

Estimation à noyau adaptatif dans des espaces multidimensionnels hétérogènes

Aurélie BUGEAU¹, Patrick PÉREZ¹

¹IRISA INRIA, Rennes, France
aurelie.bugeau@irisa.fr, perez@irisa.fr

Résumé – Les méthodes d’estimation à noyau, telles que le mean shift, ont un inconvénient majeur : le choix de la taille du noyau. La sélection de cette taille devient vraiment difficile dans le cas de données multidimensionnelles et hétérogènes. Nous présentons une solution à ce problème. La taille est choisie itérativement pour chaque type de données, en cherchant parmi un ensemble de tailles prédéfinies celle qui donne localement les résultats les plus stables. La sélection itérative nécessite l’introduction d’un nouvel estimateur. La méthode est validée dans le contexte de la segmentation d’image couleur et de mouvement.

Abstract – Kernel estimation techniques, such as mean shift, suffer from one major drawback: the kernel bandwidth selection. This selection becomes a real challenge in case of multidimensional heterogeneous features. This paper presents a solution to this problem. The selection is done iteratively for each type of features, by looking for the stability of local bandwidth estimates within a predefined range of bandwidths. A new estimator that permits the iterative computation is introduced. The validity of the method is demonstrated in the context of color image segmentation and motion segmentation.

1 Introduction

Les algorithmes mean shift sont très populaires en classification de données et segmentation d’image grâce à leurs performances et leurs faibles coûts de calcul [3][5]. Ces méthodes présentent l’avantage de fournir automatiquement le nombre de classes. Cependant, elles sont basées sur l’estimation d’une densité à noyau dont la taille doit être définie préalablement.

Les méthodes mean shift usuelles utilisent un noyau de taille fixe pour tout l’espace de données. Cependant, lorsque les données varient de manière significative le long de l’espace caractéristique, des noyaux de taille variable sont préférables.

La sélection de la taille de noyau se fait généralement en essayant d’équilibrer le biais et la variance. Dans [6] les auteurs combinent deux méthodes pour résoudre le problème de sélection des noyaux de taille variable. La première, basée sur la loi “plug-in” de [7], a de très bonnes propriétés statistiques. Cependant, cette loi est difficile à mettre en œuvre pour des données multidimensionnelles [9][10, p.109]. La seconde méthode, plus détaillée dans [4], est une sélection semi paramétrique. Les meilleures tailles sont choisies en utilisant un ensemble de tailles prédéfinies et en réalisant une approximation locale des données multidimensionnelles par une loi normale. Les résultats obtenus pour la segmentation sont très prometteurs. Néanmoins, cette méthode a encore un défaut : si l’espace des données est composé de différents types d’information, autrement dit d’un ensemble de données hétérogènes, l’ensemble de tailles prédéfinies devient grand et l’algorithme coûteux.

Dans ce papier, nous donnons une solution au problème du choix de la taille des noyaux dans le cas de données multidimensionnelles hétérogènes. La sélection est faite en

utilisant itérativement pour chaque type d’information, ou domaine, la méthode décrite dans [4]. Nous détaillons cette méthode et rappelons le principe des méthodes mean shift dans la section 2. Notre algorithme itératif est lui décrit dans la section 3. Un nouvel estimateur à noyau permettant ce calcul itératif y est introduit. Enfin, dans la section 4, la validité de notre méthode est montrée pour des applications à la segmentation d’images couleur et à la segmentation de mouvement.

2 Mean shift et taille du noyau

2.1 Le filtrage mean shift à noyau fixe

Le mean shift est une méthode itérative de montée de gradient convergeant vers le mode, *i.e.* le maximum local, de la densité d’un nuage de points. Le filtrage mean shift est bien décrit dans [6]. Nous rappelons ici les résultats. Étant donnés n points $\{\mathbf{x}^{(i)}\}_{i=1..n}$ dans l’espace de dimension d , \mathcal{R}^d , le vecteur mean shift pour un noyau gaussien de covariance \mathbf{H} est donné par :

$$\mathbf{m}_{\mathbf{H}}(\mathbf{x}) = \frac{\sum_{i=1}^n \mathbf{x}^{(i)} \exp(-\frac{1}{2}D^2(\mathbf{x}, \mathbf{x}^{(i)}, \mathbf{H}))}{\sum_{i=1}^n \exp(-\frac{1}{2}D^2(\mathbf{x}, \mathbf{x}^{(i)}, \mathbf{H}))} - \mathbf{x} \quad (1)$$

où $D^2(\mathbf{x}, \mathbf{x}^{(i)}, \mathbf{H}) \equiv (\mathbf{x} - \mathbf{x}^{(i)})^T \mathbf{H}^{-1} (\mathbf{x} - \mathbf{x}^{(i)})$ est la distance de Mahalanobis de \mathbf{x} à $\mathbf{x}^{(i)}$.

Supposons maintenant que l’espace de dimension d peut être décomposé en P espaces indépendants de dimensions $d_p, p = 1, \dots, P$ (où $\sum_{p=1}^P d_p = d$), chacun associé à un type d’information différent (*e.g.* position, couleur). Les espaces étant indépendants, la matrice \mathbf{H} peut s’écrire $\mathbf{H} = \text{diag}[\mathbf{H}_1 \dots \mathbf{H}_P]$ et le vecteur mean shift devient égal à

$$\mathbf{m}_{\mathbf{H}}(\mathbf{x}) = \frac{\sum_{i=1}^n \mathbf{x}^{(i)} \prod_{p=1}^P \exp(-\frac{1}{2}D^2(\mathbf{x}_p, \mathbf{x}_p^{(i)}, \mathbf{H}_p))}{\sum_{i=1}^n \prod_{p=1}^P \exp(-\frac{1}{2}D^2(\mathbf{x}_p, \mathbf{x}_p^{(i)}, \mathbf{H}_p))} - \mathbf{x} \quad (2)$$

avec $\mathbf{x}^{(i)T} = (\mathbf{x}_1^{(i)T}, \dots, \mathbf{x}_p^{(i)T})$ et $\mathbf{x}^T = (\mathbf{x}_1^T, \dots, \mathbf{x}_p^T)$. Le filtrage mean shift est obtenu en appliquant à chaque itération le déplacement donné par l'équation 1 (ou 2). Cet algorithme converge vers le mode local de la densité [6]. La classification finale de l'ensemble des données est obtenue en regroupant dans une même classe tous les points ayant convergé vers le même mode.

2.2 Le filtrage mean shift à noyau variable

Lorsque les données varient de manière significative le long de l'espace, il est préférable d'utiliser un estimateur à noyau de taille variable. Deux principaux estimateurs à noyau variable existent [10]. Pour le premier, appelé estimateur "sample point", la taille s'adapte aux points de données observés. Une bonne valeur du noyau peut considérablement réduire le biais de cet estimateur. Cependant trouver cette valeur pour des données hétérogènes s'avère très difficile. Dans [8] les auteurs ont montré que l'influence d'un tel estimateur par rapport à un estimateur à noyau fixe est très faible quand la taille de l'espace des données augmente.

Pour le second estimateur, appelé estimateur "balloon", la taille \mathbf{H} du noyau K s'adapte aux points d'estimation :

$$\hat{f}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{|\mathbf{H}(\mathbf{x})|^{1/2}} K(\mathbf{H}(\mathbf{x})^{-1/2}(\mathbf{x} - \mathbf{x}^{(i)})) \quad (3)$$

Dans [8], les auteurs ont montré que cet estimateur devient très intéressant quand la dimension des données devient supérieure à 3. C'est pourquoi nous utiliserons ce deuxième estimateur et introduisons le mean shift associé. Sa dérivée $\nabla_p \hat{f}(\mathbf{x})$ a des termes dépendants de $(\mathbf{x} - \mathbf{x}^{(i)})^2$ et $\mathbf{H}'(\mathbf{x})$. Dans le contexte du mean shift, la fonction \mathbf{H} est seulement définie de façon discrète aux points d'estimation. Il est donc possible de l'interpoler afin d'obtenir une fonction continue dont la dérivée s'annule pour chaque point d'estimation. La dérivée devient alors :

$$\hat{\nabla} f(\mathbf{x}) = \frac{1}{n} \left[\sum_{i=1}^n \mathbf{H}(\mathbf{x})^{-1} K(\mathbf{H}(\mathbf{x})^{-1/2}(\mathbf{x} - \mathbf{x}^{(i)})) \right. \\ \left. \left[\frac{\sum_{i=1}^n \mathbf{x}^{(i)} K(\mathbf{H}(\mathbf{x})^{-1/2}(\mathbf{x} - \mathbf{x}^{(i)}))}{\sum_{i=1}^n K(\mathbf{H}(\mathbf{x})^{-1/2}(\mathbf{x} - \mathbf{x}^{(i)}))} - 1 \right] \right] \quad (4)$$

Le dernier terme de l'équation précédente est le vecteur mean shift. Un algorithme de filtrage et de classification peut être écrit à l'aide de ce vecteur [1]. La démonstration de convergence est très proche de celle pour un noyau fixe. Comme pour l'estimateur à noyau fixe, les équations précédentes peuvent facilement se réécrire dans le cas de données multidimensionnelles.

2.3 Sélection de la taille de noyau

La qualité des résultats de classification dépend grandement du choix de la taille du noyau représentée par la matrice \mathbf{H} . Dans [4], Comaniciu propose un algorithme de sélection basé sur la propriété suivante. En supposant qu'autour d'un point les données sont normalement distribuées, il est possible de montrer que la norme du vecteur mean shift est maximisée si la taille du noyau est égale à la variance de cette distribution. La difficulté est alors

de définir et de construire le voisinage du point pour lequel la distribution normale doit être calculée. Comaniciu propose de construire un voisinage à l'aide d'une classification mean shift. Divers voisinages sont construits pour un ensemble de B noyaux prédéfinis $\{\mathbf{H}^{(b)}, b = 1 \dots B\}$, aussi appelés échelles. Chacune des classes obtenues à chaque échelle est représentée par une loi normale. Le meilleur voisinage est celui pour lequel la distribution paramétrique est la plus stable comparée aux distributions obtenues à des échelles voisines. En effet si un point est rajouté ou enlevé du voisinage la distribution paramétrique ne devrait pas bouger.

L'algorithme de sélection décrit dans [4] peut alors se décomposer en deux étapes. La première, appelée évaluation de la taille pour la classification, consiste à faire une classification mean shift à noyau fixe pour chacun des noyaux prédéfinis et à construire une distribution normale pour chacune des classes obtenues. Comaniciu utilise le mode d'une classe pour moyenne et une formule découlant de la limite en probabilité du vecteur mean shift pour calculer la covariance. Nous ne détaillerons pas ces calculs ici. Notons simplement que la formule obtenue a quelques désavantages. En particulier, la covariance peut être négative. Dans la suite nous utiliserons les formules traditionnelles pour le calcul de la moyenne et de la covariance des distributions normales.

La deuxième étape est appelée évaluation de la taille pour les données. Chaque point $\mathbf{x}^{(i)}$ de l'espace de données est associé, pour chaque échelle b , à la distribution p_b de la classe à laquelle il appartient. Les distributions obtenues pour des échelles voisines sont comparées et cela dans le but d'obtenir la distribution la plus stable. Les distributions sont comparées en utilisant la divergence de Jensen-Shannon. Pour chaque point $\mathbf{x}^{(i)}$ et pour chaque échelle b , la divergence compare les trois échelles p_{b-1}, p_b, p_{b+1} . La meilleure échelle est celle minimisant cette mesure, et la covariance finale pour le point $\mathbf{x}^{(i)}$ est la covariance associée.

Après avoir trouvé la meilleure taille de noyau en chaque point, Comaniciu l'applique au mean shift basé sur l'estimateur sample point. Cependant, dans le cas de données hétérogènes aucun résultat ne permet de lier directement la diminution du biais à la norme du gradient de l'estimateur. Pour cette raison et pour celles décrites dans la sous section précédente, nous appliquons les résultats de taille de noyau au mean shift utilisant l'estimateur balloon.

3 Cas des données hétérogènes

L'algorithme de sélection de taille décrit dans la section précédente devient coûteux dès lors que l'on considère des données hétérogènes d'assez grande dimension. En effet, si, pour chaque domaine p , B_p noyaux sont prédéfinis, la classification mean shift doit être appliquée $\prod_{p=1}^P B_p$ fois pour prendre en compte chaque possibilité. Cet algorithme de sélection n'est donc pas adapté pour des espaces de grandes dimensions composés de plusieurs parties distinctes. De plus, il n'est pas évident de savoir comment la comparaison entre trois échelles successives devrait être

adaptée au cas de données hétérogènes. Dans la suite nous proposons de trouver itérativement la meilleure taille de noyau pour chacun des domaines.

Supposons que nous voulions trouver la meilleure taille pour le premier domaine. On définit les matrices temporaires $\tilde{\mathbf{H}}_p, p = 2 \dots P$ pour chacun des autres domaines. Ces matrices sont constantes pour chaque échelle et égales à la moyenne de toutes les matrices prédéfinies : $\tilde{\mathbf{H}}_p = \frac{1}{B_p} \sum_{b=1}^{B_p} \mathbf{H}_p^{(b)}, p > 1$. L'algorithme de sélection de la section précédente est appliqué pour l'ensemble des tailles :

$$\{\tilde{\mathbf{H}}^{(b)} = \text{diag}[\mathbf{H}_1^{(b)}, \tilde{\mathbf{H}}_2, \dots, \tilde{\mathbf{H}}_P], b = 1 \dots B_1\}$$

et donne la meilleure taille $\Sigma_1(\mathbf{x}^{(i)})$ pour chaque point $\mathbf{x}^{(i)}$. On réitère l'opération pour chacun des autres domaines. Pour les espaces déjà étudiés la matrice correspondant à la meilleure taille n'est désormais plus constante : $\tilde{\mathbf{H}}_p(\mathbf{x}^{(i)}) = \Sigma_p(\mathbf{x}^{(i)})$. L'algorithme de la section précédente utilise le filtrage mean shift avec des noyaux de taille fixe pour classifier les données. Il ne peut donc plus être appliqué tel quel. Une solution est d'utiliser un nouvel estimateur, défini comme la combinaison d'un estimateur à noyau fixe et d'un estimateur balloon. Le but étant d'estimer itérativement la meilleure taille de noyau pour chaque type de données, pour $p' = 1 \dots P$, on définit l'estimateur suivant :

$$\hat{f}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \left(\frac{K(\|\mathbf{H}_{p'}^{-1/2}(\mathbf{x}_{p'} - \mathbf{x}_{p'}^{(i)})\|^2)}{|\mathbf{H}_{p'}|^{1/2}} \right) \prod_{p \neq p'} \frac{K(\|\mathbf{H}_p(\mathbf{x})^{-1/2}(\mathbf{x}_p - \mathbf{x}_p^{(i)})\|^2)}{|\mathbf{H}_p(\mathbf{x})|^{1/2}} \quad (5)$$

qui est une densité pour $\mathbf{x}_{p'}$ conditionnellement à $(\mathbf{x}_p)_{p \neq p'}$. Il est de nouveau possible ici de considérer $\mathbf{H}'(\mathbf{x}) = 0$. En utilisant un noyau gaussien et en introduisant

$$P_{\mathbf{H}_{p'}}(\mathbf{x}, \mathbf{x}^{(i)}) = k(\|\mathbf{H}_{p'}^{-1/2}(\mathbf{x}_{p'} - \mathbf{x}_{p'}^{(i)})\|^2) \prod_{p \neq p'} k(\|\mathbf{H}_p(\mathbf{x})^{-1/2}(\mathbf{x}_p - \mathbf{x}_p^{(i)})\|^2) ,$$

la dérivée de l'estimateur si $p \neq p'$ est égale à

$$\nabla_p \hat{f}(\mathbf{x}) = \mathbf{H}_p(\mathbf{x})^{-1} \hat{f}(\mathbf{x}) \left[\frac{\sum_{i=1}^n \mathbf{x}_p^{(i)} P_{\mathbf{H}_{p'}}(\mathbf{x}, \mathbf{x}^{(i)})}{\sum_{i=1}^n P_{\mathbf{H}_{p'}}(\mathbf{x}, \mathbf{x}^{(i)})} - \mathbf{x}_p \right] ,$$

et à :

$$\nabla_{p'} \hat{f}(\mathbf{x}) = \mathbf{H}_{p'}^{-1} \hat{f}(\mathbf{x}) \left[\frac{\sum_{i=1}^n \mathbf{x}_{p'}^{(i)} P_{\mathbf{H}_{p'}}(\mathbf{x}, \mathbf{x}^{(i)})}{\sum_{i=1}^n P_{\mathbf{H}_{p'}}(\mathbf{x}, \mathbf{x}^{(i)})} - \mathbf{x}_{p'} \right] .$$

si $p = p'$. Le dernier terme des deux équations précédentes est le vecteur mean shift $\mathbf{m}_p(\mathbf{x})$ si $p \neq p'$, respectivement $\mathbf{m}_{p'}(\mathbf{x})$. Le filtrage mean shift peut être effectué avec ces vecteurs, et sa convergence démontrée [1].

L'algorithme itératif final pour la sélection de la taille est décrit dans l'algorithme 1. Avec l'algorithme initial, la classification mean shift devait être faite $\prod_{p=1}^P B_p$ fois afin d'étudier toutes les tailles possibles. Avec notre méthode la classification est faite $\sum_{p=1}^P B_p$ fois.

Algorithme 1 Estimation itérative de la taille du noyau

Étant donné $\{\mathbf{H}^{(b)} = \text{diag}[\mathbf{H}_1^{(b)} \dots \mathbf{H}_P^{(b)}], b = 1 \dots B_0\}$.

Pour $p' = 1 \dots P$

- Pour $b = 1 \dots B, p = 1 \dots P, p \neq p'$
 1. $\{\tilde{\mathbf{H}}^{(b)} = \text{diag}[\tilde{\mathbf{H}}_1 \dots \mathbf{H}_{p'}^{(b)} \dots \tilde{\mathbf{H}}_P], b = 1 \dots B_{p'}\}$.
 2. Classifier les données en appliquant le mean shift avec le nouvel estimateur (equation 5).
 3. Représenter paramétriquement chaque classe à l'aide d'une loi normale.
 4. Associer chaque point $\mathbf{x}^{(i)}$ à la distribution de la classe à laquelle il appartient.
 - Pour chaque point $\mathbf{x}^{(i)}$
 1. Trouver la distribution la plus stable parmi toutes les échelles pour le domaine p' étudié.
 2. La meilleure taille de noyau pour $\mathbf{x}^{(i)}$ est égale à la covariance de la distribution la plus stable.
-

4 Résultats

4.1 Segmentation d'image

Dans cette partie nous présentons des résultats de notre algorithme appliqué à la segmentation d'image couleur. L'ensemble des données est l'ensemble des pixels de l'image. A chaque point est associé un vecteur de 5 dimensions : 2 dimensions pour la position et 3 dimensions pour la couleur. Nous considérons 5 domaines $P = 5$ de chacun une dimension (les dimensions sont toutes indépendantes). Tous les résultats de segmentation d'image ont été obtenus avec 9 noyaux prédéfinis, leurs tailles variant entre 3 et 20. Les résultats sont comparés avec la méthode décrite dans la section 2 en faisant simplement varier les différents domaines conjointement. La couleur de chaque pixel des images résultats correspond à la couleur de son mode associé.

Les premiers résultats sont présentés sur l'image d'un chalet (figure 1). La figure 2 montre les classifications mean shift avec l'estimateur balloon obtenues en utilisant les tailles de noyaux résultant de l'algorithme de Comaniciu de la section 2 et de notre algorithme itératif. Avec l'algorithme de Comaniciu la classification donne 35 classes contre 37 avec notre algorithme. La différence est principalement visible au niveau des montagnes ¹. Les tailles de noyau obtenues avec l'algorithme de Comaniciu sont plus grandes ce qui entraîne une perte de certains détails.

Sur l'image d'un village (figure 3, résultats de segmentation sur la figure 4), 136 classes sont obtenues avec l'algorithme de Comaniciu de la section 2 contre 525 avec notre algorithme. Les segmentations obtenues montrent que notre algorithme est capable de conserver d'importants détails, qui sont perdus avec l'algorithme initial.

¹Les résultats sont plus visibles sur l'écran que sur une version imprimée de cet article



FIG. 1 – Image d’un chalet.



FIG. 2 – Résultats de classification mean shift avec l’estimateur balloon sur l’image d’un chalet pour a) l’algorithme de sélection de la section 2 et b) notre algorithme itératif.



FIG. 3 – Image d’un village.

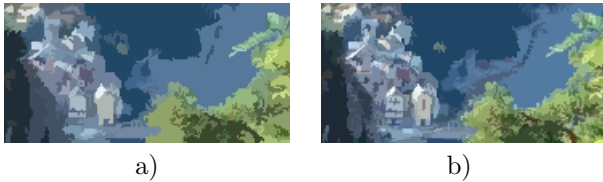


FIG. 4 – Résultats de classification mean shift avec l’estimateur balloon sur l’image d’un village pour a) l’algorithme de sélection de la section 2 et b) notre algorithme itératif.

4.2 Segmentation de mouvement

Nous validons maintenant notre méthode pour la segmentation couleur et mouvement [2]. Pour des raisons de coût de calcul, nous ne nous concentrons que sur une grille de pixels uniformément répartie sur tous les pixels en mouvement de l’image, c’est-à-dire sur les pixels n’appartenant pas au mouvement de la caméra. Chacun de ces pixels est décrit par sa position, sa couleur et son mouvement obtenu par flot optique, $P = 3$. Les tailles prédéfinies sont basées sur la taille de la grille pour la position et sur le bruit pour la couleur et le mouvement. La séquence résultat présentée est une séquence de tennis. Les classes contenant moins de 10 pixels ont été supprimées. On remarque que deux classes (une pour le joueur, une pour la raquette) ont été obtenues pour la première image et une classe (pour le joueur) pour la deuxième.

5 Conclusion

Dans ce papier nous avons présenté un algorithme pour la sélection de la taille des noyaux dans un espace multidimensionnel hétérogène. Il s’agit d’une méthode itérative

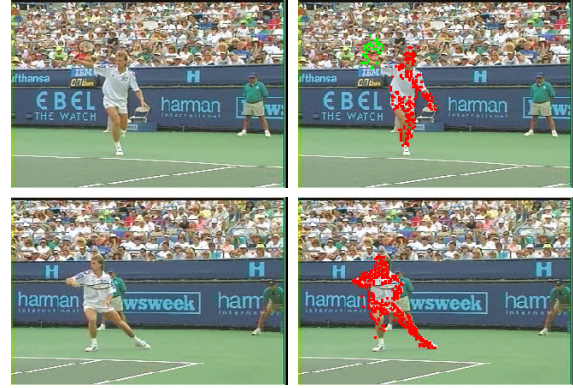


FIG. 5 – Segmentation de mouvement sur une séquence de tennis. Images 31, 212. A gauche les images initiales, à droite le résultat de la segmentation.

permettant de trouver successivement la meilleure taille pour chacun des domaines composant l’espace. Cette procédure permet une recherche plus riche de la taille optimale que l’algorithme original de Comaniciu [4]. Le caractère itératif de notre méthode a nécessité l’introduction d’un nouvel estimateur. Il serait intéressant d’étudier plus en détail ses propriétés statistiques. L’approche a été illustrée sur deux problèmes d’analyse d’image bas-niveau. Notons que pour de nombreuses applications, l’ensemble des tailles prédéfinies peut directement être obtenu à partir de pré-traitements sur les images. Dans ce cas l’algorithme devient totalement non supervisé.

Références

- [1] A. Bugeau and P. Perez. Bandwidth selection for kernel estimation in mixed multi-dimensional spaces. *Technical report, IRISA*, (PI 1852), 2007.
- [2] A. Bugeau and P. Perez. Detection and segmentation of moving objects in highly dynamic scenes. *Proc. Conf. Comp. Vision Pattern Rec.*, 2007.
- [3] C. Christoudias, B. Georgescu, and P. Meer. Synergism in low level vision. *Proc. Int. Conf. Pattern Recognition*, 04 :40150, 2002.
- [4] D. Comaniciu. An algorithm for data-driven bandwidth selection. *IEEE Trans. Pattern Anal. Machine Intell.*, 25(2) :281–288, 2003.
- [5] D. Comaniciu and P. Meer. Distribution free decomposition of multivariate data. In *SSPR/SPR*, pages 602–610, 1998.
- [6] D. Comaniciu and P. Meer. Mean shift : A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Machine Intell.*, 24(5) :603–619, 2002.
- [7] S. Sheather and M. Jones. A reliable data-based bandwidth selection method for kernel density estimation. *J. Royal Statist. Soc.*, 53 :683–690, 1991.
- [8] G. Terrell and D. Scott. Variable kernel density estimation. *The Annals of Statistics*, pages 1236–1265, 1992.
- [9] M. P. Wand and M. C. Jones. Multivariate plug-in bandwidth selection. *Comput. Statist.*, 9 :97–116, 1994.
- [10] M. P. Wand and M. C. Jones. *Kernel Smoothing*. Chapman & Hall, London, U.K., 1995.