

Estimation de l'inclinaison d'un document arabe manuscrit numérisé par analyse temps-fréquence des histogrammes de projection

Skew angle estimation of scanned handwritten Arabic documents using a time-frequency analysis of the projection histograms

Nazih Ouwayed¹, Abdel Belaïd¹ et François Auger²

¹Université Nancy 2, LORIA, équipe READ, Vandœuvre-Lès-Nancy, France
nazih.ouwayed@loria.fr, abdel.belaid@loria.fr

²Université de Nantes, IREENA site de Saint-Nazaire, France
francois.auger@univ-nantes.fr

Manuscrit reçu le 10 novembre 2009

Résumé et mots clés

Nous présentons dans cet article une nouvelle méthode de détermination de l'inclinaison d'un document manuscrit arabe à l'aide d'une représentation temps-fréquence énergétique de la classe de Cohen. Cette méthode consiste à calculer d'abord les histogrammes de projection obtenus pour différents angles, puis à déterminer la valeur maximale de la représentation temps-fréquence de la racine carrée de ces histogrammes. L'orientation du document est alors estimée par l'angle de projection fournissant la valeur maximale la plus élevée. La méthode proposée a été testée sur 864 documents inclinés avec 9 représentations temps-fréquence différentes. Les résultats sont présentés et analysés à la fin de cet article.

Documents manuscrits, distributions d'énergie, classe de Cohen, histogramme de projection, estimation de l'angle d'orientation.

Abstract and key words

Ancient Arabic textual archives contain a heavy volume of handwritten documents that need to be scanned and indexed. Some of these documents are skewed, making their recognition and indexing difficult because straight lines are more suitable for the word extraction by recognition systems. We are looking for a method that can robustly estimate this orientation, whatever the size of the document. The scientific literature already proposes some solutions for image document skew angle estimation. The projection techniques seem the most appropriate ones but need to be adapted to Arabic documents. In fact, in Arabic script, the words are made of PAWs (Parts of Arabic Words) which are almost vertical or oblique and which may distort the calculation of local orientation. This prevents to apply local techniques like nearest neighbors, because of the alignment irregularity, or global techniques such as the Hough Transform because of the difficulty of locating voting points. Although these techniques fit well to printed documents, they remain inadequate to handwritten documents, in which the interline distance is random and the skew angle can be large. Kavallieratou et al. employed Cohen's class distributions on Latin documents. This Cohen's class contains all the quadratic time-frequency distributions that are covariant under time- and frequency-shifts. The members of this class are identified by a particular kernel $\phi_{dD}(\tau, \xi)$, which determines their theoretical properties and their practical readability.



In Kavallieratou's paper, the relationship between the distributions properties and the experimental results are not highlighted. We propose in this article to look for the most relevant properties related to the skew angle estimation problem and to find, thanks to them, the best distribution to use.

To estimate the orientation angle, we propose to compute a time-frequency representation of the analytic signal $x_a(t)$ of the centered squared root of the projection histogram $x(t)$ of the document. The projection angle corresponding to the histogram with the highest maximum value of its time-frequency representation is considered as an estimation of the document orientation. To study the effectiveness of our approach, we have experimented it on 864 Arabic handwritten documents. These documents have different sizes, contain several types of writing, layout (with 1 or 2 columns), a mix of text and tables, etc. The experiments were prepared after a manual orientation of the documents into different angles ranging from -75° to $+90^\circ$. We found that the Wigner-Ville distribution reaches the highest estimation rate (100%). The other distributions yield a lower estimation rate, either because they do not satisfy properties that are important for the skew angle estimation problem, such as the scale invariance property and the support conservation, because their localization of the signal components is not sufficiently precise to provide a skew angle estimation with the maximum of the representation, or because the parameters of these distributions are not fitted to the analysed histogram profiles. The skew angle estimator using the Wigner-Ville distribution is also compared to the projection analysis and Fourier Transform methods.

Handwritten documents, energy distributions, Cohen's class, projection histograms, skew angle estimation.

Remerciements

Nous tenons à remercier la Bibliothèque Nationale de Tunis, la Bibliothèque Nationale de médecine des U.S.A et les Archives et la Bibliothèque Nationales d'Egypte, qui ont permis au public d'accéder librement aux documents qui ont constitué notre source principale de données.

1. Introduction

Lors de la numérisation d'un document, une inclinaison du texte par rapport à l'axe horizontal de l'image se produit très souvent, en raison d'un mauvais positionnement des pages ou d'une erreur de manipulation. Une phase d'estimation de l'orientation d'un document est donc indispensable avant de passer aux étapes ultérieures de segmentation et de reconnaissance.

De nombreuses approches ont été déjà proposées pour la détermination de cette inclinaison. Dans plusieurs travaux [1, 40], les histogrammes de projection de l'ensemble du document ou d'une bande de celui-ci sont calculés pour différents angles. Puis les différences entre les pics et les vallées de ces histogrammes sont calculées pour estimer l'angle d'inclinaison du document. D'autres travaux [27, 38] utilisent la classification par k-ppv, où les composantes connexes sont regroupées en k-groupes. L'orientation au sein de chaque groupe est d'abord déterminée, puis celle de l'image est obtenue en faisant leur moyenne. On trouve également des travaux qui calculent des valeurs cumulées des corrélations entre chaque paire de lignes consécutives [10, 39]. La transformée de Hough a été utilisée

dans [42, 45], associée à des points votants, c'est-à-dire des points probablement colinéaires. Ces points peuvent être les centres de gravité ou les pixels les plus bas [42] des composantes connexes. L'inconvénient de cette transformation est qu'elle nécessite un grand espace mémoire et sa durée d'exécution est rédhibitoire. La valeur maximale du module carré de la transformée de Fourier a aussi été utilisée pour estimer cet angle d'inclinaison [41, 48], ainsi que moindres carrés, par une recherche de l'axe le plus représentatif des données [4, 27]. Deux points au hasard sur cet axe permettent alors d'estimer l'orientation. Dans [11, 17], les opérations morphologiques comme la dilatation, l'érosion, l'ouverture et la fermeture sont utilisées pour transformer les lignes de texte en bandes noires de pixels et pour supprimer les connexions entre ces bandes. Deux points dans chaque bande (ligne) sont identifiés pour mesurer l'inclinaison de la ligne et l'angle moyen de toutes ces inclinaisons est considéré comme l'orientation de l'ensemble du document. Sauvola *et al.* [44] ont également estimé l'orientation en cherchant des alignements (appelés « runs ») qui correspondent à des lignes de texte espacées par des zones qui ne contiennent pas de pixels noirs. Cette méthode a été appliquée pour les documents latins imprimés. Elle ne pourrait probablement pas

fournir de bons résultats pour des documents arabes manuscrits, car ces derniers ne possèdent généralement pas d'interligne constante, et parce que les lignes sont très proches et se touchent même dans la plupart des cas. D'autres approches utilisent les caractéristiques particulières de l'alphabet de chaque langue. Par exemple, dans [39], la ligne supérieure qui se trouve dans tous les caractères des alphabets indiens (Devnagari et Bangla) est utilisée pour estimer l'orientation.

Dans cet article, nous présentons une nouvelle méthode d'estimation de l'angle d'inclinaison de documents arabes manuscrits, qui s'appuie sur une analyse temps-fréquence des histogrammes de projection du document. Kavallieratou *et al.* [30, 31] ont déjà utilisé la distribution de Wigner-Ville (DWV) pour estimer l'orientation globale des documents latins imprimés ou manuscrits. Dans le présent article, nous avons souhaité déterminer d'abord les propriétés des représentations temps-fréquence qui semblent souhaitables pour une telle application, afin d'établir ensuite une liste de représentations, dont les performances ont enfin été évaluées.

Cet article est organisé de la manière suivante : dans la section 2, nous expliquons comment calculer l'orientation en analysant les histogrammes de projection. Dans la section 3, nous rappelons ensuite brièvement la définition et les propriétés des représentations temps-fréquence de la classe de Cohen. L'algorithme proposé est exposé en détail dans la section 4. Enfin, nous présentons des résultats expérimentaux qui permettent une évaluation comparative des performances de la méthode proposée et nous terminons par une conclusion.

2. Calcul de l'orientation par analyse des histogrammes de projection

La technique classique d'analyse des histogrammes de projection [1, 40] permet de percevoir plusieurs caractéristiques du contenu d'un document écrit, en particulier la densité du document, les interlignages, la distribution des blocs et bien sûr l'inclinaison. Pour estimer cette dernière, deux cas peuvent se présenter :

Cas 1 : le document ne contient qu'une seule ligne. L'orientation est estimée en calculant les histogrammes pour des angles de projection espacés de quelques degrés, retenant celui qui est le plus concentré (voir figure 1, où l'inclinaison correcte est de 0°).

Cas 2 : le document contient plusieurs lignes. L'orientation est déterminée par l'histogramme de projection qui a le plus grand écart moyen entre les pics et les vallées, possédant donc les pics les plus élevés et les vallées les plus basses (voir figure 2, où l'inclinaison correcte est de 0°).

En général, on suppose que le document analysé comporte plusieurs lignes, et on retient uniquement le deuxième cas.

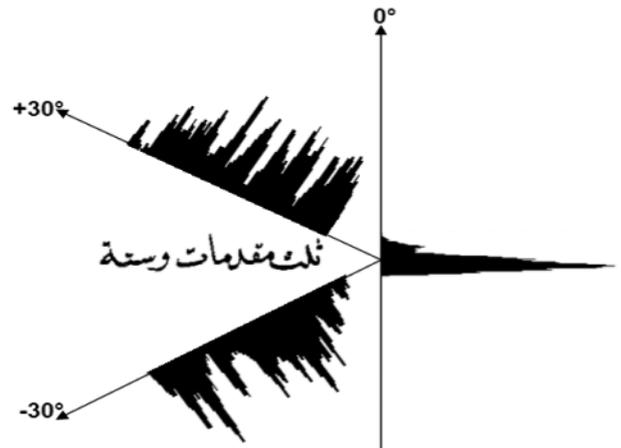


Figure 1. Histogrammes de projection d'un document ne comportant qu'une seule ligne, calculés pour des angles de projection de $+30^\circ$, 0° et -30° .

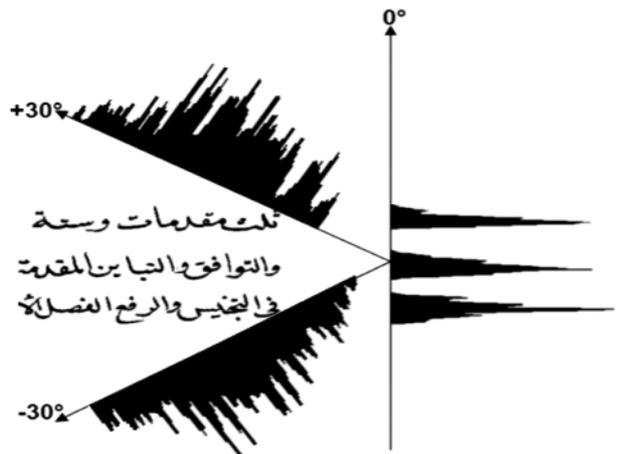


Figure 2. Histogrammes de projection d'un document comportant plusieurs lignes, calculés pour des angles de projection de $+30^\circ$, 0° et -30° .

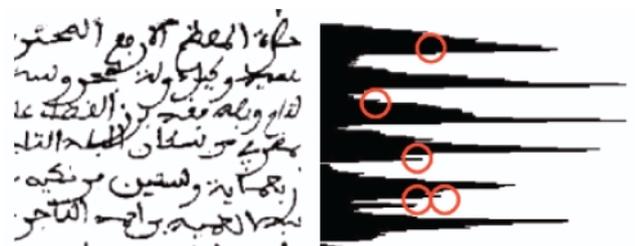


Figure 3. Histogramme de projection (à droite) du document arabe manuscrit complexe reproduit à gauche, calculé pour un angle de projection de 0° . Les cercles désignent des maxima locaux indésirables.

Cette approche classique ne conduit pas toujours à une estimation correcte de l'orientation. C'est le cas notamment pour des documents manuscrits complexes qui présentent beaucoup de

bruit, de symboles diacritiques, de chevauchements et de connexions entre les lignes. L'irrégularité de l'écriture et la présence de hampes et de jambages créent des maxima locaux dans les histogrammes de projection (voir par exemple la figure 3), et diminuent la valeur de la différence moyenne entre les pics et les vallées. Pour pouvoir disposer d'un estimateur plus robuste, nous avons donc envisagé d'utiliser des représentations temps-fréquence de ces histogrammes.

3. Les distributions temps-fréquence

Les approches classiques de traitement du signal comme la transformée de Fourier ne parviennent pas à décrire simultanément la variation d'un signal par rapport au temps et à la fréquence. Les distributions temps-fréquence énergétiques vont au delà de ce qu'autorisent ces approches classiques en analysant la «non-stationnarité» d'un signal et en distribuant son espace à la fois temporel et fréquentiel. D'après [25], l'énergie E_x d'un signal $x(t)$ a pour expression

$$S E_x = \int_{-\infty}^{+\infty} |x(t)|^2 dt = \int_{-\infty}^{+\infty} |\widehat{x}(f)|^2 df \quad (1)$$

où $\widehat{x}(f)$ est la transformée de Fourier du signal $x(t)$. Cette énergie étant une quantité quadratique, il est légitime de souhaiter que les distributions temps-fréquence énergétiques conservent cette propriété.

En 1966, Cohen [14, 18] a montré qu'un nombre important de distributions temps-fréquence énergétiques peuvent être vues comme des cas particuliers d'une expression générale :

$$C_x(t, f) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \phi_{r-d}(\tau, \xi) A_x(\tau, \xi) e^{j2\pi(t\xi - f\tau)} d\xi d\tau, \quad (2)$$

où $A_x(\tau, \xi)$ est la fonction d'ambiguïté à bande étroite du signal, définie par :

$$A_x(\tau, \xi) = \int_{-\infty}^{+\infty} x(t + \tau/2) x^*(t - \tau/2) e^{-j2\pi\xi t} dt$$

La classe de Cohen contient toutes les distributions temps-fréquence quadratiques qui vérifient la propriété d'invariance par translation temporelle ou fréquentielle [19, 25, 26]. Chacun des membres de cette classe se distingue par un noyau $\phi_{r-d}(\tau, \xi)$, qui joue un rôle déterminant dans la qualité des images fournies et dans les propriétés [2, 13, 15, 20, 21, 26] qu'il vérifie. L'utilisation de ces distributions est très répandue, notamment dans des applications de reconnaissance des formes [7, 16] ou pour le traitement des signaux sismiques [8].

Dans notre contexte d'application, un ensemble de propriétés nous ont semblé nécessaires ou souhaitables. Elles sont regroupées dans le tableau 1 et commentées ci-contre :

Tableau 1. Propriétés des distributions temps-fréquence de la classe de Cohen considérés comme nécessaires ou souhaitables dans ce contexte d'application.

No	Propriété	Expression
P1	Valeur réelle	$\forall t, f, C_x(t, f) = C_x^*(t, f)$ où $C_x^*(t, f)$ est le conjugué de $C_x(t, f)$
P2	Covariance vis-à-vis des translations temporelles et fréquentielles	$\forall t_1, f_1,$ $y(t) = x(t - t_1) e^{j2\pi f_1 t} \Rightarrow$ $C_y(t, f) = C_x(t - t_1, f - f_1)$
P3	Covariance vis-à-vis des changements d'échelle	$\forall a > 0,$ $y(t) = a^{-1/2} x(t/a) \Rightarrow$ $C_y(t, f) = C_x(t/a, af)$
P4	Conservation de l'énergie	$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} C_x(t, f) dt df$ $= \int_{-\infty}^{+\infty} x(t) ^2 dt$
P5	Positivité	$\forall x(t), \forall t, f, C_x(t, f) \geq 0$
P6	Conservation des valeurs nulles du signal	$x(t_0) = 0 \Rightarrow$ $\forall f, C_x(t_0, f) = 0$
P7	Conservation du support du signal	si $\forall t, t > T \Rightarrow x(t) = 0,$ alors $\forall f, t > T \Rightarrow C_x(t, f) = 0$

Propriété P1 : valeur réelle. Cette propriété permet d'associer à chaque point du plan temps-fréquence un nombre réel et non un nombre complexe, ce qui réduit la quantité de mémoire nécessaire au stockage de la distribution. Cela permet aussi de construire facilement une représentation graphique de la distribution temps-fréquence, par exemple par une image indiquant par des niveaux de gris la valeur de la distribution en chaque point. Enfin, cela permet d'utiliser une relation d'ordre pour comparer les valeurs entre elles et en extraire le maximum. Par exemple, la valeur maximale de la distribution de Wigner-Ville du signal analytique [6] du signal de la figure 4.a est 47.117.

Propriété P2 : covariance vis-à-vis des translations temporelles et fréquentielles. Cela veut dire qu'un simple décalage du texte dans l'image se traduira par un décalage de la représentation temps-fréquence de l'histogramme sans changer la valeur de son maximum. Ceci est intéressant car elle montre que la distribution n'est pas sensible à la localisation des lignes dans l'image. Par exemple, sur la figure 4.c, les deux images sont obtenues par translation verticale du document. Elles conduisent à des versions translattées des histogrammes de projection et donc à des versions translattées des représentations temps-fréquence, qui ont donc le même maximum absolu.

Propriété P3 : covariance vis-à-vis des changements d'échelle. Grâce à cette propriété, un agrandissement ou une réduction du document se traduira par une dilatation ou une compression de son histogramme, qui se traduira par une modi-

Tableau 2. Distributions temps-fréquences retenues dans cette étude, avec leurs expressions et leurs propriétés.

Distribution	Expression	Propriétés						
		P1	P2	P3	P4	P5	P6	P7
Wigner-Ville	$\int_{-\infty}^{+\infty} x(t + \tau/2) x^*(t - \tau/2) e^{-j2\pi f\tau} d\tau$	x	x	x	x			x
Pseudo Wigner-Ville	$\int_{-\infty}^{+\infty} h(\tau) x(t + \tau/2) x^*(t - \tau/2) e^{-j2\pi f\tau} d\tau$	x	x		x			x
Pseudo Wigner-Ville lissée	$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(t - \nu) h(\tau) x(\nu + \tau/2) x^*(\nu - \tau/2) e^{-j2\pi f\tau} d\tau d\nu$	x	x		x			
Margenau-Hill	$\frac{1}{2} \int_{-\infty}^{+\infty} [x(t + \tau)x^*(t) + x(t)x^*(t - \tau)] e^{-j2\pi f\tau} d\tau$	x	x	x	x			x x
Pseudo Margenau-Hill	$\frac{1}{2} \int_{-\infty}^{+\infty} h(\tau) [x(t + \tau)x^*(t) + x(t)x^*(t - \tau)] e^{-j2\pi f\tau} d\tau$	x	x		x			x x
Spectrogramme	$\left \int_{-\infty}^{+\infty} x(u) h^*(u - t) e^{-j2\pi fu} du \right ^2$	x	x		x			x
Choi-Williams	$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\exp(-\frac{\nu^2}{4\tau^2/\sigma})}{\sqrt{4\pi\tau^2/\sigma}} x(t - \nu + \tau/2) x^*(t - \nu - \tau/2) e^{-j2\pi f\tau} d\nu d\tau$	x	x	x	x			
Born-Jordan	$\int_{-\infty}^{+\infty} \int_{- \tau /2}^{ \tau /2} \frac{1}{ \tau } x(t - \nu + \tau/2) x^*(t - \nu - \tau/2) e^{-j2\pi f\tau} d\nu d\tau$	x	x	x	x			x
Zhao-Atlas-Marks	$\int_{-\infty}^{+\infty} \int_{- \tau /2}^{ \tau /2} h(\tau) x(t - \nu + \tau/2) x^*(t - \nu - \tau/2) e^{-j2\pi f\tau} d\nu d\tau$	x	x					x

fication en conséquence de sa représentation temps fréquence. Par exemple, le maximum de la représentation de Wigner-Ville du signal analytique du signal $y(t) = 2x(4t)$ (voir figure 4.b) est le même que celle de $x(t)$.

Propriété P4 : conservation de l'énergie. Cette propriété permet de garantir que l'énergie du signal analysé est exactement distribuée dans le plan temps-fréquence.

Propriété P5 : positivité. Cette propriété permet d'interpréter plus facilement les valeurs de la représentation comme une mesure d'énergie.

Propriété P6 : conservation des valeurs nulles du signal. Cette propriété permet d'avoir une représentation nulle à toutes les « fréquences » aux « instants » où l'histogramme de projection est nul. Ceci permet de retrouver dans la représentation temps-fréquence les vallées de cet histogramme de projection (voir figure 4.d).

Propriété P7 : conservation du support du signal. Ceci permet d'avoir une représentation temps-fréquence nulle à l'extérieur de l'intervalle dans lequel l'histogramme de projection prend des valeurs non nulles (la représentation temps-fréquence ne déborde pas du support de l'histogramme).

Pour ce problème d'estimation de l'inclinaison d'un document numérisé, les propriétés P1, P2, P3 et P4 nous paraissent fortement souhaitables, alors que les propriétés P5, P6 et P7 nous semblent moins indispensables, mais susceptibles de fournir des estimations plus précises.

Les propriétés retenues nous ont permis d'établir ensuite une liste de représentations susceptibles de fournir de bonnes estimations de l'inclinaison. Ces représentations sont énumérées dans le tableau 2, avec leurs expressions et leurs propriétés.

4. Application des distributions temps-fréquence pour l'estimation de l'orientation



Lorsqu'on calcule l'histogramme de projection d'un document suivant une direction de projection, on obtient, lorsque cette direction correspond à l'orientation du document, un histogramme dans lequel chaque ligne d'écriture conduit à un maximum local nettement marqué (voir figure 2). Chaque paragraphe du document se répercute dans l'histogramme de projection par une succession périodique de pics et de vallées dont la période est sensiblement constante. Cette succession périodique constitue alors une composante de l'histogramme délimitée dans le « temps » par la taille du paragraphe et qui oscille à une « fréquence » déterminée par la largeur de l'interligne.

Comme tous les pixels s'accroissent aux mêmes endroits, les maxima locaux ont des niveaux d'énergie plus élevés que pour les autres directions de projection, ce qui explique que l'on peut estimer l'orientation d'un document en cherchant l'angle de projection pour laquelle la distribution temps-fréquence obtenue à partir de cet histogramme localise la plus grande quantité d'énergie en un point quelconque du plan temps-fréquence. À titre d'illustration, la figure 5 montre l'évolution de la valeur maximale de la distribution de Wigner-Ville lorsque le nombre de pics et de vallées augmente et lorsque les vallées se creusent dans le signal.

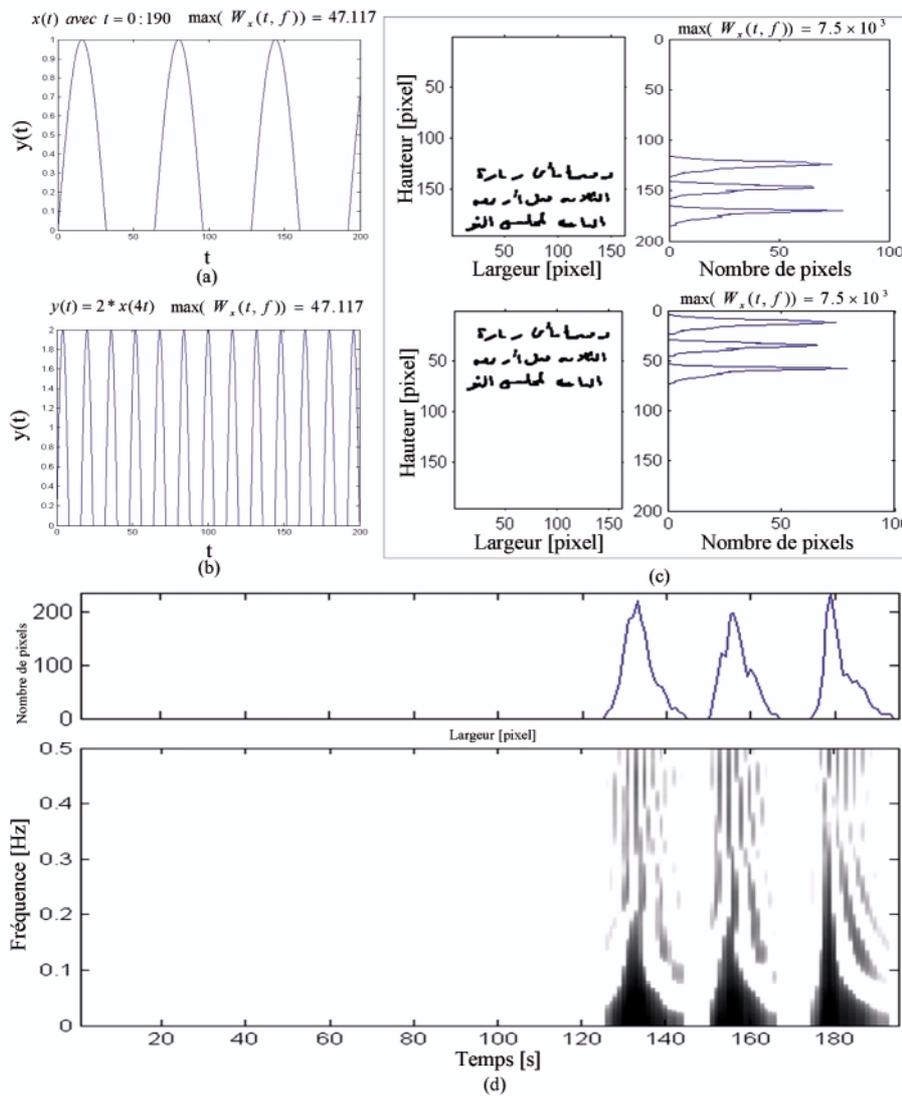


Figure 4. Illustration des propriétés des distributions de la classe de Cohen et plus particulièrement de la distribution de Wigner-Ville (figures a, b et c) et de la distribution pseudo Margenau-Hill (figure d). Le signal $x(t)$ utilisé dans les figures a et b est égal à $\max(\sin(2\pi t/64), 0)$.

L'utilisation des représentations temps-fréquence de la classe de Cohen nous semble particulièrement justifiée dans ce contexte. En effet, l'histogramme de projection d'un document numérisé, obtenu pour une direction de projection donnée, est un signal non-stationnaire, constitué de composantes qui correspondent aux paragraphes du document et qui oscillent à une fréquence déduite de la largeur de l'interligne. Lorsque la direction de projection est égale à la direction principale du document, l'énergie est alors très localisée en « temps » (la direction perpendiculaire à la direction de projection) et en « fréquence » (la fréquence spatiale associée), ce qui justifie d'estimer cet angle en recherchant la direction de projection pour laquelle la représentation temps-fréquence possède un maximum global. Ceci nécessite l'emploi d'une représentation à valeurs réelles, donc quadratique, car les représentations linéaires (telles que la transformée

de Fourier court-terme [23] ou la transformée en ondelettes continue [5]) sont à valeurs complexes. Enfin, même si l'acquisition de l'image du document fait intervenir un agrandissement ou une réduction, qui s'apparentent à une compression ou à une dilatation, le document en lui-même est caractérisé par un interligne choisi par son auteur, dont la variation conduit, sur l'histogramme de projection, davantage à une modulation de fréquence qu'à un changement d'échelle : une augmentation de l'interligne ne conduit pas nécessairement à une augmentation de la taille des caractères. C'est pour cette raison que les représentations temps-fréquence, si possible également invariantes par changement d'échelle, nous semblent plus pertinentes pour cette application que des représentations affines [37]. Les représentations temps-fréquence utilisées étant quadratiques, c'est à la racine carrée des histogrammes que nous les

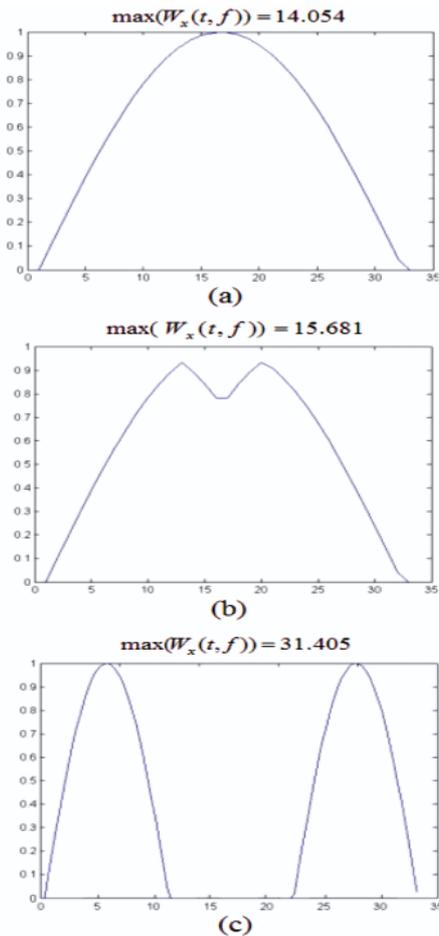


Figure 5. Exemples de valeurs maximales de la distribution de Wigner-Ville obtenues pour différents histogrammes de projection.

appliquons. Les valeurs des représentations possèdent alors la même dimension physique que les histogrammes et correspondent donc au nombre de pixels noirs situés sur une droite parallèle à la direction de projection et espacés d'un certain interligne. La propriété de conservation de l'énergie permet alors de conserver l'intégrale de l'histogramme, c'est-à-dire le nombre total de pixels noirs contenus dans l'image. La propriété de positivité correspond au souhait de ne pas avoir de valeurs négatives lorsque le nombre total de pixels noirs est distribué dans le plan temps-fréquence. Les propriétés de conservation des valeurs nulles ou des supports correspondent au souhait d'avoir des effectifs nuls dans le plan temps-fréquence aux endroits où l'histogramme ne dénombre aucun pixel. Le calcul de la représentation temps-fréquence de la racine carrée des histogrammes et non pas des histogrammes, comme il est fait dans [30, 31], renforce donc la signification des valeurs de la représentation et de ses propriétés. Les estimations de l'inclinaison obtenues dans les deux cas ne diffèrent cependant pas de manière significative. Pour estimer l'angle d'inclinaison, la méthode que nous proposons consiste donc à calculer le signal analytique $x_a(t)$ de la racine carrée de l'histogramme de projection $x(t)$ du document,



Figure 6. Document arabe manuscrit incliné de $+14.7^\circ$. La position du centre de gravité des pixels noirs de ce document, utilisé pour calculer les histogrammes de projection, est indiquée par une croix.

auquel a été retranchée sa valeur moyenne. Le signal analytique est obtenu à partir du signal $x(t)$ en lui retirant ses fréquences négatives [6]. L'histogramme $x(t)$ est obtenu en projetant le document suivant un axe qui passe par le centre de gravité de l'image, qui constitue un point déduit du contenu de l'image, et non pas de ses dimensions et de son cadrage, et incliné d'un certain angle par rapport à l'axe horizontal de l'image, utilisé comme référence angulaire (voir figure 8). Nous calculons ensuite la représentation temps-fréquence de ce signal analytique. L'angle pour lequel la valeur maximale de cette représentation temps-fréquence est la plus élevée constitue l'estimation de l'angle d'inclinaison du document. À titre d'exemple, la figure 7 montre l'évolution, en fonction de l'angle de projection, du maximum de la distribution de Wigner-Ville calculée à partir des histogrammes de projection du document de la figure 6. Les coordonnées du point (t, f) du plan temps-fréquence où la valeur maximale de la représentation est obtenue pour l'inclinaison estimée pourraient être utilisées pour estimer l'emplacement du centre de gravité du paragraphe principal du document et la largeur de son interligne. Pour réduire le coût de calcul et la quantité de mémoire nécessaire, une technique de recherche du maximum en trois phases est utilisée. Dans la première phase, les histogrammes de projection sont calculés pour des angles allant de -85° à $+90^\circ$

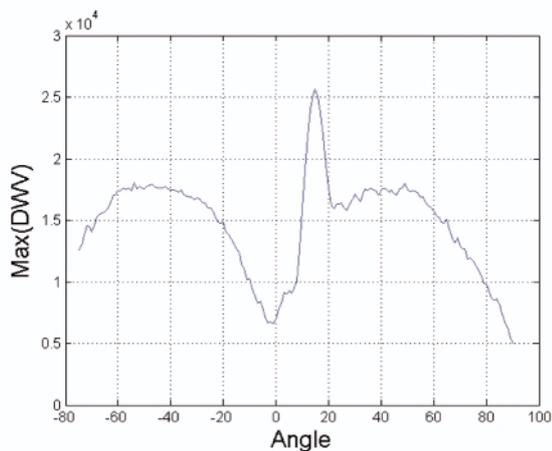


Figure 7. Courbe du maximum de la distribution de Wigner-Ville de la racine carrée des histogrammes de projection du document de la figure 6, en fonction de l'angle de projection.

avec un pas de 5°. Dans la seconde phase, les angles de projection utilisés varient de -4° à +4° autour de l'inclinaison $\hat{\theta}_1$ estimée à la première phase, avec un pas de 1°. Dans la troisième phase, l'estimation est affinée en utilisant des angles de projection entre -0.9° et +0.9° autour de l'inclinaison $\hat{\theta}_2$ obtenue à

la phase précédente, avec un pas de 0.1°. Dans chacune des trois phases, l'inclinaison est estimée en cherchant l'angle de projection pour lequel la plus grande valeur maximale de la représentation temps-fréquence est obtenue.

Tableau 3. Estimations de l'angle d'inclinaison du document de la figure 6 obtenues avec les représentations temps-fréquence retenues dans cette étude.

Distribution	Angle estimé
Wigner-Ville	+14.7°
Pseudo Wigner-Ville	+13.1°
Pseudo Wigner-Ville lissée	+14.3°
Margenau-Hill	+13.7°
Pseudo Margenau-Hill	+15.1°
Spectrogramme	+12.9°
Choi-Williams	+14.2°
Born-Jordan	+12.8°
Zhao-Atlas-Marks	+14.6°
histogrammes de projection	+14.9°
Transformée de Fourier	+13.8°

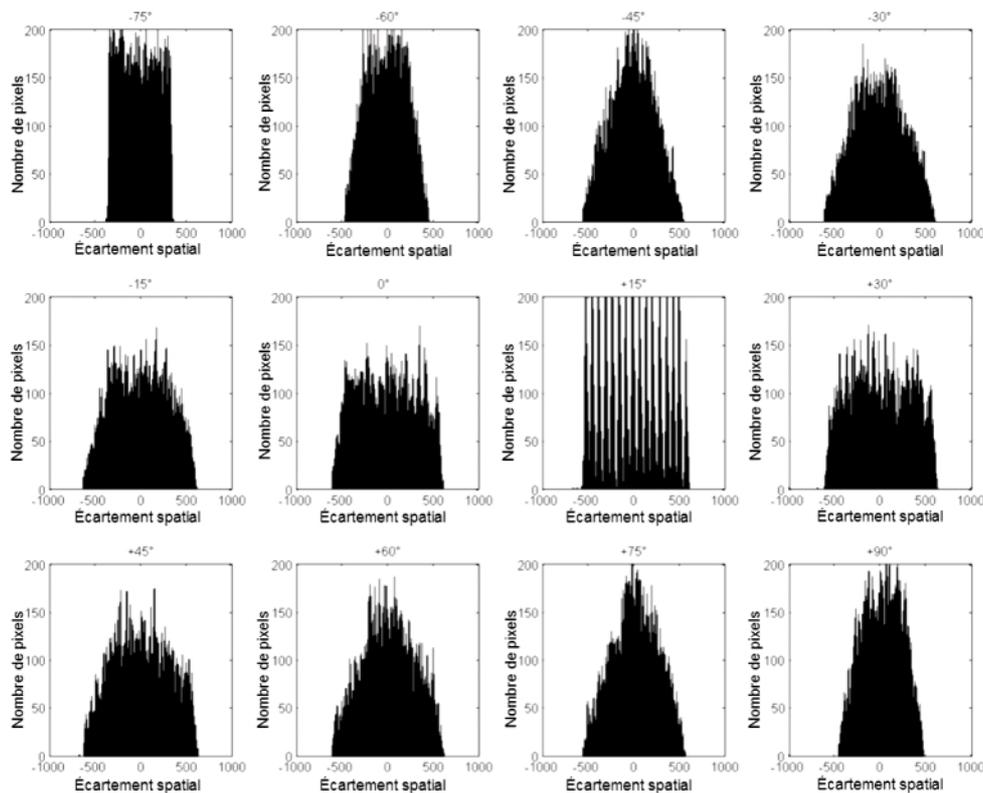


Figure 8. Quelques histogrammes de projection du document de la figure 6.

Tableau 4. Taux d'estimation correcte de l'inclinaison obtenus avec les représentations temps-fréquence retenues dans cette étude. Le temps moyen correspond à la durée moyenne de traitement d'une image de 500 × 500 pixels.

Distribution	taux d'estimation (%)	temps moyen (s)
Wigner-Ville	100	4.2
Pseudo Wigner-Ville	92	7.1
Pseudo Wigner-Ville lissée	97	6.9
Margenau-Hill	97	7.2
Pseudo Margenau-Hill	98	8.1
Spectrogramme	92	7.9
Choi-Williams	97	7.6
Born-Jordan	90	7.8
Zhao-Atlas-Marks	94	8.5
histogrammes de projection transformée de Fourier	93	6.4
	89	4.0

tailles, contiennent plusieurs types d'écriture, de mise en page (une ou deux colonnes) et des juxtapositions de texte et de tableaux. Lors de la préparation des essais, les documents ont été inclinés manuellement, avec des angles connus variant entre -75° et +90°. Les essais ont été effectués sur un PC sous Windows XP, équipé d'un micro processeur Pentium M de 1.4GHz et d'une mémoire cache de 1 Go, avec Matlab R2007b complété par la boîte à outils temps-fréquence tftb [3].



Figure 9. Partie d'un document arabe manuscrit. La largeur et la hauteur de cette image sont de 300 pixels et son inclinaison est de 0°.

5. Résultats expérimentaux et discussion

Pour évaluer et étudier l'efficacité de notre approche, nous l'avons testée sur un corpus 864 documents arabes manuscrits provenant de différentes bibliothèques. Ces documents ont différentes

Tableau 5. Estimations de l'inclinaison des six documents de la figure 10 obtenues à l'aide des neuf représentations temps-fréquence retenues dans cette étude.

documents de la figure 10		(a)	(b)	(c)	(d)	(e)	(f)
taille du document (largeur × hauteur), en pixels		521*565	553*621	825*749	573*461	417*577	569*577
résolution de la numérisation (dpi)		96	72	72	72	100	100
angle exact		-36.0°	+30.0°	-67.0°	+55.0°	+10.0°	-17.0°
Wigner-Ville	angle estimé	-36.0°	+30.0°	-67.0°	+55.0°	+10.0°	-17.0°
	temps d'exécution (s)	5.1	5.9	10.0	6.0	4.6	5.3
Pseudo Wigner-Ville	angle estimé	-34.5°	+26.0°	-69.0°	+57.0°	+13.0°	-19.0°
	temps d'exécution (s)	7.3	7.9	15.0	8.0	7.4	7.6
Pseudo Wigner-Ville lissée	angle estimé	-37.0°	+30.0°	-67.0°	+55.0°	+11.0°	-15.5°
	temps d'exécution (s)	7.1	7.4	13.1	7.6	6.9	7.2
Margenau-Hill	angle estimé	-36.0°	+30.0°	-66.0°	+55.0°	+11.0°	-17.0°
	temps d'exécution (s)	7.6	8.1	12.1	8.5	7.1	8.5
Pseudo Margenau-Hill	angle estimé	-36.0°	+29.5°	-66.5°	+55.5°	+10.0°	-17.5°
	temps d'exécution (s)	8.4	7.1	16.1	7.1	8.3	8.3
Spectrogramme	angle estimé	-37.0°	+30.5°	-67.0°	+51.0°	+12.0°	-15.0°
	temps d'exécution (s)	8.6	8.0	14.3	7.9	8.9	8.5
Choi-Williams	angle estimé	-36.0°	+30.0°	-65.0°	+55.0°	+11.0°	-17.0°
	temps d'exécution (s)	7.9	8.6	13.2	8.5	8.2	7.8
Born-Jordan	angle estimé	-35.5°	+28°	-66.5°	+53.5°	+12.5°	-19.5°
	temps d'exécution (s)	8.4	8.8	15.2	8.4	8.4	8.1
Zhao-Atlas-Marks	angle estimé	-35°	+30.5°	-67.5°	+55.5°	+10.5°	-17.5°
	temps d'exécution (s)	9.1	9.6	16.3	9.7	8.9	9.4

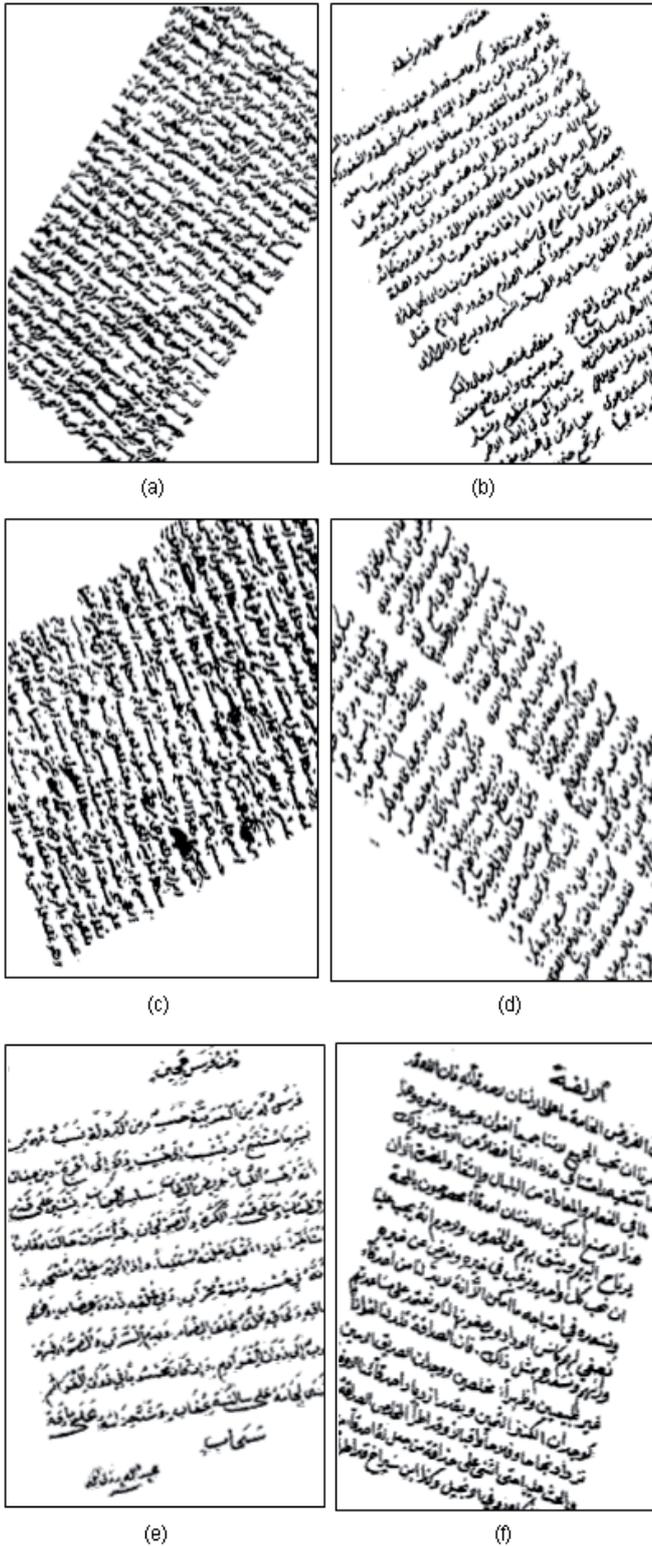


Figure 10. Échantillon de 6 documents parmi les 864 documents traités.

Le tableau 3 montre les estimations de l'inclinaison du document de la figure 6 obtenues avec les représentations temps-fréquence retenues dans cette étude. L'ensemble de ces estimations se situent entre -2.1° et $+0.5^\circ$ de la valeur exacte, égale

à 14.7° . Pour ce document, la distribution de Wigner-Ville semble être la seule à donner un résultat plus intéressant que l'analyse des histogrammes de projection. L'estimation de l'inclinaison obtenue en cherchant l'angle de projection qui maximise le module carré de la transformée de Fourier de la racine carrée de l'histogramme de projection [41, 48] est moins bonne que les estimations obtenues avec cinq des neuf représentations temps-fréquence retenues. Le tableau 5 illustre l'efficacité de notre algorithme sur un échantillon de 6 documents représentatifs (voir figure 10) choisis parmi les 864 documents traités.

Le tableau 4 présente la synthèse des résultats obtenus sur l'ensemble des 864 documents. Le taux d'estimation correcte correspond au pourcentage de cas où l'inclinaison estimée est à $\pm 0.5^\circ$ de l'inclinaison exacte, intervalle correspondant aux besoins réels des algorithmes d'extraction de lignes et de reconnaissance automatique des caractères, qui suivent dans la chaîne de traitement. Le temps moyen, mesuré en secondes, correspond à la durée de traitement d'une image de 500×500 pixels. Il dépend de la taille du document et de la représentation utilisée. Comme on peut le voir dans ce tableau, le meilleur taux d'estimation correcte est obtenu avec la distribution de Wigner-Ville et est égal à 100%. Les résultats obtenus avec les autres représentations temps-fréquence sont plus faibles, soit parce qu'elles ne vérifient pas des propriétés qui nous semblent nécessaires pour cette application, comme les propriétés d'invariance par changement d'échelle et de conservation des supports, soit parce que la localisation des composantes du signal qu'elles fournissent n'est pas suffisante pour délivrer une estimation correcte de l'inclinaison par leur valeur maximale, soit parce que les paramètres de réglage de ces représentations ne sont pas adaptés. Ces paramètres sont

- a) la fenêtre de lissage temporel $g(t)$;
- b) la fenêtre de lissage fréquentiel $h(\tau)$;
- c) la largeur σ du noyau de la distribution Choi-Williams.

Pour simplifier les comparaisons, toutes les fenêtres de lissage ont été choisies du type Kaiser-Bessel [29]. Les meilleurs résultats ont alors été obtenus pour des longueurs des fenêtres de lissage temporel de 19 points et des longueurs des fenêtres de lissage fréquentiel de 91 points. De meilleurs résultats pourraient peut-être être obtenus en adaptant ces longueurs à la dimension des histogrammes, qui varie en fonction de l'angle de projection. La valeur du paramètre de réglage de la représentation de Choi-Williams qui a fourni les meilleurs résultats est $\sigma = 3.6$.

Le tableau 4 permet également de voir que sur les neuf représentations étudiées, six fournissent de meilleurs résultats que la méthode d'analyse des histogrammes de projection. Ceci montre que l'analyse temps-fréquence apporte bien une solution aux problèmes dus à la présence de maxima locaux dans les histogrammes de projection. Le faible résultat obtenu avec le module carré de la transformée de Fourier semble montrer qu'il est important de s'appuyer sur le caractère non-stationnaire des histogrammes. Cette méthode convient sans doute mieux à des documents latins et/ou imprimés, pour lesquels l'inclinaison est beaucoup plus facile à estimer.

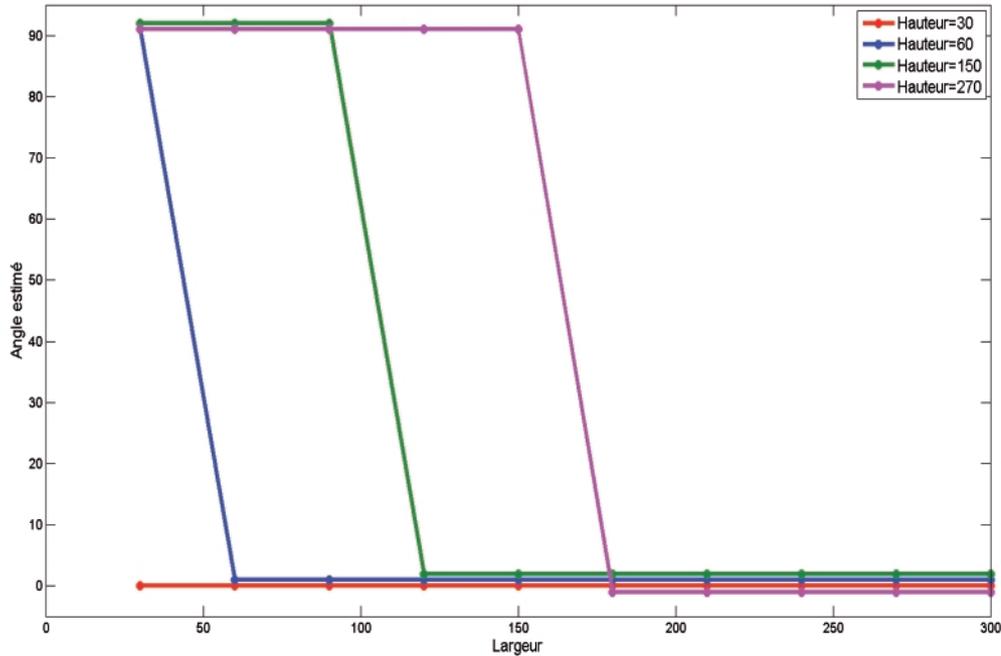


Figure 11. Évolution de l'estimation de l'angle d'inclinaison du document de la figure 9 en fonction de la largeur du document, pour quatre valeurs différentes de la hauteur. L'inclinaison est estimée en utilisant la distribution de Wigner-Ville.

Nous avons également pu mettre en évidence que l'algorithme proposé fournit une estimation aberrante lorsque la hauteur de l'image est plus grande que sa largeur. Les histogrammes de projection dénombrent alors davantage de pixels dans le sens de la hauteur que dans le sens de la largeur. La figure 11 montre que lorsque la largeur et la hauteur sont égales, l'estimation obtenue est correcte.

6. Conclusion et perspectives

Nous avons proposé dans cet article une approche originale qui permet d'estimer de manière robuste l'inclinaison d'un document arabe manuscrit numérisé. À cause des interlignes variables, des chevauchements et de la présence de hampes et de jambages dans les documents arabes manuscrits, ce problème d'estimation est difficile, et les techniques classiques conçues et utilisées principalement pour les caractères latins ne fournissent pas de résultats satisfaisants. L'approche proposée est basée sur la maximisation de la représentation temps-fréquence du signal analytique de la racine carrée des histogrammes de projection. Ceci permet de distribuer dans un plan « temps-fréquence » le nombre de pixels noirs de l'image et de chercher pour quel angle de projection cette population est la plus concentrée dans ce plan. En nous basant sur une liste de propriétés pertinentes pour

ce problème, nous avons retenu neuf représentations temps-fréquence de la classe de Cohen. La distribution de Wigner-Ville s'est avérée fournir les meilleurs résultats. Ceci s'explique sans doute par le fait que cette représentation ne possède aucun paramètre de réglage, qu'elle fournit une localisation extrêmement concentrée des composantes des signaux et qu'elle vérifie deux propriétés importantes pour cette application : l'invariance par changement d'échelle et la conservation des supports. À partir de cet estimateur, nous avons pu construire [34, 35, 36] des algorithmes d'analyse et d'extraction des lignes de documents manuscrits complexes, constitués de plusieurs zones ayant des inclinaisons différentes.

Références

- [1] T. AKIYAMA and N. HAGITA, Automatic entry system for printed documents. *Pattern Recognition*, 23: 1141-1154, 1990.
- [2] F. AUGER and C. DONCARLI, Quelques commentaires sur des représentations temps-fréquence proposées récemment. *Traitement du Signal*, 9(1): 3-25, 1992.
- [3] F. AUGER, O. LEMOINE, P. GONCALVES, and P. FLANDRIN. *Time-frequency toolbox for MATLAB, user's guide and reference guide*. <http://tftb.nongnu.org>, 1996.
- [4] A. BELAID and Y. BELAID, *Reconnaissance des formes : méthodes et applications*. InterEditions, 1992.
- [5] T. BLU and J. LEBRUN, Analyse temps-fréquence linéaire II : représentations de type ondelettes. In F. Hlawatsch and F. Auger, editors, *Temps-fréquence, concept et outils*, pages 101-138. Hermès, Traité IC2, 2005.

- [6] B. BOASHASH, Estimating and interpreting the instantaneous frequency of a signal – part I: Fundamentals. *Proc. IEEE*, 80(4): 519-538, 1992.
- [7] B. BOASHASH, B. LOVELL, and L. WHITE, Time frequency analysis and pattern recognition using singular value decomposition of the Wigner-Ville distribution. *Adv. Algorithms Architect. Signal Process., Proc. SPIE 828*, pages 104-114, 1987.
- [8] P. BOLES and B. BOASHASH, The cross Wigner-Ville distribution – a two dimensional analysis method for the processing of vibroseis seismic signals. *Proc. IEEE ICASP 87*, pages 904-907, 1988.
- [9] M. BORN and P. JORDAN, Zur quantenmechanik. *Z. Phys.*, 34: 858-888, 1925.
- [10] A. CHAUDHURI and S. CHAUDHURI, Robust detection of skew in document images. *IEEE Trans. on Image Process.*, 6(2): 344-349, 1997.
- [11] S. CHEN and R. M. HARALICK, An automatic algorithm for text skew estimation in document images using recursive morphological transforms. *Proc. Int. Conf. on Image Processing, Austin, USA*, 1: 139-143, 1994.
- [12] H. CHOI and W. J. WILLIAMS, Improved time-frequency representation of multicomponent signals using exponential kernels. *IEEE Trans. Acoust., Speech, Signal Process.*, 37(6): 862-871, 1989.
- [13] T. CLASSEN and W. MECKLENBRAUKER, The wigner distribution – a tool for time frequency analysis. *Parts I-III, Philips J. Res.*, 35, Part I: No 3, p. 217-250 ; Part II: No 4/5, p. 372-389 ; Part III: No 6, p. 372-389, 1980.
- [14] L. COHEN, Generalized phase-space distribution functions. *J. Math. Phys.*, 7(5): 781-786, mai 1966.
- [15] L. COHEN, Time-frequency distributions – a review. *Proc. IEEE*, 77(7): 941-948, 1989.
- [16] G. CRISTOBAL, J. BESCOS, and J. SANTAMARIA, Application of Wigner distribution for image representation and analysis. *Proc. IEEE Eighth Int. Conf. Pattern Recogn.*, 23: 998-1000, 1986.
- [17] A. K. DAS and B. CHANDA, A fast algorithm for skew detection of document images using morphology. *Int. Journal on Document Analysis and Recognition*, 4(2): 109-114, January 2001.
- [18] B. ESCUDIÉ and J. GRÉA, Sur une formulation générale de la représentation en temps et en fréquence dans l'analyse des signaux d'énergie finie. *CR. Acad. Sci. Paris*, 283: 1049-1051, 1976.
- [19] P. FLANDRIN, *Temps-fréquence*. Hermes, Paris, 1993.
- [20] P. FLANDRIN, *Time-Frequency/Time-Scale Analysis*. Academic Press, San Diego, CA, 1999.
- [21] P. FLANDRIN, N. MARTIN, F. AUGER, C. DEMARS, T. DOLIGEZ, LAMBERT-NEBOUT, J. MARS, F. MOLINARO, J. OVARLEZ, and O. RIOUL, Méthodes temps-fréquence: Fiches synthétiques. *numéro spécial de la revue Traitement du Signal*, 9(1): 77-113, 1992.
- [22] P. FLANDRIN and W. MARTIN, A general class of estimators for the Wigner-Ville spectrum of nonstationary processes. In A. Bensoussan and J.-L. Lions, editors, *Systems Analysis and Optimization of Systems, Lecture Notes in Control and Information Sciences*, volume 62, pages 15-23. Springer, Berlin, 1984.
- [23] R. GRIBONVAL, Analyse temps-fréquence linéaire I: représentations de type Fourier. In F. Hlawatsch and F. Auger, editors, *Temps-fréquence, concept et outils*, pages 69-100. Hermès, Traité IC2, 2005.
- [24] R. HIPPENSTIEL and P. M. D. OLIVIERA, Time varying spectral estimation using the instantaneous power spectrum (ips). *IEEE Trans. Acoust., Speech, Signal Process.*, 38(10): 1752-1759, 1990.
- [25] F. HLAWATSCH and F. AUGER, *Temps-fréquence: concepts et outils*. Hermes, Lavoisier, Paris, 2005.
- [26] F. HLAWATSCH and G. F. BOUDREAUX-BARTELS, Linear and quadratic time-frequency signal representation. *IEEE Signal Process. Mag.*, 9(2): 21-67, 1992.
- [27] S. INGLIS, Lossless document image compression. *Thèse de doctorat, Université de Waikato*, New Zealand, Mars 1999.
- [28] H. F. JIANG, C. C. HAN, and K. C. FAN, A fast approach to the detection and correction of skew documents. *Pattern Recognition Letters*, 18: 675-686, 1997.
- [29] J. F. KAISER and R. W. SCHAFER, On the use of the Io-Sinh window for spectrum analysis. *IEEE Trans. on Acoustics, Speech and Signal Process.*, ASSP-28(1), 1980.
- [30] E. KAVALLIERATOU, N. FAKOTAKIS, and G. KOKKINAKIS, Skew angle estimation in document processing using Cohen's class distributions. *Pattern Recogn. Lett.*, 20: 11-13, 1999.
- [31] E. KAVALLIERATOU, N. FAKOTAKIS, and G. KOKKINAKIS, Skew angle estimation for printed and handwritten documents using the Wigner-Ville distribution. *Image and Vision Computing*, 20: 813-824, 2002.
- [32] O. KENNY and B. BOASHASH, An optical signal processing for time-frequency signal analysis using the Wigner-Ville distribution. *J. Elec. Electron Eng.*, pages 152-158, 1988.
- [33] H. MARGENAU and R. W. HILL, Correlation between measurements in quantum theory. *Proc. Theor. Phys.*, 26: 772-738, 1961.
- [34] N. OUWAYED and A. BELAID, Multioriented text line extraction from handwritten arabic documents. *The Eighth IAPR Workshop on Document Analysis Systems (DAS 2008)*, pages 339-346, 2008.
- [35] N. OUWAYED and A. BELAID, Une approche générale pour l'extraction des lignes des documents arabes anciens multiorientés. In *12^{ème} Colloque International sur le Document Électronique (CIDE.12)*, Canada, 10 2009.
- [36] N. OUWAYED, A. BELAID, and F. AUGER, Cohens class distributions for skew angle estimation in noisy ancient arabic documents. In *10th Int. Conf. on Document Analysis and Recognition (ICDAR'2009) – Third Workshop on Analytics for Noisy Unstructured Text Data (AND'2009)*, Spain, 07 2009.
- [37] J.-P. OVARLEZ, P. GONCALVÈS, and R. BARANIUK, Analyse temps-fréquence quadratique III: la classe affine et autres classes covariantes. In F. Hlawatsch and F. Auger, editors, *Temps-fréquence, concept et outils*, pages 201-236. Hermès, Traité IC2, 2005.
- [38] U. PAL and B. B. CHAUDHURI, An improved document skew angle estimation technique. *Pattern Recognition Letters*, 17: 899-904, 1996.
- [39] U. PAL and B. B. CHAUDHURI, Skew angle detection of digitized indian script documents. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(2): 182-186, 1997.
- [40] T. PAVLIDIS and J. ZHOU, Page segmentation and classification. *Computer Vision Graphics and Image Processing*, 54(2): 484-496, 1992.
- [41] G. PEAKE and T. TAN, A general algorithm for document skew angle estimation. *IEEE Int. Conf. Image Process.*, 2: 230-233, 1997.
- [42] W. POSTL, Detection of linear oblique structures and skew scan in digitized documents. *Proceedings of the Eighth International Conference on Pattern Recognition, IEEE CS Press, Los Alamitos, CA*, pages 687-689, 1986.
- [43] A. W. RIHACZEK, *Principles of High-Resolution Radar*. Artech House, Norwood, MA, 1992.
- [44] J. SAUVOLA, D. DOERMANN, and M. PIETIKAINEN, Locally adaptive document skew detection. *Proc. SPIE Document Recognition and Retrieval IV*, 3027: 96-108, 1997.
- [45] S. N. SRIHARI and V. GOVINDRAJU, Analysis of textual image using the Hough transform. *Machine Vision Applications*, 2: 141-153, 1989.
- [46] L. STANKOVIC, A method for time-frequency analysis. *IEEE Trans. Signal Process.*, 42(1): 255-229, 1994.
- [47] E. P. WIGNER, On the quantum correction for thermodynamic equilibrium. *Phys. Rev.*, 40: 749-759, 1932.
- [48] P. Y. YIN, Skew detection and block classification of printed documents. *Image Vis. Comput.* 19 (8), pages 567-579, 2001.
- [49] Y. ZHAO, L. E. ATLAS, and R. J. MARKS, The use of cone-shaped kernels for generalized time-frequency representations of nonstationary signals. *IEEE Trans. Acoust., Speech, Signal Process.*, 38(7): 1084-1091, 1990.



Nazih Ouwayed

Nazih Ouwayed a obtenu son diplôme de Maîtrise en informatique de l'Université libanaise, Liban, en 2005 et son Master 2 recherche en images et calculs de l'Université de La Rochelle, France, en 2006. Actuellement, Il est étudiant en doctorat et Ater à l'Université Henri Poincaré. Depuis 2006, il poursuit ses travaux de recherches dans l'équipe READ au laboratoire LORIA-INRIA sur la segmentation des documents manuscrits Arabes. Ses intérêts de recherche portent sur l'estimation d'orientation, l'extraction de lignes, la séparation des lignes chevauchées et connectées et la détection de caractères et de mots.



Abdel Belaïd

Abdel Belaïd a reçu sa thèse de 3^{ème} cycle en Informatique en 1979 et sa thèse d'état, dans la même discipline, en 1987 à l'université Henri Poincaré Nancy I. Après 17 ans comme Chargé de recherche au CNRS, il a intégré l'université Nancy 2 en 2002 comme Professeur. Il a dirigé pendant 7 ans le Master Sciences Cognitives. Il est par ailleurs responsable d'équipe de recherche au LORIA concernant la reconnaissance de l'écriture et l'analyse de documents. Il est co-auteur d'un livre intitulé « Reconnaissance des formes : méthodes et applications » et auteur de plus de 130 articles. Ses thèmes de recherche concernent le traitement d'images, la reconnaissance des formes textuelles en utilisant des modèles stochastiques et structurels.



François Auger

François Auger est né à Saint-Germain-en-Laye (78) en 1966. Il a obtenu un diplôme d'ingénieur en mécanique (option automatique et traitement du signal) de l'École Nationale Supérieure de Mécanique de Nantes en 1988, et un doctorat en traitement du signal de l'École Centrale de Nantes en 1991. Depuis 1993, il est maître de conférences à l'Université de Nantes et effectue la plupart de ses enseignements au département Mesures Physiques de l'IUT de Saint-Nazaire et dans le Master Recherche « Systèmes électroniques et Génie électrique » de Polytech'Nantes. Dans le cadre de son activité de recherche à l'IREENA, il s'intéresse actuellement à la commande des dispositifs d'électronique de puissance, à l'implantation d'algorithmes de traitement du signal ou de contrôle-commande sur des circuits logiques programmables, à l'analyse spectrale et à l'analyse temps-fréquence.

