
Exploitation de l'échelle d'écriture pour améliorer la reconnaissance automatique des textes manuscrits arabe

Edgard Chammas* — Chafic Mokbel* — Laurence Likforman-Sulem****

* University of Balamand, Lebanon

** CNRS-LTCl, Telecom ParisTech, Université Paris-Saclay, France

RÉSUMÉ. Les documents manuscrits arabes présentent des défis spécifiques pour la reconnaissance du fait de la nature de l'écriture cursive et d'autres facteurs, comme la taille de l'écriture. Une des plus grandes bases étiquetées des documents manuscrits arabes, la base de données NIST-OpenHaRT inclut de grandes variabilités dans la taille du texte inter et intra mots et lignes. Nous proposons plusieurs approches pour faire face à ces variations lors des phases d'apprentissage et de reconnaissance. Nos premières expériences montrent que la reconnaissance est largement affectée par la taille d'écriture et offrent ainsi une voie pour l'amélioration des performances en tenant compte de l'échelle d'écriture. En phase de reconnaissance, nous avons redimensionné chaque donnée de test à plusieurs tailles prédéfinies, puis nous avons combiné les scores de reconnaissance associés à chacune des tailles. Cette approche a permis des gains notables de performance. De plus, nous avons intégré des données artificiellement redimensionnées pour adapter les modèles HMM à différentes échelles. Nous avons aussi obtenu des gains de performance par deux méthodes différentes de combinaison (ROVER, treillis) des résultats des modèles adaptés. Nous fournissons des résultats de reconnaissance montrant les avantages de l'exploitation de la taille d'écriture.

MOTS-CLÉS : base de données OpenHaRT, Reconnaissance de l'écriture Arabe, échelle de l'écriture, données synthétiques, Adaptation.

1. Introduction

Les styles des scripteurs caractérisent les documents manuscrits non contraints. La forme des caractères, la taille de l'écriture, les espacements des mots et des lignes varient d'un scripteur à un autre (Lawgali, 2015). Ceci forme un défi pour les systèmes de lecture automatique qui doivent être robustes à cette variabilité. Pour faire face à la variation de la taille de l'écriture, une normalisation de l'image du texte pourrait être effectuée lors d'une phase de pré-traitement (España-Boquera *et al.*, 2010) (Lawgali, 2014). Cependant, nous croyons qu'en essayant de normaliser le texte manuscrit, nous risquons de perdre des informations précieuses et importantes pour la reconnaissance (en particulier pour l'écriture arabe). Par conséquent, nous ne cherchons pas à normaliser l'image du texte. Au lieu de cela, nous étudions la façon dont la taille d'écriture affecte généralement les performances de reconnaissance de notre système à base de HMM.

Compte tenu des variations dans la taille de l'écriture, un système de reconnaissance doit être entraîné sur un grand nombre de textes à différentes échelles. Avec l'absence de données suffisantes pour effectuer cet apprentissage, des données synthétiques pourraient être générées afin de couvrir le spectre des tailles d'écriture dans l'ensemble d'apprentissage. Dans (Varga *et al.*, 2003), des données d'apprentissage synthétiques sont générées pour améliorer la performance d'un système de reconnaissance à base de HMM pour la reconnaissance des images de lignes de texte. De même les auteurs dans (Cano *et al.*, 2002) (Mori *et al.*, 2000) rapportent une amélioration notable de la reconnaissance de caractères isolés en augmentant l'ensemble d'apprentissage par des données synthétiques. Des transformations similaires et des méthodes de perturbation ont également été appliquées avec succès au niveau de la reconnaissance de caractères (Simard *et al.*, 1993) (Thien *et al.*, 1997).

Nous avons essayé d'exploiter l'échelle d'écriture au niveau des lignes de texte tout en admettant qu'il est difficile de quantifier avec précision cette échelle à partir de l'image de la ligne de texte (Slimane *et al.*, 2013). Pour ce faire deux approches sont proposées. Une première approche consiste à ne pas changer le modèle HMM appris sur les données d'apprentissage à disposition et de transformer l'image de la ligne de texte à reconnaître en la multipliant par divers facteurs d'échelle. Les images obtenues ainsi sont ensuite reconnues par le modèle initial et les différents résultats sont ensuite combinés. Dans la section 4, nous montrons comment, par cette approche, nous pourrions idéalement améliorer le taux de reconnaissance absolu de notre système HMM de près de 15%. Dans la section 5, nous montrons que la contribution à la performance en reconnaissance provient principalement de la composante verticale de l'échelle de l'écriture. Sur cette base, nous avons développé un algorithme pour mesurer la taille de l'écriture pour une ligne de texte

donnée (section 6). Dans la seconde approche, nous avons utilisé la mesure d'échelle pour classer les données d'apprentissage en trois classes en fonction de leurs tailles respectives. En partant du modèle de reconnaissance initial, trois modèles ont été construits par adaptation, un modèle pour chaque classe d'échelle. Dans la section 7, nous montrons les résultats de reconnaissance pour les différents modèles ainsi que la méthode proposée pour la combinaison de leurs scores.

2. Système de reconnaissance

Notre système de reconnaissance de base a été utilisé dans l'évaluation NIST-OpenHART en 2013 (NIST, 2013). Il est basé sur les Modèles de Markov Cachés (HMMs) (Morillot *et al.*, 2013). Le choix des paramètres de système est spécifique à l'ensemble de données NIST-OpenHART, et basée sur des expérimentations préalables. Nous adoptons la topologie Bakis pour modéliser les lettres avec transition d'état de droite à gauche, correspondant à la direction de l'écriture pour la langue arabe. Deux topologies de modèles différentes ont été utilisées en fonction des lettres. Des modèles de 5 états sont utilisés pour les petites lettres et les ponctuations tandis que des modèles de 8 états sont utilisés pour les grandes lettres. A chaque état est associé un mélange de distributions de probabilités gaussiennes. Des mélanges avec 48 densités de probabilité gaussiennes ont été utilisés dans les expériences.

Des modèles de lettres dépendant du contexte ont été construits (Bianne-Bernard *et al.*, 2011) en considérant les contextes gauche et droit de chaque lettre. Ce choix de modélisation permet de considérer la variation des formes de lettres dans différents contextes. Cependant, la modélisation contextuelle présente quelques défis. Le nombre de modèles devient énorme quand le contexte est considéré. Pour résoudre ce problème, nous réduisons le nombre de paramètres HMM en les regroupant par état selon un arbre de décision. Couvrant l'ensemble de tous lettres arabes et leurs variations, 181 modèles de lettres indépendants du contexte ont été utilisés pour initialiser le processus. Cela comprend toutes les variations des lettres arabes, mais aussi certains signes de ponctuation et les caractères latins.

Des vecteurs de paramètres de 40 coefficients (20 paramètres géométriques et statistiques (El-Hajj *et al.*, 2005) augmentés par leurs dérivées premières) sont extraits par une approche à fenêtre glissante sur l'image de chaque ligne de texte (de droite à gauche). Les coefficients géométriques sont liés aux nombres de pixels, à des valeurs de concavité dans les différentes cellules de la fenêtre et à la position de la ligne de base de l'écriture et de la position moyenne des pixels par rapport à cette ligne de référence. Les coefficients statistiques reflètent la densité des pixels dans des différentes cellules et directions (El-Hajj *et al.*, 2005). Nous avons utilisé Baum-Welch pour estimer les paramètres de chaque modèle.

3. Ensemble de données

Nous avons utilisé l'ensemble de données "NIST Phase 1 Training" de la base NIST-OpenHaRT (146005 images de lignes de texte) dans toutes les expériences présentées dans cet article. Le logiciel HTK a été utilisé à la fois pour l'apprentissage et la reconnaissance (Young *et al.*, 2006). L'apprentissage est effectué sur ~ 94% de l'ensemble de données, à savoir 136529 images de lignes de texte.

85 images des 6% restants de la base ont été considérées pour le test. Elles représentent 1212 images de lignes de texte. Cette découpe de la base de données vient répondre aux contraintes de la puissance de calcul disponible.

4. Exploitation de l'échelle d'écriture

Afin de mesurer l'impact de l'échelle d'écriture une première expérience a été effectuée en redimensionnant les données originales (échelle 1.0) de test avec 6 facteurs d'échelle différents (0.7, 0.8, 0.9, 1.1, 1.2, 1.3). Le choix de ces facteurs d'échelle qui pavent l'intervalle de 0.7 à 1.3 est arbitraire et vise seulement à identifier l'importance de l'échelle de mesure. Le tableau 1 donne les performances obtenues pour chaque échelle ainsi que le pourcentage de fois qu'une échelle produit le meilleur résultat de reconnaissance et le pourcentage des fois qu'elle produit la meilleure reconnaissance seule. Le meilleur résultat s'entend comme le meilleur taux de reconnaissance au niveau ligne: il est égal à 1-taux d'erreur, où le taux erreur est le taux d'erreur mot incluant les insertions, substitutions et destructions. Ainsi un même bon taux de reconnaissance peut être obtenu à différentes échelles. Parfois une échelle produit un taux de reconnaissance meilleur que tous les autres. Nous distinguons dans les colonnes 2 et 3 du Tableau 1 ces deux cas de figure.

Nous n'avons pas intégré de modèle de langage dans le système HMM de reconnaissance afin d'évaluer l'effet de la taille d'écriture sur le modèle optique seulement. D'après le tableau 1, on constate que certaines échelles produisent une meilleure reconnaissance que l'image d'origine (qui correspond à l'échelle 1.0), ouvrant la porte à des améliorations possibles. En outre, en combinant les résultats obtenus à des échelles différentes, nous pourrions idéalement, en choisissant manuellement l'échelle qui produit la meilleure reconnaissance, améliorer la performance en passant d'un taux de bonne reconnaissance (1 – taux d'erreur) de 44.05% à 59.10% ; ce qui équivaut à une amélioration significative de presque 15% en absolu du taux de bonne reconnaissance du système.

Facteur d'échelle	% de fois où la meilleure reconnaissance est obtenue à cette échelle	% de fois que la reconnaissance à cette échelle est meilleure que toutes les autres	Performance du système de reconnaissance (1 – taux d'erreur)
0.7	41%	7%	40.68%
0.8	46%	5%	43.64%
0.9	50%	5%	44.70%
1.0	49%	4%	44.05%
1.1	44%	3%	41.95%
1.2	40%	3%	39.24%
1.3	35%	3%	36.55%

Tableau 1. Tableau montrant le pourcentage de fois qu'une certaine échelle a été parmi celles qui ont produit la meilleure reconnaissance (seconde colonne), le pourcentage de fois une certaine échelle a atteint seule la meilleure reconnaissance parmi toutes les autres échelles (troisième colonne) et la performance du système (quatrième colonne)

Nous avons essayé l'approche "Recognizer Output Voting Error Reduction" (ROVER) (Fiscus *et al.*, 1997) pour combiner les résultats de reconnaissance sur les 7 images redimensionnées par les facteurs d'échelle. Ceci a résulté en un taux de reconnaissance de 48,94% sur les données de test, ce qui représente une augmentation de 4.89% en valeur absolue par rapport à la performance de base.

5. Composant d'échelle verticale

Nous voulions évaluer l'apport du redimensionnement d'échelle selon chacune des directions, horizontale et verticale, à l'amélioration des performances de reconnaissance. Pour ce faire des expériences similaires à (Mori *et al.*, 2000) ont été conduites. Cependant, les images ont été redimensionnées seulement dans la direction verticale.

En considérant les résultats donnés au tableau 2, la combinaison des scores des différentes échelles pourrait idéalement conduire à un taux de bonne reconnaissance de 57,47%. En utilisant ROVER pour la combinaison des scores obtenus le taux de reconnaissance atteint 47,57%. Il est important de rappeler que le résultat obtenu par

ROVER en combinant les scores à différents facteurs d'échelle est légèrement meilleur à 48.94%. Par conséquent, il est raisonnable de conclure que le redimensionnement vertical est le plus bénéfique pour améliorer les performances de reconnaissance.

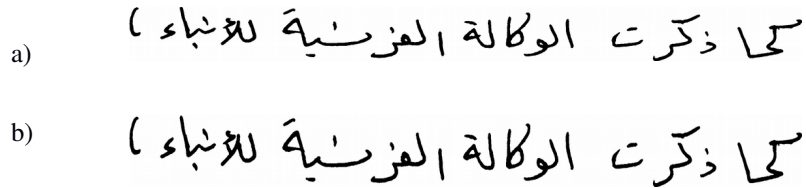


Figure 1. L'image originale de ligne de texte (a) et sa version qui est verticalement élargie à échelle 1.3 (b).

Facteur d'échelle de redimensionnement vertical	% des fois où la meilleure reconnaissance obtenue à cette échelle	% des fois la reconnaissance à cette échelle meilleure qu'à toutes les autres	Performance du système de reconnaissance (1 – taux d'erreur)
0.7	42%	3%	38.94%
0.8	51%	4%	43.07%
0.9	54%	3%	44.75%
1.0	51%	2%	44.05%
1.1	53%	3%	44.79%
1.2	51%	3%	44.14%
1.3	49%	3%	43.12%

Tableau 2. Pour les images mises à l'échelle uniquement dans la direction verticale, le tableau montre le pourcentage de fois où une certaine échelle a été parmi celles qui ont produit la meilleure reconnaissance (seconde colonne), le pourcentage des fois où une certaine échelle a atteint seule la meilleure reconnaissance parmi toutes les autres échelles (troisième colonne) et la performance du système (quatrième colonne)

Du point de vue de l'extraction des paramètres, nous utilisons une approche de fenêtre glissante où la largeur de la fenêtre est de 8 pixels et la hauteur égale à la hauteur de l'image de ligne de texte. Il est important de noter que, redimensionner l'image dans la direction horizontale est équivalent à la modification de la largeur de la fenêtre. En contrepartie, la variation de la taille de l'image dans la direction verticale (figure 1) produit un effet différent. Par conséquent, nous avons poursuivi nos expériences en s'intéressant uniquement à l'échelle verticale de l'écriture.

6. Mesure de l'échelle

Les échelles horizontale et verticale de l'écriture sont connues pour être interdépendantes (Pick *et al.*, 1983) (Van Galen *et al.*, 1998) (Burton *et al.*, 1990). Nous pouvons faire l'hypothèse ici que l'étirement vertical du texte offrirait un bon indice pour faire des hypothèses sur l'échelle de l'écriture. Nous proposons ainsi une méthode pour mesurer l'échelle de l'écriture. On se base pour cela que la longueur de l'écriture, c'est-à-dire en suivant les lignes de cette écriture, est plus grande lorsque l'échelle est plus grande, et vice versa. Pour déterminer cette longueur, on propose de déterminer la squelette du texte manuscrit et calculer sa longueur. Néanmoins, la longueur obtenue dépend aussi du contenu du texte. On propose ainsi de normaliser en divisant par la largeur de l'image du texte après avoir supprimé les espaces. Cette normalisation n'est pas parfaite car la largeur du texte varie aussi avec l'échelle de l'écriture. Néanmoins, la variation est moins importante que la variation de la longueur du texte. En définitif, la mesure proposée nous permet d'affecter chaque ligne de texte à l'une des trois classes d'échelle (Petite, Moyenne, Grande). Le nombre de classes d'échelle peut varier et le choix de trois classes prend en considération à la fois la non précision de la mesure d'échelle et la quantité de données relativement limitée disponible pour nos expériences.

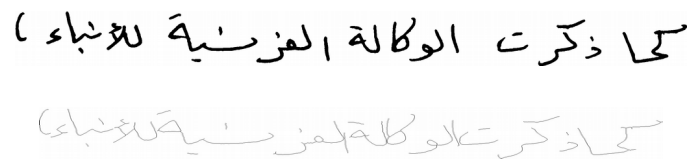


Figure 2. L'image d'une ligne de texte et de son squelette où les espaces ont été supprimés.

Ainsi, les étapes suivies pour mesurer l'échelle d'écriture pour une image de texte en ligne sont les suivantes:

1. Binariser l'image de la ligne de texte (Otsu, 1979)
2. Déterminer le squelette du texte par amincissement "image thinning" (Guo *et al.*, 1989)
3. Calculer la longueur du squelette
4. Réduire la taille de l'image squelette en supprimant les espaces entre les composantes connexes (Figure 2)
5. Produire la mesure de l'échelle m_{ech} comme le rapport de la longueur du squelette divisée par la largeur de l'image du texte non espacé

7. Adaptation d'échelle

Grâce à notre algorithme, nous mesurons l'échelle pour chaque ligne de texte dans notre ensemble d'apprentissage ce qui permet de les classer en 3 classes (Petite, Moyenne et Grande) à l'aide de l'algorithme de classification non supervisée "natural breaks optimization" de Jenks (Jenks, 1967). Les 3 classes sont définies en respectant les critères suivants:

$$Echelle = Petite \quad si \quad m_{ech} \leq 2.043$$

$$Echelle = Moyenne \quad if \quad 2.043 < m_{ech} \leq 2.432$$

$$Echelle = Grande \quad if \quad 2.432 < m_{ech}$$

Ensuite, le modèle de reconnaissance initial a été adapté en utilisant les données classifiées pour produire trois modèles HMM (modèle pour écriture petite, modèle pour écriture moyenne, modèle pour écriture grande).

Parce que les données d'apprentissage ont été réparties sur les 3 classes pour réapprendre les 3 modèles correspondants, le volume de données par modèle est devenu réduit. Pour y remédier nous avons étendu l'ensemble de données d'apprentissage de chaque classe par l'ajout de données synthétiques résultant du redimensionnement des données des autres classes. Par exemple, nous réduisons les grandes images et élargissons les petites images pour étendre le nombre d'images de taille moyenne. Ou nous réduisons les moyennes et grandes images (d'un facteur prédéfini pour chaque classe), afin d'étendre l'ensemble des petites images, etc.

Pour calculer les facteurs avec lesquels les images d'une certaine classe seront réduites ou élargies, la mesure d'échelle moyenne est calculée sur les données.

Classe d'échelle	Moyenne de l'échelle mesurée
Petite	1.89
Moyenne	2.22
Grande	2.64

Tableau 3. La moyenne de l'échelle mesurée calculée pour chaque classe d'échelle calculée à partir de l'ensemble des données d'apprentissage classifiées suivant leur taille.

Pour transformer une image I de classe X en une autre image J de classe Y , nous redimensionnons I par le facteur suivant:

$$\frac{E(X)}{E(Y)}$$

où $E(X)$ et $E(Y)$ sont les valeurs moyennes des mesures d'échelle pour la classe X et la classe Y respectivement. Une fois les modèles adaptés, des expériences de reconnaissance ont été conduites. Les résultats obtenus sur les données de test sont indiqués dans le tableau 3.

Modèle	% de fois la où meilleure reconnaissance obtenue avec ce modèle	% de fois la reconnaissance avec ce modèle est meilleure qu'avec tous les autres	Performance du système de reconnaissance (1 – taux d'erreur)
Petit	61%	7%	41.66%
Moyen	67%	7%	43.60%
Grand	58%	8%	40.11%
Initial	68%	7%	44.05%

Tableau 4. Tableau montrant le pourcentage de fois où le modèle d'une certaine échelle a été parmi celles qui ont produit la meilleure reconnaissance (seconde

colonne), le pourcentage de fois où le modèle d'une certaine échelle a atteint seul la meilleure reconnaissance parmi tous les autres modèles (troisième colonne) et la performance du système (quatrième colonne)

La combinaison des résultats des différents modèles d'échelle pourrait idéalement atteindre une performance de 52.21%; ce qui équivaut à un gain de 8.16% en taux de reconnaissance par rapport à la performance initiale.

8. Combinaison des treillis

Dans l'expérience précédente (Section 7), nous avons construit 4 systèmes de reconnaissance associés à des échelles différentes. Nous proposons ici une combinaison de ces systèmes basée sur les treillis de reconnaissance. Pour un système donné, nous calculons les N-meilleures résultats de reconnaissance (avec $N = 5$) et nous générons un treillis pour représenter ces résultats. Les treillis des différents systèmes sont ensuite combinés dans un seul réseau avant d'effectuer un décodage de Viterbi. Cette combinaison par treillis permet d'atteindre une performance de 45.89%; ce qui équivaut à un gain de 1.84% en taux de reconnaissance par rapport à la performance initiale.

9. Conclusions et perspectives

Dans cet article, nous avons démontré l'effet de la taille de l'écriture sur la performance du système de reconnaissance. Nous avons mis en évidence qu'une exploitation réussie de l'échelle de l'écriture pourrait apporter des gains en performance et ouvre une porte à l'amélioration. Deux approches sont proposées, la première considère l'échelle d'écriture à la phase de reconnaissance et la seconde la considère lors de l'apprentissage tout en intégrant avec succès les données synthétiques pour fournir plus d'informations au système. Les deux approches produisent des gains notables. En intégrant la notion d'échelle lors de la reconnaissance, une combinaison idéale conduirait à une augmentation de 15% de la performance pour notre système. Nous avons appliqué la combinaison ROVER et obtenu un gain en taux de reconnaissance absolu de 4.89% par rapport à la performance originale. Un algorithme pour mesurer la taille de l'écriture est aussi proposé et utilisé pour classifier les données en 3 classes d'échelle ce qui permet d'adapter 3 modèles spécifiques par échelle d'écriture. L'expérience a montré qu'une combinaison idéale conduirait à un gain de 8.16% en performance. Pour cela, une nouvelle méthode de combinaison à base de treillis a été proposée. Elle permet un gain de 1.84% par rapport à la performance initiale.

L'écart dans les gains de performance entre les expériences menées dans le cadre de l'approche de considération de l'échelle à la phase de reconnaissance et celles

réalisées lors de la considération de l'échelle à la phase d'apprentissage pourrait être dû au fait que nous avons juste utilisé une seule itération pour adapter nos nouveaux modèles. Cela pourrait aussi être lié à l'algorithme utilisé pour la mesure de la taille de l'écriture. Nous poursuivrons l'évaluation de la fiabilité de cet algorithme et rechercherons des améliorations possibles. Il est important aussi de se pencher sur d'autres méthodes de combinaison de score pour mieux exploiter les avantages de la taille d'écriture.

Enfin, après avoir travaillé au niveau de ligne de texte, nous envisageons d'étendre notre recherche pour exploiter l'échelle de l'écriture à des niveaux plus fins (mot, lettre, vecteur de paramètres, etc.).

10. Bibliographie

- Bianne-Bernard A-L., Menasri F., El-Hajj R., Mokbel C., Kermorvant C., Likforman-Sulem L., « Dynamic and Contextual Information in HMM Modeling for Handwritten Word Recognition », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No 10, pp. 2066-2080, 2011.
- Burton A.W. , Pick H.L., Holmesand C., Teulings H.L., (1990). « The independence of horizontal and vertical dimensions in handwriting with and without vision », *Acta Psychologica*, 75, 201-212, (1990).
- Cano J., Perez-Cortes J-C., Arlandis J., Llobet R., « Training Set Expansion in Handwritten Character Recognition », 9th Int. Workshop on Structural and Syntactic Pattern Recognition, pp. 548-556, 2002.
- El-Hajj R., Likforman-Sulem L., Mokbel C., « Arabic Handwriting Recognition Using Baseline Dependant Features and Hidden Markov Modeling », 8th International Conference on Document Analysis and Recognition, 2005.
- Espana-Boquera S., Castro-Bleda M.J., Gorbe-Moya J., Zamora-Martinez F., « Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN Models », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010.
- Fiscus J., « A post-processing system to yield reduced word error rates: Recognizer Output Voting Error Reduction (ROVER) », *IEEE Workshop on Automatic Speech Recognition Understanding*, pp. 347– 354, 1997.
- Guo Z., Hall R.W., « Parallel thinning with two-subiteration algorithms ». *Comm. ACM*, 32:359-373, January 1989.
- Jenks G. F., « The Data Model Concept in Statistical Mapping », *International Yearbook of Cartography* 7: 186–190. 1967.
- Lawgali A., « A Survey on Arabic Character Recognition », *International Journal of Signal Processing, Image Processing and Pattern Recognition*, Vol. 8, No. 2, pp. 401-426, 2015.
- Lawgali A., M. Angelova, A. Bouridane, « A Framework for Arabic Handwritten Recognition Based on Segmentation », *International Journal of Hybrid Information Technology*, Vol.7, No.5, pp.413-428, 2014.

- Morillot O., Oprean C., Likforman-Sulem L., Mokbel C., Chammas E., Grosicki E., « The UOB-Telecom ParisTech Arabic Handwriting Recognition and Translation Systems for the OpenHaRT 2013 Competition », NIST OpenHaRT 2013.
- Mori M., Suzuki A., Shio A., Ohtsuka S., « Generating New Samples from Handwritten Numerals Based on Point Correspondence », 7th Int. Workshop on Frontiers in Handwriting Recognition (IWFHR 2000), pp. 281-290, Sept 2000.
- NIST, NIST 2013 Open Handwriting Recognition and Translation Evaluation Plan - version 1.7, pp. 1-9, 2013. [Online]. Available: <http://www.nist.gov/itl/iad/mig/hart2013.cfm>
- Otsu N., « A threshold selection method from gray-level histograms ». IEEE Transactions on Systems, Man, and Cybernetics. 9 (1): 62-66, 1979.
- Pick H.L. Jr., Teulings H.L., « Geometric transformations of handwriting as a function of instruction and feedback ». Acta Psychologica 54, 327-340, 1983.
- Simard P., Cun Y. L., Denker J., « Efficient Pattern Recognition Using a New Transformation Distance », Advances in Neural Information Processing Systems, S. Hanson, J. Cowan, and C. Giles, eds., vol. 5, pp. 50-58, 1993.
- Slimane F., Kanoun S., Abed H. El, Alimi A. M., Ingold R., Hennebert J., ICDAR 2013 Competition on Multi-font and Multi-size Digitally Represented Arabic Text, 12th International Conference on Document Analysis and Recognition, 2013.
- Thien M. Ha, Bunke H., « Off-Line, Handwritten Numeral Recognition by Perturbation Method », IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(5):535-539, May 1997.
- Van Galen G.P., Weber J., « On-line size control in handwriting demonstrates the continuous nature of motor programs ». Acta Psychologica. 100, 195-216, (1998).
- Varga T., Bunke H., « Generation of Synthetic Training Data for an HMM-based Handwriting Recognition System », 7th International Conference on Document Analysis and Recognition, p. 618, 2003.
- Young S.J., Evermann G., Gales M.J.F., Hain T., Kershaw D., Moore G., Odell J., Ollason D., Povey D., Valtchev V., Woodland P.C., « The HTK Book, Version 3.4 », Cambridge Univ. Eng. Dept., 2006.