

Fusion et classification d'images multi-sources par SVM avec sélection des caractéristiques dans l'espace à noyau

SU RUAN¹, NAN ZHANG^{1,2,3}, STEPHANE LEBONVALLET¹, QINGMING LIAO², YUEMIN ZHU³

¹ Laboratoire CReSTIC

9 rue de Québec, BP 00000, 10026 Troyes Cedex, France

² Image processing laboratory

Department of Electronic Engineering, Tsinghua University, 100084 Beijing, China

³ CREATIS

INSA de Lyon, 69621 Lyon cedex, France

¹ su.ruan@univ-reims.fr, zhangnan@image.ee.tsinghua.edu.cn, stephane.lebonvallet@univ-reims.fr

² liaoqm@tsinghua.edu.cn

³ Yue-Min.Zhu@creatis.insa-lyon.fr

Résumé - L'objectif de cette étude porte sur la classification d'une même scène observée par différents types d'images, ce qui génère de grandes quantités de données à traiter. Nous avons donc choisi d'utiliser la classification SVM (Séparateur à Vaste Marge) qui est réputée pour traiter des données de grande dimension. Bien que les différentes sources d'information permettent de donner des informations complémentaires pour relever des ambiguïtés, en même temps elles introduisent certaines informations redondantes. Notre idée pour la fusion de ces données repose sur l'extraction de l'information utile sur ces données pour une classification efficace. La sélection des caractéristiques les plus discriminantes est effectuée dans l'espace à noyau du SVM, car la sélection peut se faire linéairement dans cet espace. Cela permet en même temps de réduire la dimension des données afin d'obtenir une meilleure classification. Les critères de sélection sont basés sur la séparabilité des classes. Nous proposons un système basé sur la classification SVM avec la sélection des caractéristiques pour classer une tumeur cérébrale à partir de trois types d'images IRM 3D. Notre système permet de suivre l'évolution d'une tumeur tout au long d'un traitement thérapeutique.

Abstract - This paper concerns the classification of a scene observed by different types of images, that generates large amounts of data to be processed. We have chosen to use the classification SVM (Support Vector Machine) which is efficient to deal with large data. Although different sources of information can provide complementary information to lift up ambiguities, at the same time they introduce some redundant information. Our idea to fuse these data is to extract the useful information from these data to obtain an effective classification. The selection of which features are most discriminative is carried out in the kernel space of SVM, because the selection can be done linearly in this space. This also helps to reduce the size of data to get better a classification. The selection criteria are based on the separability of classes. We propose a system based on SVM classification with the selection of characteristics to classify a brain tumor from three types of 3D MRI images. Our system can follow up the evolution of a tumor during a therapeutic treatment.

1 Introduction

Dans le cas de la segmentation d'un objet à partir de multi-types d'images, une fusion des informations est nécessaire pour améliorer les résultats de la segmentation. Surtout dans le domaine médical, plusieurs modalités d'images sont obligatoires pour obtenir un diagnostic correct. De manière générale, la fusion peut se faire à plusieurs niveaux différents: au niveau de l'extraction des caractéristiques qui combine les caractéristiques avant d'appliquer le classificateur [1], au niveau de la classification qui fusionne les résultats de chaque classificateur [2][3][4], au niveau de décision qui prend en compte la décision de chaque classificateur [5][6]. Chaque méthode a ses avantages selon le contexte d'application. Les différentes sources d'information permettent de donner des informations complémentaires pour relever certaines ambiguïtés, mais en même temps elles

introduisent certaines informations redondantes. Nous proposons donc une méthode de fusion au niveau de l'extraction des caractéristiques qui sélectionne les caractéristiques les plus discriminantes pour le classificateur à partir de l'ensemble des différents types d'images. On se trouve confronté à la disponibilité d'une masse importante d'informations à partir de laquelle les caractéristiques utiles seront extraites. Nous avons choisi d'utiliser la classification SVM [7] qui est réputée pour traiter des données de grande dimension. Notre idée repose sur la sélection des caractéristiques les plus discriminantes dans l'espace à noyau du SVM. Cela permet en même temps de réduire la dimension des données afin d'obtenir une meilleure classification. PCA (Principal Components Analysis) est la méthode classique de réduction de dimension qui offre l'avantage de la facilité d'interprétation des résultats [8]. Toutefois l'hypothèse de linéarité n'est pas toujours réaliste dans

le cas de la classification. Dans les récents travaux de Wang [9], l'auteur a proposé un critère de sélection dans l'espace à noyau permettant de choisir les caractéristiques les plus discriminantes pour séparer les classes de manière linéaire. Un système de suivi s'inspirant de ces travaux est proposé ici. Notre système permet de suivre l'évolution d'une tumeur tout au long d'un traitement thérapeutique à partir de trois types d'images IRM 3D.

2 Fusion et classification

Pour segmenter l'objet observé par plusieurs types d'images, la méthode de classification par SVM est utilisée. Un vecteur de caractéristique est d'abord construit pour chaque point à partir des intensités des points dans une fenêtre centrée sur le point courant, la moyenne et l'écart-type des intensités, ainsi que les coefficients des ondelettes [11] de cette fenêtre. Comme c'est la même scène observée par plusieurs types d'images, on a donc un vecteur de ce type par type d'image. Au total, la dimension du vecteur de caractéristique à chaque point est multipliée par le nombre de types d'images utilisés, ce qui peut rendre la dimension très grande. La sélection des caractéristiques à partir de ce vecteur est ensuite effectuée, permettant d'extraire un certain nombre parmi eux selon un critère présenté par la suite. En utilisant seulement les caractéristiques sélectionnées, le classificateur décide finalement si un point appartient à l'objet à segmenter.

2.1 Sélection des caractéristiques

La sélection des caractéristiques consiste à trouver p caractéristiques à partir de d ($d \gg p$) caractéristiques d'origine selon certains critères permettant de diminuer les erreurs de classification quand seulement ces p caractéristiques sont utilisées. Le problème devient donc de trouver la solution en optimisant des critères de séparation de classes. Transformer le vecteur de caractéristique dans l'espace à noyau peut être écrit par l'équation suivante :

$$k(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle \quad (1)$$

où $\mathbf{x}_i, \mathbf{x}_j$ sont des données d'entrée, κ est une fonction noyau, $\phi(\cdot)$ est une transformation entre l'espace d'entrée et l'espace à noyau. La sélection des caractéristiques est réalisée dans l'espace à noyau au lieu d'utiliser les données d'entrée directement comme dans les méthodes classiques afin d'éviter la non séparabilité des celles-ci. Deux mesures sont d'abord calculées : la matrice de dispersion de chaque classe S_w et celle entre les classes S_B . Leurs déterminants sont définis comme suites [9] :

$$\text{tr}(S_w^\phi) = \text{tr} \left[\sum_{i=1}^c \sum_{j=1}^{m_i} (\phi(x_{ij}) - \mathbf{m}^\phi)(\phi(x_{ij}) - \mathbf{m}^\phi)^T \right]$$

et

$$\text{tr}(S_B^\phi) = \text{tr} \left[\sum_{i=1}^c n_i (\mathbf{m}_i^\phi - \mathbf{m}^\phi)(\mathbf{m}_i^\phi - \mathbf{m}^\phi)^T \right] \quad (2)$$

où c est le nombre de classes, n_i est le nombre de points de la classe i . \mathbf{m} représente le vecteur de moyenne. Le critère de séparabilité entre classes est défini par le ratio :

$$J^\phi = \frac{\text{tr}(S_B^\phi)}{\text{tr}(S_w^\phi)} \quad (3)$$

Le noyau gaussien est utilisé dans notre étude, qui est défini par :

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right) \quad (4)$$

où σ est la largeur du noyau gaussien. Le paramètre σ est identique que celui du SVM utilisé pour la classification par la suite. Notons α un vecteur ayant la même dimension que le vecteur de caractéristiques, et n'ayant comme composantes que les valeurs 0 ou 1. La sélection des caractéristiques est donc représenté par $\mathbf{x}(\alpha) = \alpha \otimes \mathbf{x}$. Le noyau gaussien devient la forme suivante :

$$k(\alpha \otimes \mathbf{x}, \alpha \otimes \mathbf{z}) = \exp \left[-\sum_{i=1}^d \frac{\alpha_i^2 (\alpha_i - \alpha_i)^2}{2\sigma^2} \right] \quad (5)$$

En théorie on peut maximiser (3) pour trouver directement le vecteur α . Mais le problème est que ce critère dépend du paramètre de noyau, ce qui influence fortement le choix des caractéristiques. Une solution est proposée dans [9] qui considère σ comme faisant partie du paramètre à optimiser. Notons $\eta_i = \alpha_i^2 / 2\sigma^2$, η_i : $i^{\text{ème}}$ composant du vecteur η . Finalement la sélection des caractéristiques est équivalente à trouver le vecteur η qui maximise le critère (3). La méthode du gradient conjugué est utilisée pour l'optimisation. Une grande valeur η_i signifie une importante contribution de la i -ème caractéristique à la séparation. En seuillant le vecteur η on peut retrouver le vecteur α qui représente les caractéristiques les plus importantes. Le vecteur η peut être initialisé de manière aléatoire avec de faibles valeurs qui sont proches de zéro.

2.2 Classification SVM

Le principe de SVM, basé sur le théorème de l'apprentissage statistique, consiste tout d'abord à transformer les entrées dans un espace de plus grande dimension à l'aide du noyau, et puis à chercher une frontière de décision linéaire dans cet espace [7]. Le noyau utilisé dans l'étape précédente est celui du SVM. Cela permet d'éviter d'ajuster le paramètre σ . L'ajustement de celui-ci est souvent une étape fastidieuse. Etant donné un ensemble d'apprentissages $\{(x_i, y_i)\}_{i=1}^n$, où y est l'étiquette de décision ($y=1$ pour la classe 1, et -1 pour la classe 2), la décision de la fonction de l'hyperplan est définie comme suite [7] :

$$f(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^n a_i y_i k(\mathbf{x}, \mathbf{x}_i) + b \right) \quad (6)$$

Le vecteur a et le coefficient b sont obtenus par l'apprentissage. L'implémentation LIBSVM [10] est utilisée ici pour l'apprentissage.

3 Système de suivi

Nous proposons d'un système de suivi, basée sur la classification présentée précédemment, pour quantifier l'évolution d'une tumeur sous un traitement thérapeutique. Les images d'examen, comprenant l'IRM pondéré en T2, Rho (Densité de Proton) et Flair de taille 512x512 voxels pour 24 coupes chacun, sont obtenues tous les quatre mois. L'apprentissage manuel est effectué seulement une fois sur un type d'image du premier examen (Image Flair dans notre system) avec 60 points sélectionnés aléatoirement dans la zone tumorale et 60 en dehors de la zone. L'apprentissage est ensuite adapté automatiquement pour les examens suivants à partir des résultats précédents. Le vecteur de caractéristiques pour chaque point est noté $S_r = f(x; \mu, \sigma^2, c)$ pour le type image r ($r = T2, Rho$ ou FLAIR). Nous avons d'abord testé l'influence de la taille de fenêtre sur la sélection. Cette taille fixe la dimension du vecteur de caractéristiques. Le tableau 1 montre qu'à partir de la taille 11x11, le nombre de caractéristiques sélectionnées devient stable. Nous avons donc choisi cette taille pour notre application. Avec la taille de fenêtre 11x11, le vecteur de caractéristiques S_r contient 11x11 valeurs d'intensités, la moyenne et l'écart type de ses intensités, et 144 coefficients d'ondelettes calculés selon [11]. Cela fait total $3 \times 267 (=801)$ caractéristiques pour les trois types d'images utilisés sur un point. Nous devons traiter $512 \times 512 \times 24$ points dans notre application. D'où l'importance de réduire la dimension du vecteur de caractéristiques. Les caractéristiques d'un point à l'intérieur de la région tumorale issue des trois types d'images sont normalisées et illustrées sur la Figure 1. La figure 2 montre le vecteur a trouvé par notre méthode qui représente les caractéristiques sélectionnées parmi les 801. 20% de caractéristiques sont finalement sélectionnées, ce qui permet de gagner 50% de temps de calcul. On voit que les caractéristiques les plus discriminantes se trouvent sur les images T2 et Flair. Cela est conforme à l'aspect visuel.

Après la classification SVM, il est possible que certains points à l'extérieur de la région tumorale sont manquantes. Pour les retrouver, un critère fondé sur la distance d (entre le point et le contour) et la probabilité l du point courant p appartenant à la tumeur est proposée. Supposons que la distribution de l'intensité de la région tumorale obéit à une distribution normale. En utilisant les résultats de classification, on peut calculer facilement la moyenne et l'écart type de la distribution normale. La décision du point p appartenant à la tumeur est définie comme:

$$\sum_{i=1}^3 I_i(p) > th \Rightarrow p \in tumor \quad (7)$$

avec le seuil $th = [1 + 0.1 \times (d - 1)] \times t$.

où d indique la distance la plus courte entre le point et le contour. t est une constante qui est égale à 0.6 dans notre expérience. i représente le type d'image.

Pour montrer l'efficacité de notre méthode de fusion, nous avons effectué une étude de comparaison entre les résultats obtenus avec et sans la sélection des caractéristiques en utilisant les tracés manuels des experts comme les « ground truth ». Cette étude montre que la fusion par l'étape de sélection permet d'améliorer nettement les résultats de classification (Tab.2). Les erreurs ont été divisées par 3.

Notre système de suivi a été testé sur deux patients pendant un peu plus d'un an. L'intervalle des examens est environ 4 mois. Les variations volumiques du suivi sur les deux patients sont calculées quantitativement (Tab.3). Une diminution de la taille de tumeur peut être observée pour le premier patient au cours du temps, par contre la situation n'est pas stable pour le deuxième patient. Les mesures quantitatives des tailles volumiques de tumeurs au cours du temps permettent de fournir les informations très importantes pour évaluer les traitements thérapeutiques en cours.

Tab 1 : nombre de caractéristiques sélectionnées en fonction de la taille de fenêtre utilisée

| Taille | 5x5 | 7x7 | 9x9 | 11x11 | 13x13 | 15x15 |
|--------|-----|-----|-----|-------|-------|-------|
| Nombre | 6 | 9 | 13 | 26 | 28 | 29 |

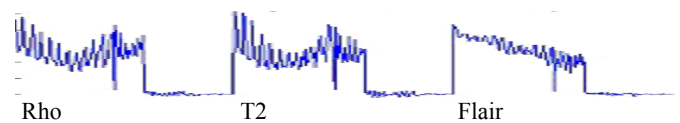


Figure 1 : Le vecteur des caractéristiques d'un point dans la région tumorale.

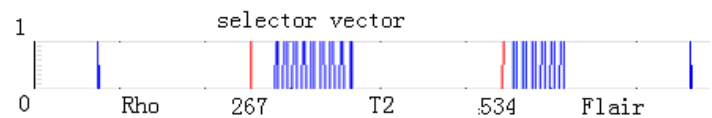


Figure 2 : le vecteur a trouvé qui représente les caractéristiques sélectionnées avec la valeur = 1, sinon = 0

Tab 2 : Trois mesures sont utilisées pour comparer les performances de classification entre l'utilisation de l'étape de la sélection et sans l'utilisation : VP (vrai positif), FN (faux négatif) et FP (faux positif).

| Mesures (%) | VP | FN | FP | Total |
|----------------|------|-----|------|-------|
| Sans sélection | 79.1 | 8.3 | 17.8 | 26.1 |
| Avec sélection | 96.3 | 5.4 | 2.8 | 8.2 |

Tab 3 : Les variations volumiques des tumeurs entre deux examens. Le symbole '-' signifie la diminution, '+' augmentation.

| Variation volumique (%) | E2-E1 | E3-E2 | E4-E3 | E5-E4 |
|-------------------------|-------|-------|-------|-------|
| Patient 1 | -13.4 | -6.4 | -3.2 | -6.5 |
| Patient 2 | 17.7 | -2.9 | 9.5 | x |

La figure 3 et la figure 4 montrent les résultats de la segmentation du patient 1 et du patient 2 respectivement sur une coupe des trois types d'IRM. Le contour de la tumeur du deuxième patient est plus complexe (irrégulier) que celui du premier patient. Toutefois, notre système permet de segmenter la tumeur correctement. Ces résultats ont été confirmés par des experts, ce qui montre l'efficacité du système proposé.

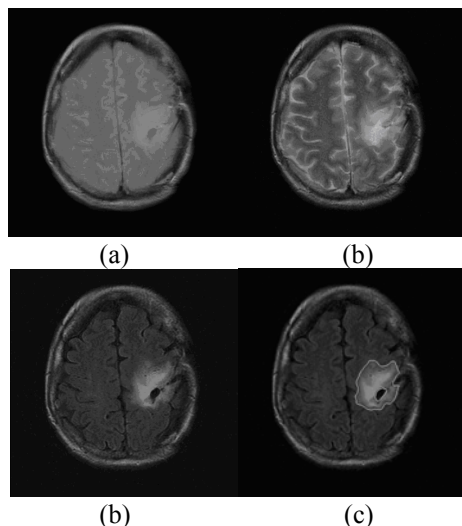


Figure 3 : Résultats de segmentation du patient 1 : les images Rho (a) T2 (b), et Flair (c) et le contour de la tumeur obtenu superposé sur l'image Flair.

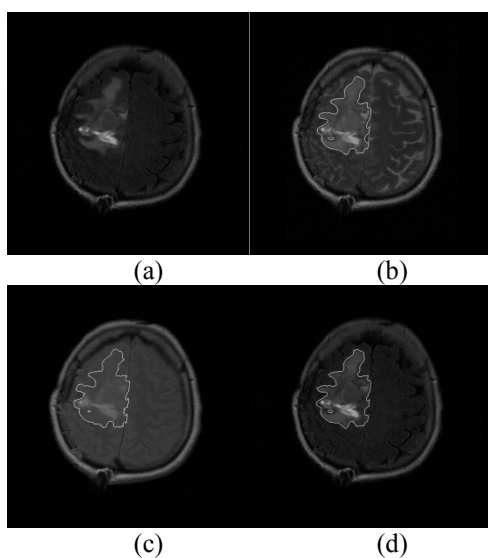


Figure 4 : Résultats de segmentation du patient 2 : une image Flair (a), les résultats de classification correspondants montrés sur les images T2 (b), Rho (c) et Flair (d).

4 Conclusion

Nous avons proposé une approche de fusion de données de la manière suivante : les données sont projetées dans un espace de grande dimension à l'aide d'un noyau dans lequel la sélection des caractéristiques les plus discriminantes est effectuée en utilisant les

critères des dispersions intrinsèques et extrinsèques des classes. Les caractéristiques sélectionnées sont ensuite utilisées par un SVM dont le noyau est le même que pour l'étape de sélection. Cette approche a été utilisée sur des images IRM cérébrales multi-séquences pour suivre l'évolution des tissus tumoraux au cours des traitements thérapeutiques. Les résultats obtenus montrent qu'elle permet de réduire non seulement le temps de calcul mais aussi les erreurs de classification grâce à la pertinence des données sélectionnées.

5 Références

- [1] X. Jing, Y. Yao, and D. Zhang. "Face and palmprint pixel level fusion and kernel DCV-RBF classifier for small sample biometric recognition", *Pattern Recognition*, 40(11):3209–3224, 2007.
- [2] A. Ross, K. Nandakumar, and A. Jain. *Handbook of Multibiometrics (International Series on Biometrics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
- [3] W. B. Dou, S. Ruan, Y. P. Chen, D. Bloyet, and J.-M. Constans, "A framework of fuzzy information fusion for the segmentation of brain tumor tissues on MR images", *Image and Vision Computing*, vol.25 (2), pp.164-171, Feb. 2007.
- [4] J.L. Marroquin, B. C. Vemuri, S. Botello, F. Calderon, and A. Fernandez-Bouzas, "An accurate and efficient Bayesian method for automatic segmentation of brain MRI", *IEEE Trans. Med. Imaging*, vol.21(8). pp.934–945, Aug. 2002.
- [5] L. Kuncheva, J. Bezdek, and R. Duin, "Decision templates for multiple classifier fusion: an experimental comparison", *Pattern Recognition*, 32(2):299–314, 2001.
- [6] J.L. Marroquin, B. C. Vemuri, S. Botello, F. Calderon, and A. Fernandez-Bouzas, "An accurate and efficient Bayesian method for automatic segmentation of brain MRI", *IEEE Trans. Med. Imaging*, vol.21(8). pp.934–945, Aug. 2002.
- [7] V. Vapnik, "The nature of statistical learning theory", Springer, 1995.
- [8] T. Jolliffe, "Principal Component Analysis". Springer 1986.
- [9] L.Wang, "Feature Selection with Kernel Class Separability", *IEEE trans. PAMI*, pp. 1534-1546, 2008.
- [10] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for Support Vector Machines", <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.
- [11] S. Chaplot, and N. R. Jagannathan, "Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network", *Biomedical Signal Processing and Control*, vol.1 (1), pp.86–92, 2006.