



**HAL**  
open science

## Classification des k-ppv par sous-voisinages emboîtés

Bruno Taconet, Abderrazak Zahour, Saïd Ramdane, Wafa Boussellaa

► **To cite this version:**

Bruno Taconet, Abderrazak Zahour, Saïd Ramdane, Wafa Boussellaa. Classification des k-ppv par sous-voisinages emboîtés. Sep 2006, pp.145-150. hal-00113585

**HAL Id: hal-00113585**

**<https://hal.science/hal-00113585>**

Submitted on 13 Nov 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Classification des k-ppv par sous-voisinages emboîtés

Bruno Taconet<sup>1</sup> – Abderrazak Zahour<sup>1</sup> – Saïd Ramdane<sup>1</sup> – Wafa Boussellaa<sup>2</sup>

<sup>1</sup> Equipe GED – Université du Havre  
Place R. Schuman 76610 Le Havre

<sup>2</sup> REsearch Group on Intelligent Machines (REGIM)  
Université de Sfax, ENIS, DGE  
BP. W-3038 - Sfax – Tunis

<sup>1</sup> bruno.taconet[zahour][ramdane]@univ-lehavre.fr

<sup>2</sup> Wafa.boussellaa@gmail.com

**Résumé :** *La méthode des plus proches voisins est une méthode de classification géométrique très utilisée en reconnaissance de formes, en raison de sa simplicité et de sa robustesse. Les caractéristiques sont exploitées dans un espace métrique de représentation, généralement  $R^n$  muni de la distance euclidienne. La méthode de Keller repose sur une fonction de décision floue qui met en jeu les distances des prototypes au point inconnu. Nous proposons une nouvelle règle de décision floue dans laquelle la contribution d'un prototype ne se fait pas individuellement, mais collectivement, en considérant son sous-voisinage, selon la règle : un prototype est d'autant plus influent qu'il est plus proche du point inconnu, et qu'il existe des prototypes de la même classe plus proche que lui. Les essais de validation sont menés en comparaison notamment avec la méthode de Keller (avec initialisation nette) sur deux banques de données standard : la base des 150 iris de Fischer en auto-validation dans l'espace de représentation de dimension 4, et la base mnist de chiffres manuscrits composée d'une base d'apprentissage de 60 000 caractères et d'une base de test de 10 000 dans l'espace de représentation de dimension  $28 \times 28$ . Cependant, l'exemple de la base mnist montre que le gain reste faible, en comparaison des méthodes qui exploitent les invariances des prototypes par transformations géométriques ou par combinaison linéaire des voisins.*

**Mots-clés :** classification, k plus proches voisins, flou, Keller, iris de Fischer, mnist.

## 1 Introduction

Toute méthode géométrique de reconnaissance de formes nécessite de disposer d'un espace métrique de représentation des caractéristiques. Le choix se porte souvent en pratique sur  $R^n$  muni de la distance euclidienne. Les dimensions de cet espace peuvent être réduites par une analyse en composantes principales, laquelle peut conduire en outre à un changement d'échelle, de façon à ajuster les écarts-types selon chaque

direction. Nous n'aborderons pas ici ces questions d'analyse et de prétraitement des données, qui forment un sujet d'étude complémentaire. L'espace métrique  $R^n$  muni de la distance euclidienne est isotrope : la distance euclidienne entre deux points se conserve pour toute translation et toute rotation dans l'espace des caractéristiques. De plus, elle est naturelle au sens de la vue, nous permettant d'évaluer les distances dans un plan. La densité de probabilité, qui a la dimension d'une probabilité par unité de volume, est une fonction de chaque point de l'espace de représentation. Décider si un point appartient à une classe équivaut à retenir la classe qui a la plus forte densité de probabilité en ce point. Le problème posé en terme probabiliste est donc de trouver un estimateur fiable de la densité de probabilité relative de chaque classe en tout point de l'espace de représentation accessible par un test. Dans une optique "floue", c'est le degré d'appartenance à une classe, qui dépend des grandeurs géométriques de l'espace de représentation et de paramètres à régler, permet de décider. Aussi la fonction qui exprime le degré d'appartenance peut être regardée comme une densité de probabilité à une constante multiplicative près.

Pour estimer la densité de probabilité, on peut recourir à des méthodes globales paramétriques : on se donne une forme analytique globale ou semi-globale de la distribution, et on en estime les paramètres à l'aide de statistiques. Ainsi une distribution de probabilité peut-elle être par exemple gaussienne ou de Rayleigh, pure ou mélangée. A l'inverse, on peut considérer que la densité de probabilité est une fonction purement locale qui ne dépend que d'un voisinage du point où on cherche à l'évaluer. Dans ce cas, seuls quelques prototypes les plus proches du point inconnu seront impliqués dans la décision. En raison de l'incomplétude des données, il est courant de recourir à une formalisation floue : l'estimation de la densité de probabilité locale est remplacée le calcul du degré d'appartenance à une classe.

La méthode que nous proposons est une méthode de classification, apparentée aux k plus proches voisins, de type floue. Nous la situerons au sein de la famille de ces

méthodes, puis nous présenterons les conditions expérimentales de sa validation.

## 2 Objectifs des méthodes des k-ppv

La méthode du plus proche voisin [Fix 51][Cover 67] est la méthode locale non paramétrique la plus simple. On suppose seulement que le voisinage qui influe sur la décision de classification du point inconnu se limite à une boule contenant seulement le plus proche voisin. La densité de probabilité de toutes les classes est nulle, à l'exception de celle affectée au plus proche voisin qui est naturellement choisie pour la décision. Cette méthode en apparence très simple est robuste. Son erreur est majorée par 2 fois l'erreur minimale obtenue par la classification bayésienne, dans les conditions asymptotiques. Elle permet de reconnaître les formes lorsque la distribution des classes n'est pas convexe. L'algorithme de base est facile à écrire et à programmer. En contrepartie, elle nécessite un grand volume de données d'apprentissage, ce qui veut dire un grand nombre de points d'apprentissage à stocker et à examiner, autrement dit des ressources mémoire élevées et un temps d'exécution important dans une implémentation brute de l'algorithme.

Depuis plusieurs décennies, les chercheurs ont essayé d'améliorer la méthode, d'une part en réduisant le volume des données nécessaires et en optimisant les algorithmes pour réduire le temps de traitement (condensing), d'autre part en étendant et en enrichissant le contexte local pour affiner la décision (editing rules). En effet, par exemple, est-il opportun de choisir le premier plus proche voisin si le deuxième et le troisième plus proche voisin appartiennent à une autre classe et sont quasiment à la même distance que le premier plus proche voisin ? Par ailleurs, l'erreur ou l'incertitude sur la mesure de chaque prototype, le nombre insuffisant de points appris, les erreurs d'étiquetage du professeur, sont autant de causes qui perturbent les résultats.

Des solutions pour améliorer le taux reconnaissance ont été explorées dans deux directions complémentaires : enrichir l'analyse du voisinage et affiner le critère de décision, corriger les données d'apprentissage.

1. Etendre le voisinage à plusieurs points, dans le but d'affiner la décision, ce qui implique
  - (a) l'existence d'un critère ou d'une fonction de décision (editing rule) : vote majoritaire, pondération par rapport de distances de Dudani [DUD 76], degré d'appartenance flou de Keller[Kel85], évidence de Donooux [DON 95], voisinage par classe de Hattori [HAT 99], paramétrique de Arif[ARI 05].
  - (b) le choix du nombre de voisins  $k$ , ou l'exploitation de plusieurs valeurs de  $k$ .
2. Corriger les données d'apprentissage
  - (a) par lissage flou des étiquettes, afin de filtrer les incertitudes Keller [KEL 85],
  - (b) par modification de l'ensemble d'apprentissage, en éliminant les points jugés aberrants et/ou en

adjoignant des points jugés manquants, dans le but de corriger les erreurs et/ou omissions du professeur. En particulier, l'exploitation des invariants lors de transformations géométriques a débouché sur la méthode de la distance tangente [SIM 98], et la méthode de l'hyperplan [VIN 01], façons élégantes et très efficaces d'adjoindre implicitement un grand nombre de prototypes pour chaque prototype.

Cet article propose une nouvelle règle de décision (editing rule) dérivée de celle de Keller, et se rattache essentiellement au premier point cité au-dessus. Seule l'introduction de la distance angulaire se rattache au second point. Afin de mieux situer notre apport, nous rappelons d'abord les règles de décision classiques.

## 3 Principales règles de décision

La méthode du voteur majoritaire présuppose l'hypothèse d'une densité de probabilité uniforme à l'intérieur du  $k$ -voisinage, et une valeur de  $k$  suffisamment élevée pour que la densité estimée soit stable. En pratique, il est rare que ces deux conditions soient remplies : en effet, le nombre de points d'apprentissage est non seulement fini, mais encore il est souhaitable qu'il soit limité pour des raisons de taille des données à collecter, à stocker et à exploiter. Dans un contexte de plus faible densité, augmenter le nombre  $k$  entraîne l'accroissement du volume du voisinage, ce qui affaiblit l'hypothèse d'une distribution locale uniforme. La valeur optimale de  $k$  exprime le meilleur compromis qui concilie distribution uniforme et voisinage peu dense. Cependant, lorsque la représentation est clairsemée, il devient plus plausible de supposer que la densité de probabilité n'est pas uniforme dans le  $k$ -voisinage. Logiquement, plus un voisin est proche, plus il influe sur la décision. La méthode de pondération de Dudani [Dud76] semble la première à s'inscrire dans cette optique. La pondération  $w_i$  du point  $x_i$  s'exprime ainsi :

$$w_i = \frac{d(x, x_k) - d(x, x_1)}{d(x, x_k) - d(x, x_1)} \text{ si } d(x, x_k) \neq d(x, x_1),$$

$$w_i = 1 \text{ si } d(x, x_k) = d(x, x_1) \quad (1)$$

où  $x_1$  et  $x_k$  sont respectivement le prototype le plus proche et le plus lointain.

La méthode de Keller [Kel85] propose de calculer pour chaque point un degré d'appartenance du point inconnu, fonction des distances de  $x$  à chaque prototype  $x_j$ , selon la formulation :

$$f_{C_i} = \frac{\sum_j u_{ij} d(x, x_j)^{\frac{2}{1-m}}}{\sum_{i,j} u_{ij} d(x, x_j)^{\frac{2}{1-m}}} \quad (2)$$

où  $u_{ij}$  est le degré d'appartenance du point  $x_j$  à la classe  $C_i$ . Dans le cas d'initialisation nette,  $u_{ij}$  vaut 1 ou 0 selon que le point appris  $x_j$  appartient ou non à la classe  $C_i$ . Dans le cas d'initialisation floue,  $u_{ij}$  est un degré d'appartenance flou, associé à chaque prototype, qui

dépend d'un avis d'expertise, ou du voisinage appris du prototype. Le paramètre  $m$  règle l'influence de la distance.

La méthode de Hattori [Hat99] considère un  $k$ -voisinage pour chaque classe. La fonction de décision dépend de paramètres à ajuster et des points de ce  $k$ -voisinage.

Dans le cadre d'une approche analytique, Arif [Arif 2005] a proposé récemment une formulation du degré d'appartenance du point inconnu  $x$  à la classe  $C_i$  est proportionnel à  $1/(1+1/b_i)$ , avec :

$$b_i = \frac{N_i^\alpha \sqrt{R(x_i)} v(x, x_i)^\gamma}{d(x, x_i)^\beta} \quad (3)$$

où  $x_i$  est le point de la classe  $C_i$  le plus proche de  $x$ ,  $N_i$  est le nombre de  $k$ -voisins appartenant à la classe  $C_i$ ,  $R(x_i)$  est le rang de classement de  $x_i$  dans le  $k$ -voisinage,  $v(x, x_i)$  est une pondération dépendant du voisinage restreint à  $x_i$  et aux points  $x_j$  plus proches que lui, ayant pour expression :

$$v(x, x_i) = \frac{d(x, x_i)}{\sum_{i=1, \dots, i} d(x, x_i)} \quad (4)$$

$\alpha$ ,  $\beta$ ,  $\gamma$  sont des exposants à ajuster en optimisant le taux de reconnaissance. Arif propose finalement :

$\alpha=1$ ,  $\beta=1$ ,  $\gamma$  ajustable, afin de réduire à 1 le nombre de paramètres à ajuster.

Comme on le voit, les méthodes récentes introduisent au moins un paramètre à ajuster en plus du nombre de voisins  $k$ . Nous proposons de modifier la méthode de Keller en remplaçant la distance euclidienne par une distance qui exploite le sous-voisinage de chaque prototype.

## 4 Méthode des sous-voisinages

### 4.1 Principe de la règle de décision

La démarche que nous proposons participe du postulat suivant : dans le cas d'une faible densité, la contribution d'un prototype voisin appartenant à une même classe est d'autant meilleure que :

1. le prototype est proche,
2. il y a plus de prototypes plus proches que lui appartenant à la même classe

La contribution d'un prototype interne au voisinage se fera donc en utilisant une autre distance que la distance euclidienne, en respectant la règle : un prototype interne au sous-voisinage et de la même classe rapproche artificiellement le prototype considéré, d'autant plus que ce prototype interne est proche du point inconnu. Un prototype d'une autre classe n'affecte pas le prototype considéré.

### 4.2 Formulation de la règle de décision

Dans la suite, nous donnerons aux formules une plus grande généralité en attribuant à un prototype  $x_m$  une masse floue  $u_{im}$ , qui dénote le degré d'appartenance de  $x_m$  à la classe  $C_i$ . Dans le cas d'un étiquetage net de la

base d'apprentissage,  $u_{ij}$  vaut 1 si  $x_j$  appartient à la classe  $C_i$ , 0 sinon.

Nous déduisons des principes énoncés précédemment une formule de la distance relative à la classe  $C_i$  définie sur le sous-voisinage  $V(x, x_j)$ , notée  $D_i(x, x_j)$

$$D_i(x, x_j) = \frac{\sum_{x_m \in V(x, x_j)} u_{im} d(x, x_m)}{\text{Card}_{x_m \in C_i, V(x, x_j)}(u_{im}, X_m)} \quad (5)$$

Détaillons les constituants de cette définition de la distance sur le sous-voisinage .

$d(x, x_j)$  est la distance de l'espace métrique, usuellement la distance euclidienne dans  $R^n$ .

$\text{Card}_{x_m \in C_i, V(x, x_j)}(u_{im}, X_m)$  représente la masse des prototypes de la classe  $C_i$  dans le sous-voisinage

$D_i(x, x_j)$  est la moyenne arithmétique des distances des prototypes de la classe  $C_i$ .

Comme il est évident que :

$$d(x, x_m) \leq d(x, x_j) \quad \forall x_m \in V(x, x_j),$$

la moyenne arithmétique des distances vérifie l'inégalité qui caractérise le rapprochement virtuel :

$$d(x, x_j) \leq D_i(x, x_j) \leq d(x, x_j) \quad (6)$$

Ecrivons la formule du degré d'appartenance de  $x$  à la classe  $C_i$  avec le même formalisme que la distance de Keller (cf. relation (3)):

$$f_{C_i} = \frac{\sum_j u_{ij} D_i(x, x_j)^{-\alpha}}{\sum_i \sum_j u_{ij} D_i(x, x_j)^{-\alpha}} \quad (7)$$

Dans cette relation la distance euclidienne, qui traduit l'influence d'un point isolément, a été remplacée par une « distance de sous-voisinage ».  $\alpha$  règle l'influence de la distance, comme  $m$  dans la formule de Keller. La signification de  $\alpha$  est la suivante : si  $\alpha$  est nul, on obtient la règle du voteur majoritaire ; si  $\alpha$  est infini, on obtient la règle du plus proche voisin. En terme de stockage des données, la méthode des sous-voisinages repose, comme la méthode de Keller, sur le tableau des distances ordonnées des plus proches voisins. Seule l'exploitation de ce tableau est légèrement plus complexe car il faut calculer de façon récurrente la distance  $D_i(x, x_j)$ .

Notons enfin que, dans cette méthode, comme la méthode de Keller, le degré d'appartenance à une classe est une fonction monotone du voisinage :

$$V_{k, C_i}^{(1)}(x) \subset V_{k, C_i}^{(2)}(x) \Rightarrow f_{C_i}^{(1)}(x) < f_{C_i}^{(2)}(x) \quad (8)$$

Si, dans un premier voisinage, tous les prototypes sont soit plus proches soit aussi proches du point inconnu que dans un second voisinage, et si les points des autres classes sont invariants, le degré d'appartenance de  $x$  est supérieur dans le cas du premier voisinage. Cela assure la stabilité si l'étiquetage net de l'ensemble d'apprentissage suffit. Remarquons que cette propriété n'est pas vérifiée par exemple pour la méthode de Arif (cf. (3)): plus le rang de classement est élevé, plus le degré d'appartenance augmente. Pour ces méthodes qui ne vérifie pas la monotonie, le réglage des paramètres permet de s'adapter sans ré-étiquetage au jeu

des données d'apprentissage en « leave one out », mais rien ne garantit la stabilité des performances sur d'autres cas.

## 5 Les iris de Fischer

### 5.1 Composition

La base des 150 iris de Fischer [Fis36] {IRI 06} comprend trois variétés de 50 fleurs chacune: Setosa, Versicolor, Virginica. Les quatre caractéristiques retenues pour la classification sont: la longueur des sépales, la largeur des sépales, la longueur des pétales, la largeur des pétales, toutes exprimées en centimètres.

K	Maj	Dud	K1/4	SV1/44	K1/2	SV1/2	K1	SV 1	K 2	SV 2	K 4	SV4
1	6	6	6	6	6	6	6	6	6	6	6	6
2	6	6	6	6	6	6	6	6	6	6	6	6
3	6	6	6	6	6	6	6	6	6	6	6	6
4	6	6	6	6	6	6	6	6	6	6	6	6
5	5	6	5	5	5	5	5	5	5	5	6	6
6	6	6	6	6	6	6	6	6	6	6	6	6
7	5	6	5	5	5	5	5	5	5	5	6	6
8	5	6	5	5	5	5	5	5	6	6	6	6
9	6	6	6	6	6	6	5	5	4	5	6	6
10	5	6	5	5	5	5	5	5	6	6	6	6
11	4	6	4	4	4	4	5	4	4	5	6	6
12	5	6	5	5	5	5	5	5	5	5	6	6
13	5	6	5	5	4	4	5	4	4	6	6	6
14	4	6	4	4	4	4	4	4	5	5	6	6
15	4	6	4	4	4	4	5	4	4	6	6	6
Moy	5.20	6.00	5.20	5.20	5.13	5.13	5.27	5.07	5.20	5.60	6.00	6.00

TAB-1 TABLEAU COMPARATIF DES MAL CLASSEES DE LA BASE IRIS

Les mal classées se trouvent toujours parmi les six mêmes fleurs mal classées par le 1-ppV (colonne majoritaire,  $k=1$ ): 3 Versicolor comme Virginica, et 3 Virginica comme Versicolor. En réalité, 3 fleurs sont toujours mal classées par toutes les méthodes. On constate que les performances des deux méthodes sont proches, avec une préférence pour la méthode du sous-voisinage lorsque la puissance est plus faible, et pour la méthode de Keller (étiquetage net) lorsque la puissance est plus élevée. En effet, un exposant élevé accentue l'influence du plus proche voisin dans la méthode du sous-voisinage ce qui explique son plus mauvais score qui se rapproche de celui des 1-ppV. Mais cela laisse espérer une légère amélioration du score par la méthode du sous-voisinage en comparaison de la méthode de Keller, lorsque l'exposant n'est pas trop fort. Ce n'est qu'une indication à poursuivre l'expérience, car le nombre de mal classés est ici insuffisant pour en tirer une quelconque leçon statistique.

## 5.2 Expérimentation et résultats

Puisque seul l'ensemble d'apprentissage est connu, et que le nombre de données est assez faible, les expériences de reconnaissance sont menées en «leave one out»: 1-ppV, majoritaire, Dudani, Keller (étiquetage net), sous-voisinage (étiquetage net), pour des valeurs de l'exposant  $\alpha$  variant de 1/4 à 4. Le tableau 1 donne le taux de reconnaissance selon la méthode employée.

## 6 La base mnist

### 6.1 Composition

La base mnist mise à disposition librement sur Internet [LEC 06] est composée d'images de chiffres manuscrits de taille normalisée 28x28 pixels en 256 niveaux de gris, obtenue par la réunion de deux bases mnist d'images binaires normalisées 20x20, dont le centre de gravité a été translaté au centre de l'image avec un lissage anti-aliasing. La base mnist comprend un ensemble d'apprentissage de 60 000 chiffres manuscrits et un ensemble de test distinct de 10 000 chiffres manuscrits. Notre but n'est pas d'obtenir un taux de reconnaissance minimal pour valider une méthode de reconnaissance comme étant la meilleure, mais de construire une assez bonne méthode qui fournisse des caractéristiques permettant de comparer la méthode de Keller et la méthode des sous-voisinages.

## 6.2 Prétraitement

Nous avons fait subir un prétraitement à chaque image : un redressement vertical par translation anti-aliasing après détection d'inclinaison de l'axe principal d'inertie, puis un lissage par un filtrage linéaire symétrique  $1/8(0, 1, 0; 1, 2, 1; 0, 1, 0)$ , et enfin une translation des niveaux de gris par soustraction du niveau de gris moyen. Après transformation, un point noir de la pseudo-image  $Z$  ayant pour coordonnées  $(i, j)$  a un pseudo-niveau de gris  $z(i,j)$  négatif ; un point blanc a un pseudo-niveau de gris positif. La masse claire (négative) pèse autant que la masse sombre (positive). Nous avons ainsi construit une banque prétraitée de 60 000 pseudo-images de référence et 10 000 pseudo-images de test.

## 6.3 Distance angulaire

Dans cet espace de dimension  $28 \times 28$ , au lieu d'employer la distance euclidienne, nous avons défini et utilisé une distance basée sur le produit scalaire normalisé entre deux pseudo-images  $Z_1$  et  $Z_2$ .

$$d(Z_1, Z_2) = 2 \sqrt{0.5 \left( 1 - \frac{\sum_i \sum_j z_1(i,j) z_2(i,j)}{\sqrt{\sum_i \sum_j z_1(i,j)^2} \sqrt{\sum_i \sum_j z_2(i,j)^2}} \right)}$$

Lorsqu'elle est petite, cette distance s'interprète géométriquement comme l'angle exprimé en radian entre les deux vecteurs. Dans le cas présent, les vecteurs ont comme composantes les valeurs des cellules des deux pseudo-images dans l'espace de dimension  $28 \times 28$ , car le produit scalaire normalisé calcule le

Euclidienne sans lissage	244
Angulaire après lissage	197

TAB-2. MAL CLASSES PAR 1-ppv : INFLUENCE DU TYPE DE DISTANCE

k-ppv	4	5	6	7	8	9	10
Keller	176	177	177	183	175	184	183
Sous-Vois	174	174	175	178	172	178	177
Ecart	2	3	2	5	3	6	6

TAB-3. . MAL CLASSES : INFLUENCE DE LA METHODE, POUR  $\alpha=2$

Nous constatons une supériorité très légère, mais uniforme, de la méthode des sous-voisinages sur la méthode de Keller. L'emploi de l'une comme l'autre méthode permet d'obtenir un gain significatif (de l'ordre de 0.2 %) par rapport au simple plus proche voisin dans les mêmes conditions, dont les fausses reconnaissances s'élèvent à 197 sur 10 000 (soit 1,97% d'erreur).

Par ailleurs, pour une méthode de traitement simple, sans couche neuronale ni machine à support de vecteurs, les résultats en taux de reconnaissance absolu sont tout à fait honorables, si on les compare avec ceux des méthodes simples. Notamment, l'emploi de la distance angulaire après prétraitement permet un gain

de 0.47% sur le classique 1-ppv avec distance euclidienne. Cependant, le gain reste faible et les performances médiocres en comparaison des méthodes qui exploitent les invariances par transformation géométrique d'un prototype voisin (méthode de la distance tangente : 1,1 % d'erreur), [LEC 98] ou par combinaison linéaire des prototypes voisins (méthode de l'hyperplan ou HKNN : 1,2% d'erreur [VIN 01]). Un tableau comparatif actualisé est disponible à [LEC 06].

## 6.4 Expérimentation et résultats

Afin de bien contrôler la programmation et de gagner un peu de temps dans l'exécution, nous avons programmé en langage C. En effet, il faut calculer 60 000 distances par échantillon de test entre deux images de dimension  $28 \times 28$ . Le temps d'exécution est légèrement supérieur à 1 seconde par échantillon de test, soit plusieurs heures pour tester les 10 000 échantillons. Le tableau 2 montre l'influence de la distance dans la méthode du plus proche voisin : dans le premier cas, la distance euclidienne après redressement sans lissage est appliquée, dans le second cas, la distance angulaire après redressement lissage. Le tableau 3 présente les résultats comparatifs entre la méthode de Keller et notre méthode des sous-voisinages pour un exposant  $\alpha$  de valeur 2, avec la distance angulaire.

## 7 Conclusion

Nous avons présenté une méthode de classification de type k-ppv, proche de la méthode de Keller : la

différence est que la distance considérée entre le point inconnu et un prototype est la moyenne des distances de tous les points du sous-voisinage. Cette méthode demande les mêmes ressources mémoire et quasiment le même temps d'exécution que la méthode de Keller. Confirmant une légère indication donnée par les essais sur la banque de donnée des iris, la méthode des sous-voisinages a fourni des résultats très légèrement supérieurs à ceux de la méthode de Keller sur la totalité de la vaste base mnist. Pour étendre ces premiers résultats, il resterait à expérimenter la méthode avec un ré-étiquetage flou. Cependant, le gain demeure très faible en regard de celui obtenu par l'exploitation des invariances géométriques. Ainsi, au regard de l'expérience menée sur la base mnist, il semble que la méthode floue présentée ici se justifie seulement lorsque les invariances des prototypes par transformations géométriques ne sont pas connues ou ne sont pas exploitées volontairement.

Document Recognition”, *Proceedings of the IEEE*, vol. 86, n° 11, 1998, pp. 2278–2324.

## 8 Bibliographie

[FIX 51] FIX E. AND HODGES J.L. “Discriminatory analysis, non parametric discrimination. Consistency properties.” *Technical Report 4, USAF School of Aviation medicine*, Randolph field, TX, 1951

[COV 67] COVER T.M. AND HART P.E., “Nearest neighbour pattern classification.” *IEEE Trans. on Inform. Theory*, Vol 13(1) pp. 21-27, 1967

[DUD 76] DUDANI S. A. “The distance weighted k-nearest neighbour rule”, *IEEE Trans SMC*, Vol. 6, pp. 325-327, 1976

[KEL 85] KELLER J.M., GRAY M.R., GIVENS J.A. “A fuzzy k-nearest neighbour algorithm” *IEEE Trans. SMC*, 15(4), pp. 580-585, 1985

[DON 95] DONOEUX T. “A k-nearest neighbour classification rule based on Dempster-Shaffer theory”, *IEEE Trans. on Systems, Man and Cybernetics*, Vol 25(5), pp. 804-813, 1995

[HAT 99] HATTORI K., TAKAHASHI M. “A new nearest neighbour rule in the pattern classification problem”, *Pattern Recognition*, Vol. 32, pp. 425-432, 1999

[ARI 05] ARIF M. “Fusion de données ; ultime étape de reconnaissance de formes, application à la fusion et à l'authentification ”, *thèse d'Informatique*, Université de Tours, Nov. 2005

[SIM 98] SIMARD P.Y., LECUN Y.A., DENKER J.S., VICTORRI E. “transformation invariance in pattern recognition” *Lecture Notes In computer sciences*, 1524, 1998

[VIN 01] VINCENT P., BENGIO Y. “K-local hyperplane and convex distance”, *Technical report 1197*, June 20<sup>th</sup>, 2001

[FIS 36] FISCHER R. A. “The use of multiple measurements in taxonomic problems”, *annals of Eugenics*, Vol.7, pp. 179-188, 1936

[IRI-06] iris data base, URL : <http://www.ics.uci.edu/~mllearn/databases/iris/>

[LEC 06] URL : <http://yann.lecun.com/exdb/mnist/>

[LEC 98] LECUN Y., BOTTOU L., BENGIO Y., HAFFNER P., “Gradient-Based Learning Applied to