

Utilisation de la distance de Hausdorff dans un SVM ?

Ludovic Simon, Jean-Philippe Tarel

Laboratoire Central des Ponts et Chaussées (ESE)

58 Boulevard Lefebvre

75015 PARIS

Résumé :

Notre objectif est de réaliser la détection et la reconnaissance automatique des panneaux de signalisation routière à partir d'une caméra montée sur un véhicule. Faire la détection et la reconnaissance, à partir des informations de forme, est généralement beaucoup plus difficile qu'avec des informations de couleurs. L'information de forme est pourtant essentielle, en particulier pour les panneaux. La distance de Hausdorff et ses variantes sont des mesures de similarité très utilisées pour comparer des contours. Elles sont donc des candidates naturelles pour construire le noyau d'un SVM qui permettrait de faire l'apprentissage sur des formes. Le noyau pouvant s'interpréter comme une mesure de dissimilarité, une question se pose : quelle fonction choisir pour transformer la distance de Hausdorff en noyau permettant au SVM d'avoir de bonnes performances ? Dans cet article, nous présentons une étude expérimentale qui ouvre plusieurs pistes quant à l'utilisation de la distance de Hausdorff modifiée dans le cadre d'un SVM. En particulier, nous avons observé un lien entre la qualité des résultats de classification obtenus par SVM et certaines caractéristiques simples à calculer.

Mots-clés : Apprentissage, Support Vector Machines, Noyau, Distance de Hausdorff, Appariement de contours, Forme, Noyaux, Classification.

1 INTRODUCTION

De plus en plus de caméras sont utilisées pour observer la route. Ces caméras peuvent être embarquées dans un véhicule ou fixées au voisinage de la route. Une des problématiques récurrentes posée par le traitement des images de routes est celle de la détection d'objets. Pour les besoins de nos applications, notre objectif est de détecter les panneaux dans les images routières.

Un panneau, même si c'est un objet bien normalisé peut avoir des aspects assez variés à cause des variations d'éclairage et des conditions météorologiques. Cela implique une phase d'apprentissage des différents aspects de chaque panneau. Pour faire cet apprentissage, nous avons choisi d'utiliser un algorithme nommé "Support Vector Machines" (SVM) connu pour son efficacité et dont la convergence vers un minimum unique est garantie, contrairement à d'autres techniques d'apprentissage. L'utilisation du SVM nécessite le choix d'une mesure de dissimilarité, le noyau, qui doit vérifier certaines contraintes.

Dans le cas des panneaux, l'information de contours est assez importante et elle est plus difficile à traiter que l'information de couleurs. La distance de Hausdorff permet de comparer deux contours. Elle est bien connue pour ses intéressantes propriétés [Dubuisson, 1994]. Elle semble donc être une candidate intéressante pour construire une mesure de dissimilarité utilisable par un SVM.

Cet article présente une étude sur la possibilité d'utiliser une distance de Hausdorff dans un SVM pour comparer des contours. Dans les parties 2 et 3, nous présentons la distance de Hausdorff et ses variantes. Dans la partie 4, nous rappelons l'essentiel de l'approche SVM. Dans les deux parties suivantes, nous étudions expérimentalement la possibilité d'utiliser la distance de Hausdorff ou une de ses variantes dans les SVM, en introduisant la notion de noyau défini positif et conditionnellement défini positif. Plusieurs noyaux sont proposés. Enfin, nous étudions expérimentalement les performances de ces noyaux sur des problèmes de classification avec des données synthétiques.

2 DISTANCE DE HAUSDORFF ET VARIANTES

Considérant deux ensembles de points A et B , la distance de Hausdorff directe h peut être écrite comme :

$$h(A, B) = \max_{a \in A} \left(\min_{b \in B} (\|a - b\|) \right) \quad (1)$$

La distance entre un point a et un ensemble de n points $B = \{b_1, \dots, b_n\}$ est noté :

$$d(a, B) = \min_{b \in B} \|a - b\| \quad (2)$$

On remarque que la fonction (1) n'est pas symétrique, elle ne peut donc pas être une distance au sens métrique. Pourtant, elle est beaucoup utilisée pour mesurer un degré de similarité entre deux contours A et B .

La symétrie est habituellement restaurée en prenant le maximum entre $h(A, B)$ et $h(B, A)$. Comme illustré sur la Fig. 1, on obtient alors la définition de la distance indirecte de Hausdorff :

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (3)$$

Comme il a été montré dans [Dubuisson, 1994], il existe des variantes aux définitions précédentes qui donnent de meilleurs résultats pour la reconnaissance de contours. Six façons de calculer les distances directes $h(A, B)$ et $h(B, A)$ ont été proposées (d_2 , d_3 et d_4

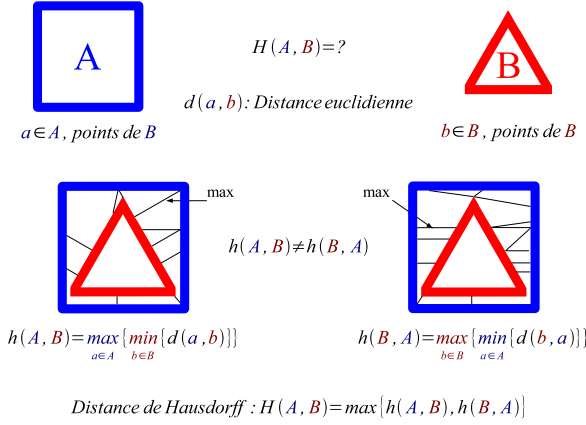


FIG. 1 – Distance de Hausdorff entre deux contours.

sont les distances de Hausdorff généralisées introduites dans [Huttenlocher, 1993] :

$$d_1(A, B) = \min_{a \in A} d(a, B) \quad (4)$$

$$d_2(A, B) = \frac{1}{K} \sum_{a \in A} d(a, B) \quad (5)$$

$$d_3(A, B) = \frac{1}{K} \sum_{a \in A} d(a, B)^2 \quad (6)$$

$$d_4(A, B) = \frac{1}{K} \sum_{a \in A} d(a, B)^{90} \quad (7)$$

$$d_5(A, B) = \max_{a \in A} d(a, B) \quad (8)$$

$$d_6(A, B) = \frac{1}{N_a} \sum_{a \in A} d(a, B) \quad (9)$$

De plus, quatre façons de construire la distance indirecte $H(A, B)$ sont proposées dans [Dubuisson, 1994] :

$$f_1(A, B) = \min(d(A, B), d(B, A)) \quad (10)$$

$$f_2(A, B) = \max(d(A, B), d(B, A)) \quad (11)$$

$$f_3(A, B) = \frac{d(A, B) + d(B, A)}{2} \quad (12)$$

$$f_4(A, B) = \frac{N_a d(A, B) + N_b d(B, A)}{N_a + N_b} \quad (13)$$

En combinant les distances d et les fonctions f , on obtient différentes variantes de la distance de Hausdorff. La combinaison de d_5 et f_2 définit la distance de Hausdorff classique (HD), présentée dans la Fig. 1. Celle de d_6 et f_2 définit la distance de Hausdorff modifiée (MHD). C'est cette dernière qui est recommandée dans [Dubuisson, 1994]. En effet, l'utilisation d'une moyenne la rend plus robuste au bruit. Aucune de ces variantes ne sont des distances métriques.

3 COMPORTEMENT DES VARIANTES

Le nombre de variantes proposées par [Dubuisson, 1994] est important. Nous avons donc réalisé une étude sur des contours simples, pour observer les avantages et les inconvénients de chaque variante.

Pour réaliser ces tests, on place un carré de référence centré à l'origine de notre repère. La forme comparée se déplace sur l'axe des abscisses. Pour ce faire, le barycentre de cette forme est positionné incrémentalement

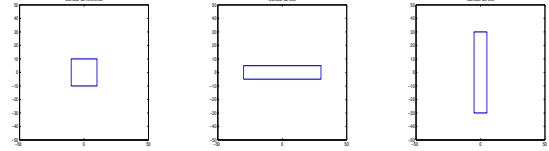


FIG. 2 – Un carré, un rectangle horizontal et un rectangle vertical ont été utilisés pour étudier le comportement des variantes de la distance de Hausdorff.

entre -50 pixels et $+50$ pixels de celui de la forme de référence. Les formes utilisées sont visibles sur la Fig. 2. La Fig. 3 illustre la comparaison entre les deux variantes Hausdorff classique (HD) et Hausdorff modifiée (MHD). Sur la première ligne de la Fig. 3, la forme comparée est un autre carré, sur la deuxième ligne, c'est un rectangle horizontal, et enfin sur la troisième ligne, c'est un rectangle vertical. Sur les deux premières lignes, première colonne, on trouve un comportement assez proche de la distance Euclidienne pour HD. Le comportement de MHD est similaire. Par contre sur la troisième ligne, le comportement de HD n'est pas satisfaisant car un aplatissement de la distance apparaît. Pour la MHD, les comportements sur la deuxième et la troisième lignes sont semblables. En résumé, le comportement de HD varie trop en fonction de l'orientation relative des formes comparées. Le comportement de MHD reste lui plus cohérent.

De façon plus générale, le fait d'utiliser d_6 et donc de faire une moyenne induit une plus grande tolérance à la présence de petites perturbations dans la forme des contours. Il est donc préférable d'utiliser d_6 au lieu de d_5 ou d_1 dans la définition de la mesure de similarité. Les distances d_2 , d_3 et d_4 ne sont pas utilisées car elles sont trop coûteuses en calculs.

La Fig. 4 montre l'influence du choix de f sur la mesure de similarité entre contours. Si le choix de f_2 et f_3 semble intéressant, l'utilisation de f_1 est à éviter du fait de l'aplatissement que l'on peut observer sur la figure en haut à droite. En effet, f_1 occasionne des comportements différents en fonction de l'orientation de la forme comparée. Entre f_2 et f_3 , on choisira f_2 , car c'est elle qui donne les comportements les plus semblables entre les deux orientations de la forme comparée.

La recommandation de [Dubuisson, 1994] est donc bien confirmée. C'est la combinaison de d_6 et f_2 qui semble la plus adaptée, nous prendrons donc la distance de Hausdorff modifiée (MHD) comme mesure de similarité dans la suite.

4 LES SUPPORT VECTOR MACHINES (SVM)

Les SVM, proposés par Vapnik en 1995 [Vapnik, 1995], sont de plus en plus utilisés en vision par ordinateur et en traitement d'images pour résoudre des problèmes de reconnaissance et de classification. Le principe du SVM est de séparer les données de façon à maximiser la marge entre les exemples positifs et négatifs [Cristianini, 2000]. Les données sont $\{(x_1, y_1), \dots, (x_l, y_l)\}$ où x_i est un

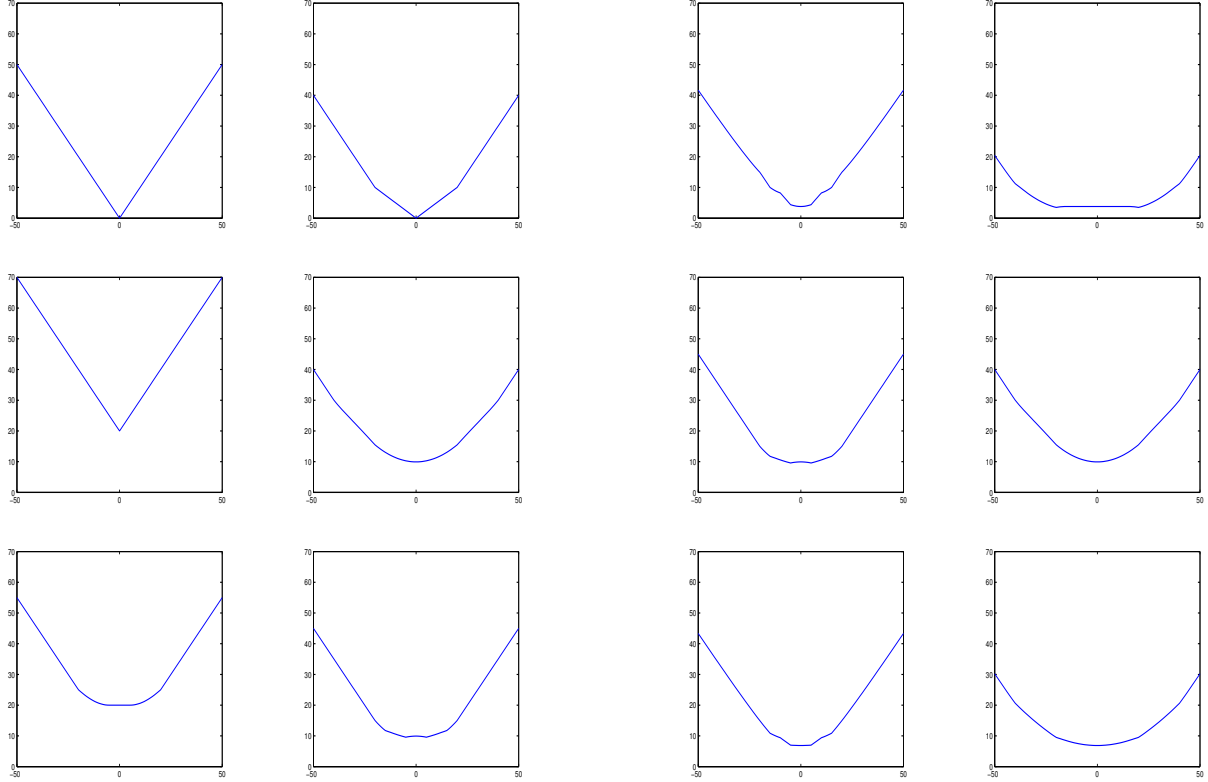


FIG. 3 – Évolution de la distance de Hausdorff et de Hausdorff modifiée entre deux contours en fonction de leurs positions relatives. La première colonne correspond à la distance de Hausdorff classique et la deuxième à celle modifiée. Sur la première ligne, les deux contours sont des carrés, sur la deuxième ligne, un carré et un rectangle horizontal, puis sur la troisième ligne, un carré et un rectangle vertical.

vecteur de \mathbb{R}^n représentant l'échantillon i et où $y_i \in \{-1, +1\}$ est son label.

Lorsque les exemples positifs et négatifs sont linéairement séparables, l'algorithme SVM estime les paramètres $w \in \mathbb{R}^n$ et $b \in \mathbb{R}$ de l'hyperplan $w \cdot x + b = 0$ qui les sépare, voir Fig. 5. On définit les vecteurs support (SV) comme les exemples les plus proches de l'hyperplan qui sont seuls utiles à caractériser cet hyperplan, voir Fig. 5.

Dans le cas où les exemples ne sont pas linéairement séparables, voir Fig. 6, un noyau $k(x, x')$ est utilisé afin de projeter les exemples dans un espace de plus grande dimension où le problème devient linéairement séparable. Toutefois, pour ce faire, le noyau choisi doit être défini positif. Cette condition assure en particulier la convergence de l'algorithme vers le minimum global.

D'autre part, on introduit un terme de régularisation L_1 , paramétré par C , qui autorise des erreurs de classification lors de l'apprentissage. Ceci est utile lorsque des exemples d'apprentissage ont des labels erronés. Une autre régularisation dite L_2 peut aussi être utilisée seule ou en combinaison avec la régularisation L_1 .

FIG. 4 – Évolution de la distance entre deux contours en fonction de leurs positions relatives. Dans la première colonne, un carré et un rectangle vertical sont utilisés, alors que, dans la deuxième colonne, c'est un carré et un rectangle horizontal. La première ligne correspond à la variante d_6 et f_1 , la deuxième ligne à d_6 et f_2 , et la dernière ligne à d_6 et f_3 .

En pratique, l'algorithme SVM consiste à minimiser, par rapport aux α_i , $1 \leq i \leq l$, le problème quadratique suivant :

$$W(\alpha) = - \sum_{i=1}^l \alpha_i + \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j k(x_i, x_j) \quad (14)$$

sous les contraintes $\sum_{i=1}^l \alpha_i y_i = 0$ et $0 \leq \alpha_i \leq C$.

Après le calcul des α_i , le b est calculé [Cristianini, 2000]. La fonction de décision du SVM est donné par le signe de :

$$g(x) = \sum_{i=1}^l \alpha_i y_i k(x_i, x) + b \quad (15)$$

5 UTILISATION DE LA DISTANCE DE HAUSDORFF DANS UN SVM ?

Le choix du noyau du SVM s'interprète comme le choix d'une mesure de dissimilarité. Pour pouvoir utiliser les distances non métriques introduites dans la partie 2, il est donc nécessaire d'appliquer une fonction m décroissante pour les transformer en mesure de dissimilarité. Dans la

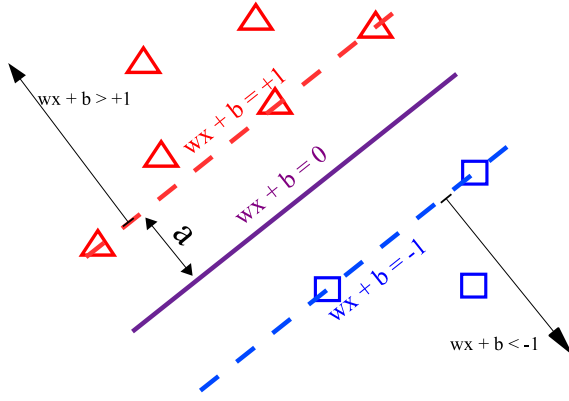


FIG. 5 – Classification linéairement séparable entre les \triangle et les \square . L’hyperplan séparateur est défini par $w \cdot x + b = 0$. La décision est donnée par le signe de $w \cdot x + b$. Les exemples situés sur les deux lignes discontinues sont des vecteurs support (SV). La marge a est la distance minimale des SV à l’hyperplan.

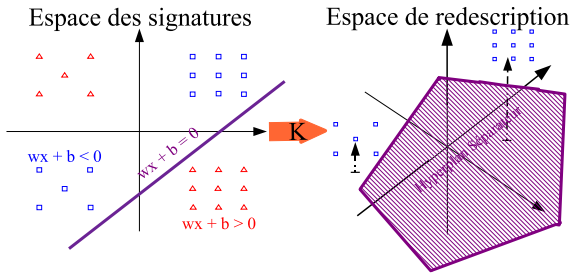


FIG. 6 – Comme les données 2D sur la gauche ne sont pas linéairement séparables, elles sont projetées dans \mathbb{R}^3 , par l’utilisation d’un noyau. Le SVM sépare alors les exemples positifs et négatifs projetés dans \mathbb{R}^3 par un hyperplan.

suite, on note seulement par s la similarité obtenue par la distance MHD entre deux contours.

Nous avons étudié les possibilités suivantes pour le choix de la fonction m :

- fonction opposée d’une puissance :

$$m_{\alpha}^{puiss}(s) = -\left(\frac{s}{\sigma}\right)^{\alpha} \quad (16)$$

- fonction exponentielle de l’opposée d’une puissance :

$$m_{\alpha}^{exp}(s) = \exp\left(-\left(\frac{s}{\sigma}\right)^{\alpha}\right) \quad (17)$$

- fonction hyperbolique à une puissance :

$$m_{\alpha}^{hyper}(s) = \frac{1}{\left(1 + \frac{s}{\sigma}\right)^{\alpha}} \quad (18)$$

- fonction T-student :

$$m_{\alpha}^{stud}(s) = \frac{1}{1 + \left(\frac{s}{\sigma}\right)^{\alpha}} \quad (19)$$

- fonction logarithme d’une puissance :

$$m_{\alpha}^{log}(s) = -\log\left(1 + \left(\frac{s}{\sigma}\right)^{\alpha}\right) \quad (20)$$

- fonction puissance d’un logarithme :

$$m_{\alpha}^{puiilog}(s) = -\log^{\alpha}\left(1 + \frac{s}{\sigma}\right) \quad (21)$$

- fonction puissance d’un logarithme de logarithme :

$$m_{\alpha}^{loglog}(s) = -\log^{\alpha}\left(1 + \log\left(1 + \frac{s}{\sigma}\right)\right) \quad (22)$$

Pour qu’un noyau k soit utilisable dans le cadre d’un SVM, une condition nécessaire est qu’il soit défini positif, c’est-à-dire que, quelque soit l’ensemble de données $\{x_1, \dots, x_l\}$, la matrice de Gram associée $K = (k(x_i, x_j))$, $1 \leq i \leq l$, $1 \leq j \leq l$ soit définie positive. En d’autres termes, la fonction noyau k sera définie positive, si pour tous l , pour tout exemple $x_1, \dots, x_l \in \chi$ et pour tout vecteur $v \in \mathbb{R}^l$, la matrice K satisfait $v^T K v \geq 0$.

Par la suite, on définira le noyau issu de l’utilisation de chaque fonction par la relation $k^{fonction}(x, x') = m_{fonction}^{fonction}(MHD(x, x'))$. Parmi les fonctions m ci-dessus, seules les fonctions m_{α}^{exp} , m_{α}^{hyper} et m_{α}^{stud} donnent des valeurs positives. Pour les autres, le noyau est à valeur négative ou nulle. De ce fait, ces derniers ne peuvent pas être des noyaux définis positifs.

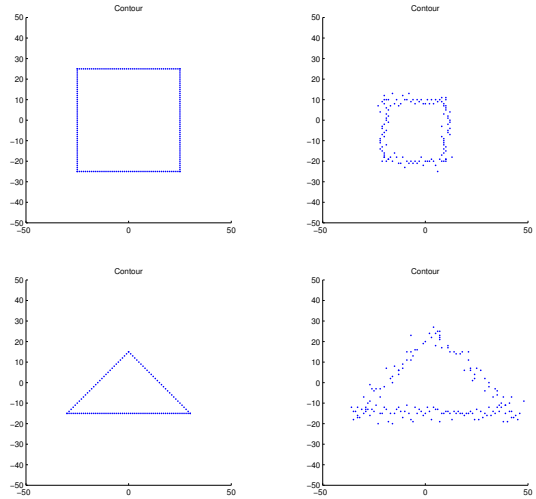


FIG. 7 – Exemples des formes générées avec un décalage tiré aléatoirement et avec aléatoirement l’absence ou la présence d’un bruit aléatoire.

A partir de [Berg, 1984, Haasdonk, 2004], on peut montrer que si la distance s est isométrique à une norme L_2 , alors k^{exp} et k^{stud} sont des noyaux définis positifs pour α entre 0 et 2. Par contre lorsque la distance n’est pas une métrique, ces noyaux ne sont pas nécessairement définis positifs. Nous avons vérifié cela expérimentalement. Pour ce faire, nous avons généré une base de 100 contours de deux types et de tailles aléatoires, comprenant 100 points chacun, comme dans la Fig. 7. Puis nous avons calculé la matrice de Gram associée à cette base. Le calcul de ses valeurs propres permet de voir si la matrice de Gram est définie positive ou non.

Nous avons fait varier la valeur de α entre 0 et 2 avec un pas de 0.1 et nous avons fait une observation intéressante. Comme le montre la Fig. 8 pour le noyau k^{exp} , plus la valeur de α diminue, moins il y a de valeurs propres négatives et plus la plus petite valeur propre augmente. Le paramètre α joue donc un rôle important. On ob-

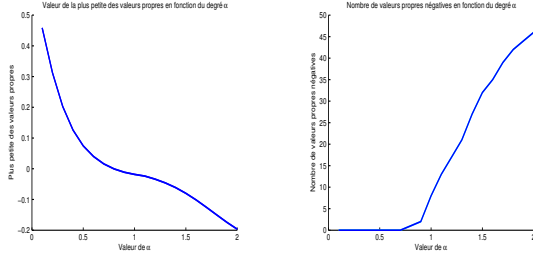


FIG. 8 – A droite, variations en fonction de α de la valeur de la plus petite des valeurs propres de la matrice de Gram associée au noyau $k^{exp}(x, x')$, et à gauche, nombre des valeurs propres négatives.

serve le même comportement pour k^{stud} . Pour le noyau $k^{hyper}(x, x')$, α a une influence moins déterminante.

6 NOYAUX CONDITIONNELLEMENT DÉFINIS POSITIFS

Comme indiqué dans [Scholkopf, 2000], il existe une autre condition nécessaire sur le noyau k qui permet de garantir le bon fonctionnement du SVM. En effet, si k est conditionnellement défini positif, le SVM est aussi assuré de converger vers le minimum global. L'ensemble des noyaux conditionnellement définis positifs est constitué des fonctions satisfaisant l'inégalité $v^T K v \geq 0$ pour tous les vecteurs v tels que la somme de ses composantes est nulle.

Il existe des liens forts entre les noyaux définis positifs et ceux conditionnellement définis positifs, voir [Berg, 1984, Scholkopf, 2000, Boughorbel, 2005]. Ainsi :

Propriété 1. Soit k un noyau symétrique sur $\mathcal{X} \times \mathcal{X}$. Alors pour tout $x_0 \in \mathcal{X}$, on définit :

$$\tilde{k}(x, x') = \frac{1}{2} [k(x, x') - k(x, x_0) - k(x', x_0) + k(x_0, x_0)] \quad (23)$$

Alors, \tilde{k} est défini positif si et seulement si k est conditionnellement défini positif.

La propriété précédente peut être utilisée pour tester expérimentalement si un noyau est conditionnellement défini positif. En effet, étant donné un noyau k , il suffit de construire \tilde{k} et de tester si la matrice de Gram associée à \tilde{k} est définie positive.

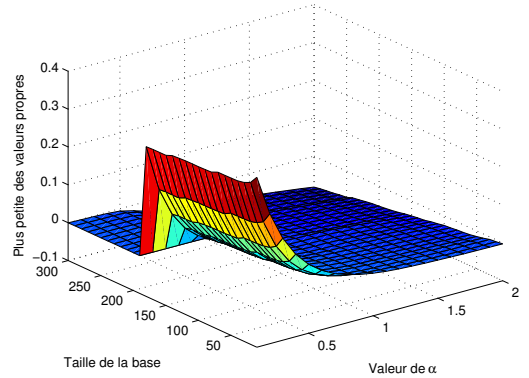
Un problème se pose alors, que choisir pour x_0 ? Son rôle n'est pas fondamental, il suffit qu'il soit fixe. Nous avons donc utilisé pour x_0 le singleton constitué de l'origine du repère.

La classe des noyaux conditionnellement définis positifs est beaucoup plus large et nous permet maintenant de considérer l'ensemble des noyaux introduits dans la partie précédente. Ainsi à partir de [Berg, 1984, Haasdonk, 2004, Boughorbel, 2005], on peut montrer

que si la distance s est isométrique à une norme L_2 , alors k^{puiss} et k^{log} sont des noyaux conditionnellement définis positifs pour α entre 0 et 2. De même, $k^{puisslog}$ et k^{loglog} le sont pour α entre 0 et 1. Par contre, lorsque la distance n'est pas une métrique, ces noyaux ne sont pas conditionnellement définis positifs.

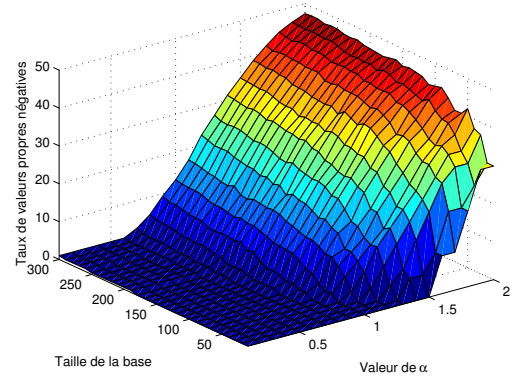
Comme dans la partie précédente, nous avons généré aléatoirement des ensembles de formes constituées de 100 points chacun, comme illustré dans la Fig. 7. Les matrices pour les différents noyaux et les différentes valeurs de α entre 0 et 2 ont été calculées. La valeur utilisée pour σ est de 80. Ayant constaté que la taille de la base l avait son importance, nous avons aussi étudié l'effet de l lorsqu'il varie entre 10 et 300.

Valeur de la plus petite des valeurs propres en fonction de la taille de la base et du degré α



(a)

Pourcentage de valeurs propres négatives en fonction de la taille de la base et du degré α

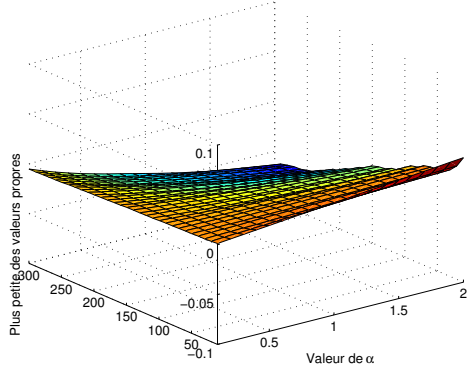


(b)

FIG. 9 – En (a), évolution de la valeur de la plus petite des valeurs propres de la matrice de Gram, en (b), évolution du taux de valeurs propres négatives, en fonction de α et de la taille de la base l , pour le noyau \tilde{k}^{loglog} .

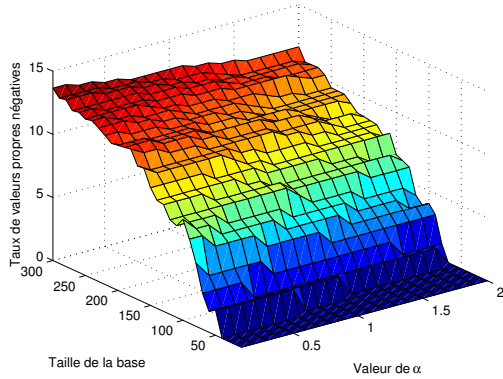
Nous avons obtenu des résultats très similaires pour tous les noyaux, seul le noyau k^{hyper} se distingue des autres. C'est pour cela que la Fig. 9 montre que les résultats pour les noyaux \tilde{k}^{loglog} et la Fig. 10 pour le noyau \tilde{k}^{hyper} . A part pour \tilde{k}^{hyper} , les noyaux ont leur plus petite valeur propre qui reste positive pour $0 < \alpha \leq 1$ et $0 < l \leq 180$. Pour \tilde{k}^{hyper} , c'est pour $0 < \alpha \leq 2$ et $0 < l \leq 40$ seulement que la plus petite valeur propre est positive. Dans tous les cas, il existe donc une taille maximum de

Valeur de la plus petite des valeurs propres en fonction de la taille de la base et du degré α



(a)

Pourcentage de valeurs propres négatives en fonction de la taille de la base et du degré α



(b)

FIG. 10 – En (a), évolution de la valeur de la plus petite des valeurs propres de la matrice de Gram, en (b), évolution du taux de valeurs propres négatives, en fonction de α et de la taille de la base l , pour le noyau \tilde{k}^{hyper} .

la base au delà de laquelle des valeurs propres négatives apparaissent. Cette taille maximum est reliée au type de noyau et à la nature du problème traité. Enfin, comme dans la section précédente le fait de diminuer α est favorable pour diminuer l'importance et le nombre de valeurs propres négatives.

Nous illustrons maintenant l'effet du choix de α sur la mesure de similarité entre deux formes, prenant l'exemple du noyau $k^{puiss}(x, x')$. Comme le montre la Fig. 11, plus α est proche de 0 et plus l'évolution de la dissimilarité entre deux contours en fonction de leurs positions relatives tend à être piquée lorsque les contours sont semblables. En revanche, si ces derniers sont différents, la Fig. 12 montre que la dissimilarité garde la même allure que pour la distance MHD de la Fig. 3, à ceci près que l'amplitude des valeurs est plus réduite à mesure que l'on s'approche de $\alpha = 0$.

7 CLASSIFICATION

Dans les parties précédentes, nous avons étudié les noyaux selon leurs propriétés intrinsèques. Parmi les noyaux que nous avons construits à partir de la dis-

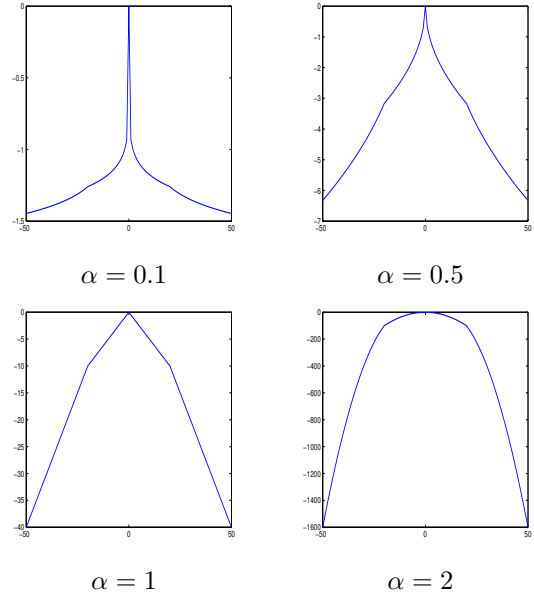


FIG. 11 – Évolution de la dissimilarité produite par $k^{puiss}(x, x')$ entre deux mêmes formes, deux carrés, en fonction de leurs positions relatives.

tance de Hausdorff modifiée, aucun n'est défini positif, ni conditionnellement défini positif. Pour autant, dans [Haasdonk, 2004] l'auteur assure qu'il est possible d'utiliser ce type de noyaux où une distance non métrique est substituée à une distance métrique, sans que cela pose d'inconvénient. Nous avons donc cherché à tester cela expérimentalement.

Afin de vérifier si les noyaux proposés permettent de faire de bonnes classifications, nous avons généré une base synthétique contenant une classe positive de 300 triangles et une classe négative de 300 carrés. Chaque contour a des caractéristiques différentes :

- de positions, avec des décalages entre -5 et $+5$ pixels,
- d'échelles, avec des tailles allant de 16 à 70 pixels de hauteurs,
- et sur certains, un bruit gaussien 2D isotrope centré et de variance égale à 20% de la taille de la forme est rajouté sur la position des points constituant le contour.

Ces perturbations sont illustrées sur la Fig. 7. Une base de $2l$ contours est utilisée, avec l entre 10 et 300 avec un pas de 10. Le paramètre α varie de 0.1 à 2, avec un pas de 0.1. Le paramètre C de la régularisation L_1 a été fixé à 10^8 . Si celui-ci est mis à l'infini, il n'y a alors plus de régularisation, et nous avons observé que le SVM ne converge pas toujours pour les α grands. La régularisation L_1 est donc absolument nécessaire.

Afin d'éviter des erreurs dues à l'aspect aléatoire du tirage de la base, on effectue une validation croisée à $k = 4$ parties (k-cross validation). Cela consiste à séparer la base en $k = 4$ parties de même taille. Puis, on choisit deux parties pour constituer la base d'apprentissage, les deux parties restantes étant utilisées pour constituer la base de test. Les résultats, le taux de SV utilisés à l'issue de l'apprentissage ainsi que les taux de reconnaissance et

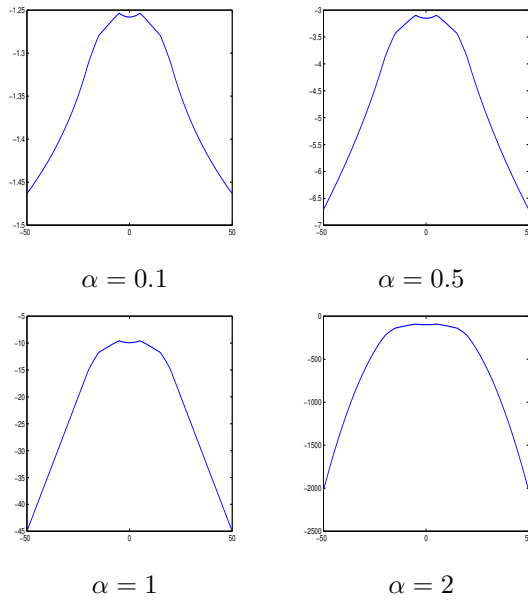


FIG. 12 – Évolution de la dissimilarité produite par $k^{puiss}(x, x')$ entre deux contours, un carré et un rectangle vertical, en fonction de leurs positions relatives.

de fausses détection sont moyennés sur les 6 permutations possibles.

Dans nos expérimentations, nous avons observé, voir la Fig. 13(a), que plus α augmente et plus le pourcentage de vecteurs support (SV) diminue. D'autre part, ce pourcentage évolue aussi favorablement vers une diminution lorsque la taille de la base d'apprentissage augmente. Les résultats sont très similaires pour tous les noyaux que nous avons introduits dans la partie 5, sauf pour le noyau $k^{hyper}(x, x')$. Ce dernier noyau ayant un comportement atypique mais cohérent avec nos conclusions, par souci de concision, il ne sera plus considéré dans la suite. Enfin, il faut noter que le taux de SV reste au dessus de 50% ce qui est assez élevé. Or un fort taux de SV diminue l'intérêt de faire un apprentissage par SVM, comparé à d'autres techniques.

Sur les bases de test, nous avons aussi évalué la qualité de l'apprentissage en mesurant le taux d'erreurs. Ce dernier est défini comme la somme du taux de fausses détections et du taux de non reconnaissance. Nous avons obtenu, voir Fig. 13(b), un bon taux de reconnaissance et un faible taux de fausses détections pour des α entre 0.1 et 0.9. Pour les valeurs de α supérieures, la qualité de l'apprentissage se dégrade à mesure que la valeur de α augmente. Nous avons observé dans des tests supplémentaires qu'à partir d'une certaine valeur de la taille l de la base, la qualité de la classification se dégrade soudainement. Cette taille seuil est fonction de la valeur de α .

Si nous comparons maintenant le taux d'erreurs de la Fig. 13(b) avec le taux de valeurs propres négatives de la Fig. 9(b), on remarque que les deux surfaces se ressemblent. Il en est de même pour le noyau $k^{hyper}(x, x')$. Une corrélation entre le taux d'erreurs et le taux de valeurs propres négatives est très importante à noter car

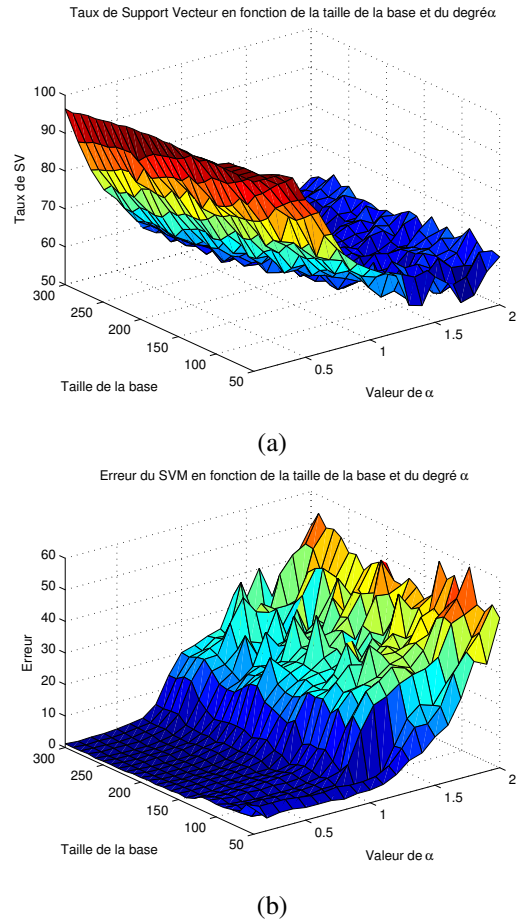
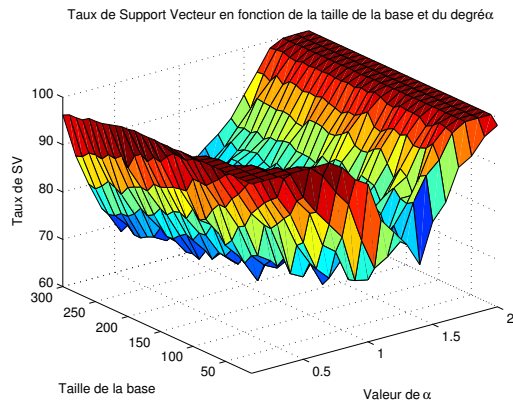


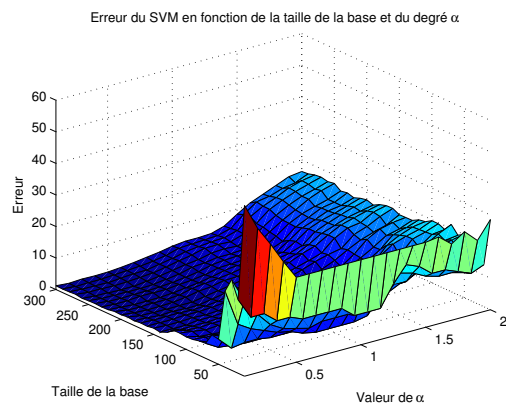
FIG. 13 – En (a), pourcentage de SV utilisés par le SVM et en (b), indice de qualité de l'apprentissage, en fonction de α et de la taille l de la base d'apprentissage, avec le noyau $k^{puiss}(x, x')$ et la régularisation L_1 .

elle permet de diagnostiquer la qualité de l'apprentissage sans avoir besoin de faire la classification, simplement en calculant les valeurs propres de la matrice de Gram. Cela permet d'espérer un gain de temps pour accélérer le réglage des hyper-paramètres comme C et σ .

Dans [Haasdonk, 2004], il est suggéré d'ajouter une régularisation L_2 en plus de celle L_1 pour augmenter les performances de la classification. Cette régularisation consiste à ajouter une valeur constante c sur la diagonale de la matrice de Gram avant apprentissage. Pour choisir la valeur de c , nous avons ajouté la valeur absolue de la plus petite valeur propre de la matrice de Gram, lorsque celle-ci est négative. Quand la plus petite valeur propre est positive, alors la matrice de Gram reste inchangée. La valeur de c est choisie automatiquement, de façon à ce que la matrice de Gram soit toujours définie positive après régularisation. Sur les mêmes données, nous avons alors obtenu pour tous les noyaux étudiés, comme le montre Fig. 14(a), une augmentation du nombre de SV utilisés par rapport à l'utilisation de la régularisation L_1 seule. Ceci n'est pas très satisfaisant comme expliqué précédemment. Par contre, comme le montre Fig. 14(b), la zone où la qualité est bonne s'est étalée selon l'axe des



(a)



(b)

FIG. 14 – En (a), pourcentage de SV utilisés par le SVM et en (b), indice de qualité de l'apprentissage, en fonction de α et de la taille l de la base d'apprentissage, avec le noyau $k^{puiss}(x, x')$ et la régularisation $L_1 + L_2$.

α . Selon l'axe l , il ne semble plus y avoir de baisse de la qualité de la classification.

8 CONCLUSIONS ET PERSPECTIVES

Après avoir rappelé la définition de la distance de Hausdorff et de ses variantes, mesures de similarité très utilisées en traitement d'images pour comparer deux contours, nous avons étudié expérimentalement leurs performances dans le cadre de l'apprentissage par SVM. Le noyau du SVM étant une mesure de dissimilarité, nous avons proposé de composer la distance de Hausdorff avec diverses fonctions décroissantes paramétrées simplement et nous avons étudié l'influence de leurs paramètres.

La distance de Hausdorff n'étant pas une distance métrique, les noyaux proposés ne sont ni définis positifs, ni conditionnellement définis positifs. Par contre, d'après nos expérimentations, ces noyaux permettent d'obtenir de bonnes performances en terme de taux de reconnaissance et de taux de fausses détections. Le seul bémol est que le taux de vecteurs support reste toujours au dessus de 50% pour toutes les tailles de la base d'apprentissage que nous avons testées. Ceci réduit l'intérêt de choisir une tech-

nique d'apprentissage dont l'avantage est normalement sa capacité à synthétiser de façon compacte la base d'apprentissage.

Nous avons observé un lien entre le taux d'erreurs lors de la reconnaissance et le taux de valeurs propres négatives de la matrice de Gram associée. Cette observation nous a permis de proposer une méthode automatique pour choisir le paramètre utilisé lors de la régularisation L_2 . Plus généralement, cela permet d'espérer des gains de temps importants pour diagnostiquer un SVM qui utilise un noyau non défini positif ou non conditionnellement défini positif. En particulier, nous pouvons espérer accélérer le réglage des hyper-paramètres.

Une autre perspective intéressante à ces travaux est d'introduire les couleurs autour des contours au niveau du noyau pour combiner les informations de formes et de couleurs et améliorer les performances en pratique.

BIBLIOGRAPHIE

- [Berg, 1984] C. Berg, J. P. R. Christensen et P. Ressel, *Harmonic Analysis on Semigroups : Theory of Positive Definite and Related Functions*, Springer-Verlag, 1984.
- [Boughorbel, 2005] S. Boughorbel, J.-P. Tarel et N. Boujemaa, Conditionally Positive Definite Kernels for SVM Based Image Recognition, Actes IEEE International Conference on Multimedia and Expo (ICME'05), Amsterdam, The Netherlands, 2005. (2005).
- [Cristianini, 2000] N. Cristianini et J. Shawe-Taylor, *An Introduction to Support Vector Machines (and other kernel based learning methods)*, Cambridge University Press, 2000.
- [Dubuisson, 1994] M. P. Dubuisson et A. K. Jain, A Modified Hausdorff Distance for Object Matching, Actes International Conference on Pattern Recognition, A :566-568, 1994.
- [Haasdonk, 2004] B. Haasdonk et C. Bahlmann, Learning with Distance Substitution Kernels, Actes 26th DAGM Symposium, Tübingen, Germany, Springer Berlin, 2004.
- [Huttenlocher, 1993] D. Huttenlocher, D. Klanderman et A. Rucklidge, Comparing images using the Hausdorff distance, IEEE Transactions on Pattern Analysis and Machine Intelligence, 15 :(9), 850-863, 1993.
- [Scholkopf, 2000] B. Scholkopf, *The kernel trick for distances*, Actes Neural Information Processing Systems, 301-307, 2000.
- [Vapnik, 1995] V. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, 1995.