau 1. *Taux de précision de l'OCR en fonction du traitement.*

moyenne divisée par l'intensité de cette composante de couleur. Ceci empêche une composante de contribuer beaucoup plus à l'image qu'une autre.

Dans notre cas, au lieu d'appliquer une translation de chaque composante couleur vers le gris moyen, on propose de réaliser cette translation vers une intensité ℓ donnée. On calculera donc le nouveau triplet (R', G', B') à partir de

$$R' = \frac{R \times \ell}{\bar{R}}, \qquad G' = \frac{G \times \ell}{\bar{G}}, \qquad B' = \frac{B \times \ell}{\bar{B}}.$$

Dans un second temps, on procède à un étirement de l'histogramme de luminance de l'image pour garantir que les valeurs les plus claires de l'image sont bien représentées par un blanc pur, et que les valeurs les plus sombres correspondent bien à un noir parfait. Il faut commencer par convertir l'image depuis l'espace RGB vers un espace possédant une composante de luminance : Lab, LCH, LSV, etc. Ensuite, on applique une méthode d'augmentation de contraste seulement dans le canal de luminance avec deux seuils de coupure pour les valeurs basses et hautes, respectivement l et h. L'image est finalement retransformée à l'espace RGB. En portant les paramètres l et h à la limite, on obtient des images qui semblent binaires, bien qu'elles soient représentées dans un espace RGB. La figure 4 donne quelques exemples de résultats de cette méthode.

2.3. *Résultats : effets des corrections sur la performance de l'OCR*

Pour évaluer l'effet des corrections proposées sur la performance d'outils OCR, nous avons réalisé 35 captures d'une même page, à l'aide d'un *smartphone*, avec plusieurs degrés de perspective et sous différentes illuminations. Pour disposer d'une référence, nous avons également scanné cette page, et nous disposons de la vérité terrain pour le texte. Pour cette expérience, nous avons utilisé l'OCR commercial ABBYY FineReader.

On peut apprécier, dans le tableau 1, l'impact des corrections proposées sur le comportement de l'OCR. FineReader montre en effet de faible performance pour les images acquises avec le *smartphone* qui présentent des déformations de perspective. Dans la plupart des cas, les erreurs sont dues à un échec de l'étape de segmentation de la page ; les zones de texte étant considérées comme des images. Cependant, après correction de la perspective, on constate un fort gain en précision. Finalement, la cor-

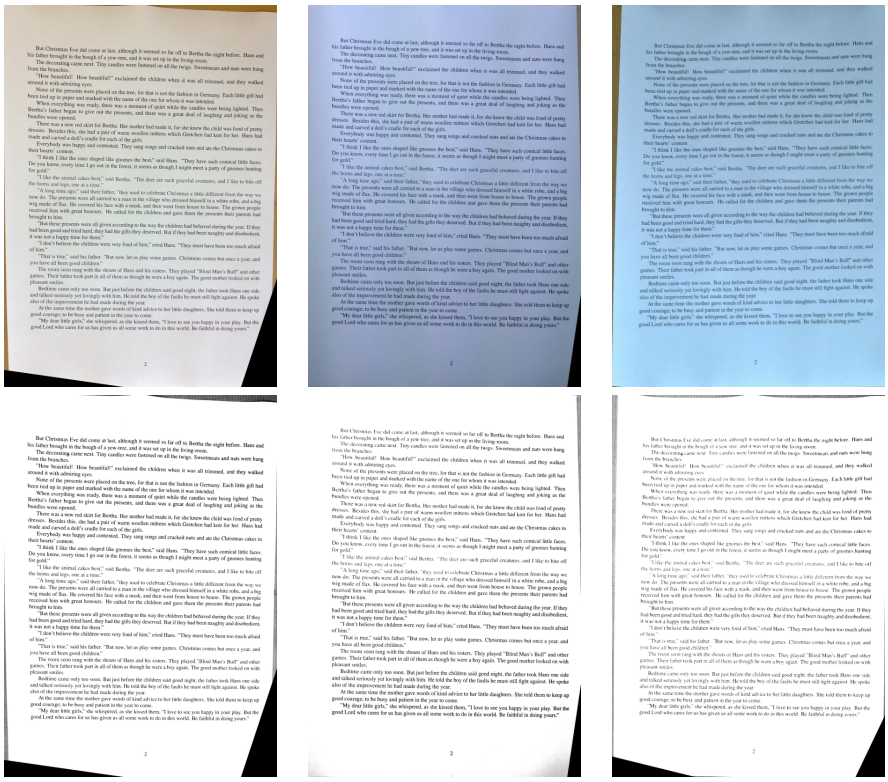


Figure 4. *Correction de l'illumination des images. Pour chaque image, la vignette du haut est la version après correction de perspective, et celle du bas la version après correction de l'illumination.*

rection de l'illumination aide également FineReader et permet d'obtenir des résultats très proches de ceux obtenus à partir de la version scannée du document.

3. Estimation de la qualité

Dans cette section, nous décrivons la méthode d'estimation de qualité (dans un objectif de traitement OCR) que nous proposons d'appliquer aux images de documents capturées à l'aide d'un téléphone mobile, ainsi que les évaluations que nous avons réalisées. Cette méthode garantie que l'image produite dispose d'une netteté suffisante, et limite donc le risque de flou de focus délicat à maîtriser par l'utilisateur.

Cette estimation de la qualité a lieu, dans le scénario que nous considérons, après le pré-traitement de l'image pour ne conserver que la partie concernant la page de document, où la perspective et l'illumination ont été corrigées. Cette étape fait alors

office de barrière avant l'envoi (souvent via un réseau payant) des images vers un service centralisé. Notons que si, dans sa version actuelle, notre méthode d'estimation de la qualité intervient après la capture, il est possible d'imaginer un fonctionnement plus interactif où des conseils pour l'amélioration de la prise de vue pourraient être fournis une fois l'image capturée, voire même pendant la capture comme proposé par (Chen *et al.*, 2013b). Si le temps de traitement est suffisamment court, ce qui est ici une condition indispensable, il pourrait même être possible, à terme, de déclencher automatiquement la capture lorsqu'un optimum est détecté dans la configuration de prise de vue.

Il s'agit donc d'une estimation de qualité s'intéressant spécifiquement aux conditions de capture, et non à la qualité intrinsèque d'un document pour laquelle d'autres méthodes existent depuis un certain temps à présent (Gonzalez *et al.*, 1998). L'estimation de la qualité d'images de documents peut-être entendu de différentes façons, selon qu'on dispose ou non d'une référence pour comparaison. (Ye et Doermann, 2013) proposent ainsi de distinguer trois formes.

Estimation de qualité avec référence complète Cette méthode est possible lorsque l'image de référence est disponible, par exemple lors du contrôle du résultat d'une compression.

Estimation de qualité avec référence partielle Cette méthode repose sur la présence de marqueurs ou de mires de calibration, par exemple, dans l'image capturée qui permet d'estimer précisément les paramètres de la capture.

Estimation de qualité sans référence Ce cas plus général est celui où la qualité de l'image doit être estimée « à l'aveugle », en se basant sur une connaissance à priori des propriétés attendues de l'image.

Dans notre cas, la capture d'un document inconnu impose une estimation de qualité sans référence, afin de prédire le niveau de fiabilité d'un système OCR.

3.1. *État de l'art*

Comme montré par (Ye et Doermann, 2013), peu de travaux ont été menés sur la prédiction du niveau de fiabilité des systèmes OCR pour des images de documents capturées avec appareil photo, caméras ou encore *smartphones*. La gestion spécifique du flou du focus n'a fait l'objet que de quelques études à notre connaissance.

Une étude récente menée par (Kumar *et al.*, 2013) compare les résultats de différentes méthodes à l'aide d'un jeu d'images de documents spécifique. Le jeu de test utilisé par ces auteurs comporte, pour 25 documents différents, une série de captures faisant varier la distance focale de l'appareil, et ainsi le flou de focus. Cette base est accessible publiquement, et nous l'avons utilisée pour l'évaluation de notre méthode.

Parmi les trois méthodes comparées par les précédents auteurs, la méthode CORNIA présentée par (Ye et Doermann, 2012 ; Ye *et al.*, 2012), présente les meilleurs résultats. Elle est cependant basée sur une représentation éparse et une sélection automatique des caractéristiques, imposant des apprentissages lourds et des traitements

coûteux difficilement intégrables dans des *smartphones*. Les méthodes Q (proposée par (Zhu et Milanfar, 2010) et basée sur une décomposition en valeurs singulières) et ΔDoM (proposée par (Kumar *et al.*, 2012) et basée sur une métrique plus simple à base de gradients) présentent quant à elles des résultats plus modestes. Il ressort de cette étude que les méthodes adaptées au traitements de documents inconnus sont actuellement soit trop lourde pour être intégrées dans les *smartphones*, soit trop peu performantes.

Une autre méthode, ne figurant pas dans l'évaluation précédente, a été proposée par (Peng *et al.*, 2011). Bien que l'information produite par cette approche, destinée à faciliter la prise de décision relative à l'image courante, soit particulièrement intéressante, la méthode proposée se base sur des caractéristiques définies manuellement considérant la forme du contour des caractères, ainsi que des rapports hauteur-largeur pour les mots et caractères. Ce type de caractéristiques limite, selon nous, le champs d'application de cette méthode à une classe de documents textuels bien connus, et ne semble pas adaptée au traitement de documents inconnus.

Il est donc nécessaire de rechercher d'autres méthodes basées sur des métriques simples et rapides à calculer permettant d'estimer efficacement la netteté d'une image de document inconnu sans être spécifique à une classe de contenu particulière. Des travaux dans cette direction existent dans d'autres communautés scientifiques, en particulier pour la mise au point d'algorithmes d'auto-focus (travaillant en temps réel). (Pertuz *et al.*, 2013) proposent ainsi une comparaison de 36 mesures de focus classées en 6 familles présentant un comportement homogène selon le type de perturbation considéré. Le comportement de ces mesures sur des images de documents n'a, à notre connaissance, jamais été étudié.

Par ailleurs, ces mesures étant généralement très rapides à calculer, nous avons souhaité étudier la possibilité de les combiner pour aboutir à une méthode globale restant rapide à calculer, mais permettant de compenser les défauts des méthodes isolées, et ainsi mieux modéliser l'évolution de la fiabilité d'un système OCR.

3.2. *Mesures de focus étudiées*

Dans (Pertuz *et al.*, 2013), les auteurs distinguent 6 familles de mesures de focus aux comportements homogènes.

Famille des méthodes « à base de gradients » Abrégée *GRA* pour *« gradient-based »*, cette famille estime la netteté des contours du contenu de l'image grâce à une analyse des gradients de cette dernière.

Famille des méthodes « à base de laplacien » Abrégée *LAP* pour *« Laplacian-based »*, cette famille vise le même objectif que la précédente, mais en utilisant la dérivée seconde ou le laplacien de l'image.

Famille des méthodes « à base d'ondelettes » Abrégée *WAV* pour *« wavelet-based »*, cette famille construit des indicateurs grâce à une transformée en ondelettes de l'image.

Mesure	Abrév.	Mesure	Abrév.
Gaussian derivative	GRA1	Ratio of wavelet coefficients	WAV3
Gradient energy	GRA2	Gray-level variance	STA3
Thresholded absolute gradient	GRA3	Gray-level local variance	STA4
Squared gradient	GRA4	Normalized gray-level variance	STA5
Tenengrad	GRA6	Histogram entropy	STA7
Tenengrad variance	GRA7	Histogram range	STA8
Energy of Laplacian	LAP1	Brenner's measure	MIS2
Modified Laplacian	LAP2	Image curvature	MIS4
Diagonal Laplacian	LAP3	Helmli and Scherer's mean	MIS5
Variance of Laplacian	LAP4	Steerable filters-based	MIS7
Sum of wavelet coefficients	WAV1	Spatial frequency measure	MIS8
Variance of wavelet coefficients	WAV2	Vollath's autocorrelation	MIS9

Tableau 2. *Noms et abréviations des 24 mesures de focus extraites de (Pertuz* et al.*, 2013) pour lesquelles nous avons recherché la combinaison optimale.*

Famille des méthodes « à base de transformée en cosinus discrète » Abrégée *DCT* pour *« discrete cosine transform-based »*, cette famille, à l'instar de la précédente, utilise l'espace de représentation induit par la transformée en cosinus discrète pour extraire des indicateurs.

Famille des méthodes « à base d'indicateurs statistiques » Abrégée *STA* pour *« statistic-based »*, cette famille utilise des analyses statistiques (moments, variance, etc.) pour extraire des indicateurs.

Famille des méthodes « diverses » Abrégée *MISC* pour *« miscellaneous-based »*, cette famille regroupe les méthodes qui ne rentrent dans aucune des catégories précédentes.

Parmi les 36 mesures proposées par ces auteurs, nous en avons sélectionné 24, indiquées dans le tableau 2, pour lesquelles le temps de calcul était raisonnable au regard de notre scénario. Cette sélection a notamment provoqué l'élimination des méthodes à base de transformée en cosinus discrète, trop coûteuses.

3.3. *Combinaison des mesures*

Parmi les mesures présentées précédemment, il est difficile de savoir à priori quelles sont celles qui présenteront la meilleure corrélation avec la mesure de la fiabilité de l'OCR, tout comme il est délicat de choisir comment combiner ces mesures pour palier leurs défauts respectifs. Pour résoudre ce problème, nous avons testé toutes les combinaisons possibles de ces 24 mesures.

Soit $n = 24$ et $S = \{FM_1, FM_2, \ldots, FM_n\}$ l'ensemble de toutes les mesures de focus FM appliquées à une image de test donnée. Nous avons évalué toutes les

combinaisons T possibles de cet ensemble. On note T_m, où $m \in [1,\ n]$, tous les sous-ensembles possibles de mesures formés en choisissant m éléments de S. En particulier, T_m^j (avec $j \in \left[1,\ \binom{n}{m}\right]$), dénote le j-ième sous-ensemble possédant exactement m éléments. Ces combinaisons représentent plus de 16 millions de configurations différentes.

Une fois un sous-ensemble de mesures de focus à combiner choisi, un double problème se pose alors : 1) comment normaliser les mesures pour qu'elles produisent des valeurs dans un domaine cohérent ; et 2) comment fusionner ces valeurs pour produire un résultat unique.

3.3.1. *Méthodes de normalisation*

Comme noté par (Kittler *et al.*, 1998), il est nécessaire de normaliser les mesures pour que les domaines de valeurs de ces dernières soient comparables. Nous avons utilisé des techniques usuelles, présentées ci-après, dans nos expérimentations. Pour un ensemble de mesures T_m^j, nous notons ${T_m^j}'$ leur variante normalisée.

Min-max Cette technique applique un facture de mise à l'échelle et transforme la mesure pour qu'elle renvoie des valeurs dans le domaine $[0, 1]$. Soient $min(T_m^j)$ et $max(T_m^j)$ les minimum et maximum, respectivement, des scores.

$$ {T_m^j}' = \frac{T_m^j - min(T_m^j)}{max(T_m^j) - min(T_m^j)} $$

Cette méthode est très sensible aux points aberrants dans les données.

Z-score Cette technique est une des plus couramment utilisées. Elle utilise la moyenne arithmétique μ et l'écart-type σ des données considérées.

$$ {T_m^j}' = \frac{T_m^j - \mu}{\sigma} $$

L'utilisation de la moyenne arithmétique et de l'écart-type rend également cette méthode sensible aux points aberrants.

Tanh Cette technique robuste et efficace est elle aussi basée sur la moyenne arithmétique et l'écart-type.

$$ {T_m^j}' = \frac{1}{2}\left\{\tanh\left[0.01 \cdot \left(\frac{T_m^j - \mu}{\sigma}\right)\right] + 1\right\} $$

MAD Les écarts médians et médians absolus utilisés par cette méthode sont insensibles aux points aberrants et aux extremums de la distribution.

$$ {T_m^j}' = \frac{T_m^j - median(T_m^j)}{median(|T_m^j - median(T_m^j)|)} $$

3.3.2. *Stratégies de fusion*

Après la normalisation des valeurs renvoyées par chacune des mesures de focus, plusieurs stratégies de fusion des mesures normalisée contenues dans un sous-ensemble T_m^j : il est possible de choisir le minimum, le maximum, la somme, le produit, la moyenne ou encore la médiane de ces valeurs. Plus formellement, on définit les stratégies de fusion suivantes :

$$combMAX = max\left(T_m^{j\,\prime}\right), \qquad combMIN = min\left(T_m^{j\,\prime}\right),$$

$$combSUM = \sum_{j=1}^{\binom{n}{m}}\left(T_m^{j\,\prime}\right), \qquad combPROD = \prod_{j=1}^{\binom{n}{m}}\left(T_m^{j\,\prime}\right),$$

$$combAVG = mean\left(T_m^{j\,\prime}\right), \qquad combMED = median\left(T_m^{j\,\prime}\right).$$

Notre choix d'utiliser, dans un premier temps, ces techniques de fusion au lieu de techniques de régression se justifie par la nécessité, dans l'application finale, de privilégier la détection et le refus de mauvaises captures à une prédiction exacte de la qualité des résultats OCR. Une combinaison linéaire aurait pour effet de rendre certaines mesures plus importantes que les autres, et notre approche vise au contraire à permettre de tirer profit des « doutes » de chacune des mesures.

Les résultats expérimentaux présentés ci-après indiquent les performances obtenues pour chaque méthode de normalisation et chaque stratégie de fusion.

3.4. *Validation expérimentale*

Nous avons utilisé la base publique proposée par (Kumar *et al.*, 2013) pour évaluer notre approche, et la comparer à celles évaluées dans cet article. Cette base comporte 175 images qui représentent les différentes captures (entre 6 et 8) réalisées avec différentes distances focales pour chacun des 25 documents. Pour chacun des documents, certaines captures sont parfaitement nettes, et d'autres présentent des niveaux de flou variables. La vérité terrain ainsi que les réponses de trois systèmes OCR (FineReader, Omnipage et Tesseract) pour chacune des images est fournie. La fiabilité de l'OCR pour chacune des images est également disponible. La figure 5 illustre le contenu de ces documents.

Dans la mesure où la base publique de (Kumar *et al.*, 2013) est la seule à proposer une variation du focus dans des images de documents, mais qu'elle ne contient pas de déformation de perspective, nous avons évalué séparément l'étape d'estimation de la qualité des images basée sur une mesure de focus et celle de normalisation de la perspective et de l'illumination présentées en section 2. Pour simuler une étape de correction de perspective idéale, nous nous sommes contentés de détecter automatiquement la zone de l'image contenant le document avec une version simplifiée de la

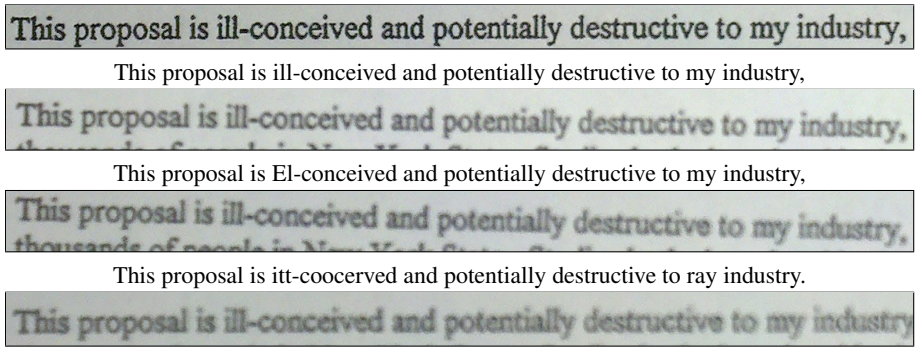


Figure 5. *Extraits d'un même document issu de la base de test de (Kumar* et al.*, 2013) montrant différents niveaux de flou, ainsi que les réponses d'un des systèmes OCR.*

Fusion	Normalisation			
	Min-max	Z-score	Tanh	MAD
combMAX	0.87439	0.86364	0.86363	0.86679
combMIN	0.89791	**0.9378**	0.93779	0.93092
combSUM	0.92056	0.92164	0.92165	0.91945
combPROD	0.87073	0.91372	0.92175	0.86989
combAVG	0.92056	0.92164	0.92165	0.91945
combMED	0.92056	0.92164	0.92165	0.91945

Tableau 3. *Meilleures médianes des coefficients de corrélation de Pearson pour les différentes techniques de normalisation et de fusion.*

méthode présentée à la section précédente ; les images de la base ne nécessitant pas de technique complexe.

Nous avons utilisé la méthode d'évaluation de l'adéquation entre les mesures et la fiabilité de l'OCR proposée par (Kumar *et al.*, 2013), à savoir le calcul de la corrélation de Pearson *(LLC)* (resp. de Spearman *(SROCC)*) séparément pour chaque document, puis en prenant la médiane de cette valeur sur l'ensemble des documents.

Le tableau 3 indique les meilleurs résultats obtenus (pour la corrélation de Pearson) parmi tous les sous-ensembles des 24 mesures, en fonction des techniques de normalisation et de fusion utilisées. Dans la plupart des cas, ces résultats optimaux ont été obtenus en ne combinant qu'entre deux et quatre mesures parmi les 24 considérées. Concernant les techniques de fusion, on peut noter que la technique la plus pessimiste *combMIN* est celle qui permet de mieux représenter le comportement de l'OCR.

Méthode	Principe	LCC médian	SROCC médian
Q	Métrique	0.8271	0.9370
ΔDOM	Métrique	0.8488	0.9524
CORNIA	Apprentissage	0.9747	0.9286
Proposée	Métrique	0.9378	0.96429

Tableau 4. *Comparaison avec l'état de l'art (Kumar* et al.*, 2013).*

Méthode	FineReader	Omnipage	Tesseract
CombMIN+Z-score	0.9378	0.8794	0.9197

Tableau 5. *Médiane des coefficients de la corrélation de Pearson pour les différents systèmes OCR.*

D'une manière générale, les mesures ayant donné les meilleurs résultats sont celles à base de gradient. La meilleure configuration était, quant à elle, formé du sous-ensemble de 4 mesures $T_4^j = \{GRA1, GRA2, GRA4, STA8\}$. Le temps moyen pour le calcul et la normalisation de ces quatre mesures 0,61 secondes sur une machine récente avec un code Matlab séquentiel et non optimisé.

Le tableau 4 compare nos résultats avec deux recensés dans (Kumar *et al.*, 2013). On remarque que notre méthode est meilleure que Q (Zhu et Milanfar, 2010) et ΔDOM (Kumar *et al.*, 2012) pour les deux mesures. Cependant, la méthode CORNIA (Ye *et al.*, 2012), basée sur un apprentissage plus lourd, reste plus performante selon ces critères.

Pour terminer, le tableau 5 synthétise les médianes des coefficients de la corrélation de Pearson obtenus pour les différents systèmes OCR. Le fait qu'on observe une très faible différence entre les systèmes FineReader et Tesseract, malgré leur très grande disparité en terme de performance réelle (voir (Kumar *et al.*, 2013), figures 4 et 6), laisse penser que la méthode d'évaluation proposée par (Kumar *et al.*, 2013) présente un biais optimiste. En effet, le calcul de la corrélation de Pearson sur l'ensemble des images (en intégrant donc les résultats moins avantageux) donne des résultats plus modestes. Pour les 175 images de la base, le coefficient global (LLC) que nous obtenons est de $0,6467$, ce qui est bien inférieur au coefficient de $0,9378$ donné par la méthode d'évaluation de l'état de l'art.

4. Conclusion

Pour faire face aux distorsions spécifiques contenues dans les images de documents numérisées avec des périphérique mobiles, nous avons proposé un prototype de

système d'acquisition permettant d'améliorer puis contrôler la qualité des images sur le mobile, avant leur transfert ou la perte de disponibilité du document.

La correction de la perspective est réalisée grâce à une détection des contours de la page, ce qui permet à la méthode de fonctionner même si la page contient peu de texte, ou si les coins ne sont pas visibles. La correction de l'illumination est réalisée grâce à une translation des composantes couleurs vers une illumination cible selon une hypothèse de monde gris, puis grâce à un étirement de l'histogramme pour la composante de luminance. L'évaluation de chacune de ces étapes sur des images réelles montre qu'elles apportent chacune un gain de terme de performance pour l'OCR.

L'estimation de la qualité de l'image est réalisée grâce à combinaison optimale de mesures de focus évaluées pour la première fois sur ce type de données. La validation de cette seconde étape sur une base publique montre que cette méthode surpasse les autres méthodes à base de métriques, et donne de bons résultats par rapport aux méthodes à base d'algorithmes plus lourds.

Le système résultant proposé est basé sur une combinaison de méthodes légères pouvant facilement être mise en œuvre sur un *smartphone*.

Remerciements

Ce projet a été financé par le consortium Valconum (`http://valconum.fr`). This project has been funded by the People Program (Marie Curie Actions) of the 7th Framework Program of the EU (FP7/2007-2013) under the grant agreement number 600388 of the REA.

5. Bibliographie

Ballard D., « Generalizing the Hough transform to detect arbitrary shapes », *Pattern Recognition*, vol. 13, n° 2, p. 111-122, 1981.

Chen F., Carter S., Denoue L., Kumar J., « SmartDCap : Semi-Automatic Capture of Higher Quality Document Images from a Smartphone », *Proceedings of the International Conference on Intelligent User Interfaces*, p. 287-296, 2013a.

Chen F., Carter S., Denoue L., Kumar J., « SmartDCap : Semi-Automatic Capture of Higher Quality Document Images from a Smartphone », *Proceedings of the 2013 international conference on Intelligent user interfaces (IUI'13)*, p. 287-296, March, 2013b.

Clark P., Mirmehdi M., « Estimating the Orientation and Recovery of Text Planes in a Single Image », *Proceedings of the 12th British Machine Vision Conference*, p. 421-430, 2001.

Fischler M., Bolles R., « Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography », *Communications of the ACM*, vol. 24, n° 6, p. 381-395, 1981.

Gonzalez J., Kanai J., Nartker T. A., « Prediction of OCR Accuracy Using a Neural Network », *in* S. L. Taylor, J. J. Hull (eds), *Document Analysis Systems II*, vol. 29, World Scientific, p. 356–370, 1998.

Jagannathan L., Jawahar C., « Perspective correction methods for camera based document analysis », *Proceedings of the First International Workshop on Camera-based Document Analysis and Recognition*, p. 148-154, 2005.

Kittler J., Hatef M., Duin R., Matas J., « On combining classifiers », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, n° 3, p. 226-239, March, 1998.

Kumar J., Chen F., Doermann D., « Sharpness estimation for document and scene images », *21st International Conference on Pattern Recognition (ICPR)*, p. 3292–3295, 2012.

Kumar J., Ye P., Doermann D., « A Dataset for Quality Assessment of Camera Captured Document Images », *International Workshop on Camera-Based Document Analysis and Recognition (CBDAR)*, 2013.

Lu S., Chen B., Ko C., « Perspective rectification of document images using fuzzy set and morphological operations », *Image and Vision Computing*, vol. 23, p. 541-553, 2005.

Peng X., Cao H., Subramanian K., Prasad R., Natarajan P., « Automated image quality assessment for camera-captured OCR », *18th IEEE International Conference on Image Processing (ICIP)*, p. 2621-2624, 2011.

Pertuz S., Puig D., Angel Garcia M., « Analysis of focus measure operators for shape-from-focus », *Pattern Recognition*, vol. 46, n° 5, p. 1415-1432, May, 2013.

Rodríguez-Piñeiro J., Comesaña-Alfaro P., Pérez-González F., Malvido-Garc ía A., « A new method for perspective correction of document images », *Proceedings of the Document Recognition and Retrieval XVIII*, p. 787410, 2011.

Ye P., Doermann D., « Learning features for predicting OCR accuracy », *21st International Conference on Pattern Recognition (ICPR)*, p. 3204–3207, 2012.

Ye P., Doermann D., « Document Image Quality Assessment : A Brief Survey », *12th International Conference on Document Analysis and Recognition (ICDAR)*, p. 723-727, 2013.

Ye P., Kumar J., Kang L., Doermann D., « Unsupervised feature learning framework for no-reference image quality assessment », *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 1098–1105, 2012.

Zhu X., Milanfar P., « Automatic parameter selection for denoising algorithms using a no-reference measure of image content », *IEEE Transactions on Image Processing*, vol. 19, n° 12, p. 3116–3132, 2010.