

Segmentation d'images en couleurs par réseau de neurones : application au domaine végétal.

Philippe FOUCHER ^{1,2} , Paul REVOLLON ¹ , Bertrand VIGOUROUX ²

1- UMR SAGAH (Sciences AGronomiques Appliquées à l'Horticulture)

Institut National d'Horticulture

2 rue Le Nôtre

49045 ANGERS CEDEX 01

Téléphone 02 41 22 54 36

Télécopie 02 41 22 54 62

2 - LISA (Laboratoire d'Ingénierie des Systèmes Automatisés)

Institut Universitaire de Technologie d'Angers

4 boulevard Lavoisier

BP 2018

49016 ANGERS CEDEX

Téléphone 02 41 73 53 20

Télécopie 02 41 73 53 30

Adresses électroniques philippe.foucher@iut.univ-angers.fr

revollon@angers.inra.fr

bertrand.vigouroux@univ-angers.fr

Mots-clés

Images couleurs, segmentation, réseaux de neurones, application agronomique.

Résumé

La segmentation en deux classes (classe "végétal" et classe "substrat de culture") d'images en couleurs de plants de forsythias est réalisée à l'aide d'un réseau de neurones (perceptron à une couche cachée). La segmentation est nettement améliorée lorsque le vecteur d'entrée du perceptron comprend non seulement les trois composantes des pixels de l'image dans l'espace colorimétrique choisi, mais aussi la moyenne et l'écart-type de ces composantes au voisinage immédiat du pixel à classifier. Le nombre de neurones de la couche cachée est ajusté en fonction des résultats obtenus sur un ensemble d'apprentissage. La qualité de la segmentation est évaluée à l'aide de trois paramètres : la spécificité, la sensibilité et l'erreur globale, par comparaison avec une segmentation effectuée par un expert.

1 - Introduction

L'emploi de la couleur en segmentation d'images est relativement récent, et constitue un sujet de recherche encore en développement. La classification des méthodes de segmentation des images colorées peut être calquée sur la classification établie pour la segmentation des images en niveaux de gris [1], avec deux complications supplémentaires : (i) les résultats de la segmentation dépendent de l'espace colorimétrique utilisé, et (ii) lorsqu'une métrique est nécessaire pour réaliser la segmentation, son choix dépend non seulement de la précision souhaitée (comme pour les images en niveaux de gris) mais aussi de l'appréciation psychovisuelle de la distance entre deux couleurs [2].

Une des stratégies possibles pour la segmentation est celle de la classification de pixels. Les réseaux de neurones se prêtent bien à une telle classification [3], [4], [5]. Ils ont été utilisés, dans le domaine végétal, pour classifier les folioles d'après leur géométrie [6] ou pour évaluer la vitesse de croissance de plants de laitue [7]. Par contre, il semble que la segmentation d'un couvert végétal par réseau de neurones n'ait pas encore été envisagée. C'est aux premiers résultats obtenus dans cette voie qu'est consacrée la présente communication. Le réseau a pour fonction de classer les pixels de l'image selon deux catégories : la catégorie "végétal" (forsythia) et la catégorie "substrat de culture".

Bien sûr, d'autres stratégies de segmentation sont envisageables, parmi lesquelles deux ont été testées, sur les mêmes images de forsythias. La première consistait à déterminer un seuil de binarisation à partir de l'histogramme bi-dimensionnel, dans l'espace colorimétrique rouge-vert, par un algorithme de poursuite des minima locaux [8]. La deuxième mettait en œuvre une analyse discriminante des triplets (R,V,B) de chaque pixel de l'image [9]. Ces deux voies se sont avérées moins performantes [9] que celle exposée ici.

Le paragraphe 2 présente le réseau de neurones utilisé (il s'agit d'un perceptron multicouche), ainsi que la méthode d'apprentissage mise en œuvre (rétro-propagation du gradient). Le paragraphe 3 décrit le matériel utilisé et précise la dimension des images, les espaces colorimétriques utilisés, la nature des vecteurs d'entrée du réseau de neurones et la méthode d'évaluation de la qualité de la segmentation. Dans le paragraphe 4 les résultats obtenus sont analysés.

2 - Le perceptron

Quelques exemples de segmentation d'images colorées par réseaux de neurones ont été publiés récemment. Les réseaux utilisés étaient des réseaux de Hopfield, configurés à l'aide de l'histogramme des couleurs [3] ou des réseaux auto-organisés du type Kohonen [4], [5]. Nous proposons ici d'utiliser un autre modèle de réseau: le perceptron.

Le perceptron multicouche est une structure constituée d'une couche d'entrée, d'une ou plusieurs couches cachées, et d'une couche de sortie. Dans de nombreux cas [10], une couche cachée suffit pour obtenir du perceptron des résultats satisfaisants (figure 1).

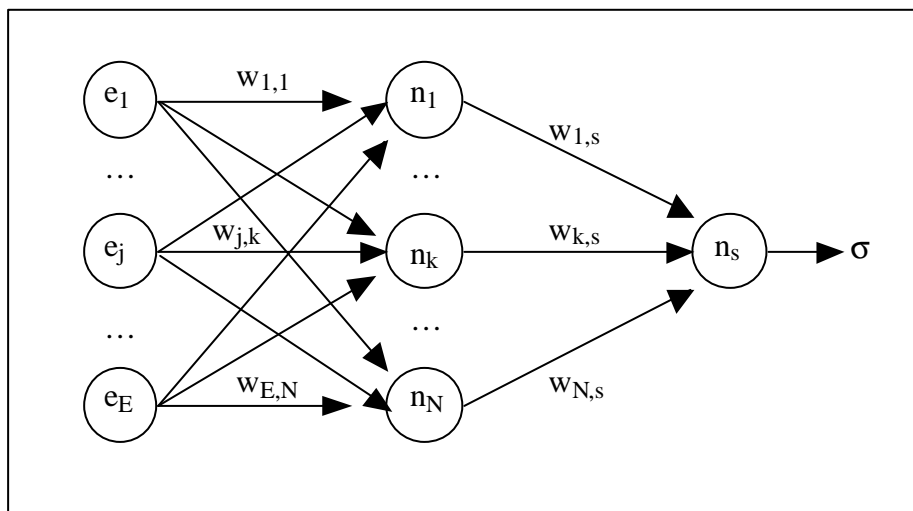


Figure 1 : perceptron à une couche cachée et un neurone de sortie.

L'entrée du perceptron est un vecteur \mathbf{e} de dimension E , et de composantes e_j où j est un entier de l'ensemble $\{1, \dots, E\}$. La couche cachée comporte N neurones ; le k -ième neurone de la couche cachée est désigné par n_k , où k est un entier de l'ensemble $\{1, \dots, N\}$. Le perceptron étant destiné à classifier des données suivant deux catégories, la couche de sortie ne contient qu'un neurone, noté n_s , qui fournit une sortie σ égale à 0 ou à 1. Chaque composante de \mathbf{e} est connectée aux N neurones de la couche cachée, et chaque neurone de cette couche est connecté au neurone de sortie. Les connexions sont affectées de poids synaptiques notés $w_{j,k}$ pour la connexion de la composante d'entrée e_j avec le neurone n_k , et $w_{k,s}$ pour la connexion entre le neurone n_k et le neurone de sortie n_s . Le signal de sortie s_k du neurone n_k s'écrit

$$s_k = f(x_k) \quad \text{avec} \quad x_k = \sum_{j=1}^E w_{j,k} e_j + \theta_k \quad (1)$$

Le signal de sortie s du neurone n_s s'écrit

$$\sigma = f(x_s) \quad \text{avec} \quad x_s = \sum_{k=1}^N w_{k,s} s_k + \theta_s \quad (2)$$

où les constantes θ_k et θ_s sont les polarisations des neurones n_k et n_s . Elles peuvent être considérées comme les poids synaptiques d'une entrée unité, ce qui permettra leur mise à jour lors de l'apprentissage. La fonction $f(u)$ est la fonction d'activation, qui doit être dérivable pour pouvoir mettre en œuvre un apprentissage par rétro-propagation du gradient. La fonction la plus couramment utilisée est la fonction sigmoïde

$$f(u) = \frac{1}{1 + \exp(-u)} \quad (3)$$

Les poids synaptiques du réseau de neurones sont ajustés au cours d'une phase d'apprentissage, à l'aide d'un lot de A vecteurs d'entrée, jugés représentatifs de l'application. L'ensemble de ces vecteurs, rangés sous la forme $[\mathbf{e}_1 \dots \mathbf{e}_i \dots \mathbf{e}_A]$ constitue une matrice \mathbf{M} de E lignes et A colonnes, appelée matrice d'apprentissage du perceptron. Une "époque" est la présentation, au réseau de neurones, de la matrice \mathbf{M} . Une époque est repérée par l'entier t appartenant à l'ensemble $\{1, \dots, Q\}$. A l'époque t , le neurone k de la couche cachée fournit A sorties $s_k^i(t)$, où i est un entier de l'ensemble $\{1, \dots, A\}$. De même, le perceptron fournit A sorties $\sigma_i(t)$. Un expert chargé d'opérer la classification fournit A réponses h_i , où les h_i appartiennent à l'ensemble $\{0, 1\}$. L'ajustement des poids synaptiques s'opère par minimisation du carré de l'erreur

$$\Delta_i(t) = \sigma_i(t) - h_i \quad (4)$$

considérée comme une fonction des poids synaptiques des connexions, par un algorithme de rétro-propagation du gradient [11], [12]. L'algorithme supporte un certain nombre de variantes. Dans celle que nous avons utilisée, les poids synaptiques sont initialisés de façon aléatoire. A l'époque t , pour chaque vecteur \mathbf{e}_i de l'ensemble d'apprentissage, les écarts de poids $\delta w_{k,s}^i(t)$ et $\delta w_{j,k}^i(t)$ sont calculés selon les formules

$$\delta w_{k,s}^i(t) = \eta \Delta_i(t) \frac{\partial \sigma_i(t)}{\partial x_s^i(t)} s_k^i(t) \quad ; \quad \delta w_{j,k}^i(t) = \eta \Delta_i(t) \frac{\partial \sigma_i(t)}{\partial x_k^i(t)} w_{k,s}^i(t) \frac{\partial s_k^i(t)}{\partial x_j^i(t)} e_j^i(t) \quad (5)$$

où η est le rythme d'apprentissage. A la fin de l'époque t , l'actualisation des poids synaptiques s'effectue à l'aide des formules

$$w_{k,s}(t+1) = w_{k,s}(t) + \sum_{i=1}^A \delta w_{k,s}^i(t) \quad ; \quad w_{j,k}(t+1) = w_{j,k}(t) + \sum_{i=1}^A \delta w_{j,k}^i(t) \quad (6)$$

La polarisation est mise à jour suivant le même principe. La procédure est itérée jusqu'à ce que l'erreur de classification devienne inférieure à une consigne fixée à l'avance, ou jusqu'à un nombre maximal Q d'époques autorisé. En fin d'apprentissage, les poids synaptiques et les polarisations sont définitivement fixés.

3 - Matériel et méthodes

Les images sont acquises en serre, au format RVB, au moyen d'une caméra tri-CCD en visée verticale de 24 plants de forsythias. Chacun des plans colorimétriques R (rouge), V (vert) et B (bleu) est numérisé sur 758×576 pixels, quantifiés sur 256 niveaux. Des 24 images ainsi obtenues sont extraites 16 imageries de résolution 180×180 pixels, constituant le lot d'apprentissage (figure 2). A chaque pixel des imageries du lot d'apprentissage peut être associé un vecteur d'entrée du réseau de neurones. Ce vecteur peut prendre en compte : ou bien les seules composantes du pixel dans l'espace colorimétrique choisi (alors chaque vecteur d'apprentissage a pour dimension $E = 3$) ; ou bien ces mêmes composantes colorimétriques, associées à des paramètres caractéristiques des pixels voisins (alors $E > 3$).



Figure 2 : les 16 imageries constituant le lot d'apprentissage.

Les composantes colorimétriques peuvent être les trois composantes R, V, B des images fournies par la caméra, ou tout autre ensemble de composantes obtenues par changement d'espace colorimétrique. Nous avons testé, outre l'espace RVB, l'espace HSI dans lequel les composantes colorimétriques H (teinte), S (saturation) et I (intensité) s'expriment selon [13] :

$$\begin{aligned}
H &= \cos^{-1} \frac{2R - V - B}{\sqrt{R^2 + V^2 + B^2 - (RV + VB + BR)}} \quad \text{si } V \geq B \quad ; \quad H = 2\pi - H \text{ sinon} \\
S &= 1 - \frac{3 \min(R, V, B)}{R + V + B} \\
I &= \frac{R + V + B}{3}
\end{aligned} \tag{7}$$

La prise en compte du voisinage de chaque pixel a été réalisée en ajoutant aux composantes colorimétriques leur moyenne et leur écart-type dans une fenêtre de dimension 7 x 7 pixels centrée sur le pixel d'intérêt. Chaque vecteur d'apprentissage a alors pour dimension $E = 9$.

Pour effectuer l'apprentissage, 10 pixels sélectionnés sur la première imagerie I_1 permettent de définir la première matrice d'apprentissage M_1 . A l'issue de ce premier apprentissage, les poids synaptiques et les polarisations du réseau de neurone sont fixés, et l'imagerie I_1 est présentée au réseau pour être segmentée. Le résultat est comparé avec une segmentation manuelle réalisée par un expert, pour en tirer une image d'erreur de segmentation. Dix pixels mal classés sont sélectionnés et ajoutés aux pixels du premier apprentissage, pour effectuer un second apprentissage. Le processus est itéré jusqu'à ce que l'erreur globale de segmentation se stabilise. Soit P_1 l'ensemble des pixels utilisés pour la dernière itération sur l'imagerie I_1 . On opère de même à partir des 15 autres images du lot d'apprentissage, puis on effectue un dernier apprentissage à l'aide des pixels de l'ensemble P défini par

$$P = P_1 \cup \dots \cup P_{16} \tag{8}$$

Le réseau de neurones est alors définitivement figé. Les 24 images globales lui sont ensuite présentées pour segmentation.

La qualité de la segmentation, jugée par comparaison avec une segmentation experte, est quantifiée à l'aide de 3 paramètres : la spécificité et la sensibilité, définies par Littman et Ritter [4], et l'erreur globale. Soient N_E l'ensemble des pixels appartenant au couvert végétal après segmentation experte, et N_N l'ensemble des pixels appartenant au couvert végétal après segmentation neuromimétique. On a

$$\text{spécificité} = 100 \frac{\text{cardinal}(N_E \cap N_N)}{\text{cardinal}(N_N)} \quad ; \quad \text{sensibilité} = 100 \frac{\text{cardinal}(N_E \cap N_N)}{\text{cardinal}(N_E)} \tag{9}$$

La spécificité précise, parmi les pixels assignés à la classe "végétal", le pourcentage de ceux correctement assignés à cette classe. La sensibilité précise, parmi les pixels appartenant effectivement à la classe "végétal", le pourcentage de ceux correctement assignés à cette classe. L'erreur globale de classification donne le pourcentage de pixels mal classés :

$$\text{erreur globale} = 100 \frac{\text{cardinal}(N_E \cup N_N) - \text{cardinal}(N_E \cap N_N)}{N_I} \tag{10}$$

où N_I désigne le nombre total de pixels de l'image.

4 - Résultats et discussion

4.1 - Phase d'apprentissage

Le perceptron décrit dans le paragraphe 2 présente, sur sa couche cachée, un nombre de neurones N non défini. La première étape du travail consiste à étudier l'influence de N sur la qualité de la segmentation. De même, le choix du nombre d'époques Q de la phase d'apprentissage doit être optimisé. Les résultats obtenus sont présentés dans le tableau 1, pour différentes valeurs de N et Q . La spécificité, la sensibilité et l'erreur

globale indiquées dans le tableau représentent la moyenne observée sur les 16 imagettes de l'ensemble d'apprentissage. Dans tous les cas, le rythme d'apprentissage du réseau a été fixé à la valeur $\eta = 0,9$. Les vecteurs d'entrée du réseau ont pour composantes les 3 composantes colorimétriques des pixels d'apprentissage dans l'espace RVB.

N	Q	Spécificité (%)	Sensitivité (%)	Erreur globale (%)
5	10	99,12	96,82	1,56
10	10	98,71	98,11	1,56
10	50	98,47	98,46	1,55
10	100	98,36	98,57	1,59
20	50	98,85	97,55	1,79

Tableau 1 : influence de N et Q sur la qualité de la segmentation.

Pour un nombre d'époques identique ($Q = 10$) les réseaux à 5 et à 10 neurones présentent la même erreur globale (1,56%). Cependant, le réseau à 5 neurones présente une plus grande disparité, entre spécificité et sensibilité, que le réseau à 10 neurones. Le réseau à 5 neurones ne sera donc pas retenu pour la suite de l'étude. De même, pour un nombre d'époques identique ($Q = 50$), le réseau à 20 neurones présente une erreur globale (1,79%) supérieure à l'erreur globale (1,55%) du réseau à 10 neurones. Seuls les réseaux à 10 neurones seront donc retenus pour la suite de l'étude.

A nombre de neurones identique ($N = 10$), les apprentissages à 10, 50 et 100 époques présentent des résultats sensiblement voisins. C'est cependant l'apprentissage à 50 époques qui présente la plus faible erreur globale (1,55%) et la plus grande similitude entre spécificité et sensibilité (98,47% et 98,46%). L'apprentissage sera donc effectué dorénavant sur 50 époques.

L'influence de l'espace colorimétrique (RVB ou HSI) et de la dimension des vecteurs d'entrée ($E = 3$ ou 9) sur les résultats de la segmentation apparaît dans le tableau 2. La prise en compte du voisinage de chaque pixel ($E = 9$) fournit une plus faible erreur globale dans chacun des deux espaces colorimétriques testés. C'est donc ce type de vecteur d'entrée que nous retiendrons par la suite, en adoptant l'espace RVB qui fournit une erreur globale de 1,55%, inférieure à celle obtenue dans l'espace HSI (1,83%).

Espace	E	Spécificité (%)	Sensitivité (%)	Erreur globale (%)
RVB	3	95,20	95,45	5,20
	9	98,47	98,46	1,55
HSI	3	98,49	95,58	2,98
	9	98,25	98,54	1,61

Tableau 2 : influence de la dimension du vecteur d'entrée et de l'espace colorimétrique sur la qualité de la segmentation.

Finalement, la segmentation sera effectuée sur un perceptron dont la couche cachée contient 10 neurones, l'apprentissage étant effectué sur 50 époques. Le vecteur d'entrée du réseau sera constitué des 3 composantes colorimétriques des pixels de l'image, dans l'espace RVB, en y ajoutant la moyenne et l'écart type de chacune de ces composantes, dans un voisinage de taille 7×7 pixels autour du pixel à classifier.

4.2 - Phase de segmentation

Les résultats de la segmentation des 24 images disponibles sont présentés dans le tableau 3, et un exemple de segmentation est fourni sur la figure 3 (image 6 du tableau 3).



(a) (b)



Figure 3: exemple d'image à segmenter (a) et résultat de la segmentation (b).

Dans tous les cas, l'erreur globale est très faible : elle s'échelonne de 0,71% à 1,75%, avec une moyenne de 1,06% sur l'ensemble des 24 images. Les spécificités et sensibilités obtenues sont également satisfaisantes, l'une et l'autre présentant une moyenne de 98,82% et 98,80% sur l'ensemble des 24 images.

5 - Conclusion

Ainsi est démontrée la pertinence de l'utilisation d'un perceptron à une couche cachée pour la segmentation d'images en couleurs de couverts végétaux. La prise en compte du voisinage de chaque pixel dans l'espace colorimétrique, pour définir le vecteur d'entrée du perceptron, accroît sensiblement les performances de la segmentation. Le choix du nombre d'époques optimal pour l'apprentissage du réseau n'est pour l'instant justifié que par les résultats obtenus. Nous pensons proposer ultérieurement une méthode automatisée du choix du nombre d'époques, permettant de s'affranchir de la série des images à segmenter. D'autres espaces colorimétriques que les espaces RVB et HSI pourraient être également essayés, en particulier les plus proches de la perception psychovisuelle, dans la mesure où la qualité de la segmentation est évaluée par comparaison avec une segmentation effectuée par un opérateur humain.

Une analyse en composantes principales [13] des images pourrait également permettre de choisir un espace colorimétrique plus favorable.

Image n°	Spécificité (%)	Sensitivité (%)	Erreur globale (%)
1	99,35	99,15	0,75
2	99,50	99,04	0,73
3	99,40	99,15	0,71
4	99,44	98,17	1,17
5	99,74	97,39	1,35
6	99,67	98,40	0,89
7	99,62	98,13	1,06
8	99,70	97,17	1,46
9	99,20	99,10	0,87
10	99,78	98,71	0,78
11	99,60	98,95	0,75
12	98,29	99,34	1,05
13	98,02	99,71	0,99
14	98,79	98,99	0,97
15	98,28	99,19	1,14
16	97,62	99,54	1,27
17	97,38	99,84	1,23
18	98,70	98,99	1,00
19	98,12	99,62	0,98
20	98,15	98,89	1,12
21	98,31	97,09	1,75
22	98,33	98,62	1,16
23	97,34	99,10	1,36
24	99,31	98,90	0,81

Tableau 3 : résultats de la segmentation.

Références

- [1] Cocquerez J.-P., Philipp S. : *Analyse d'images: filtrage et segmentation*. Masson, Paris (1995).
- [2] Lezoray O. : Segmentation d'images couleur par morphologie mathématique et classification de données par réseaux de neurones: application à la classification de cellules en cytologie des séreuses. *Thèse de doctorat*, Université de Caen, France (2000).
- [3] Campadelli P., Medici D., Schettini R. : Color Image Segmentation using Hopfield Networks. *Image and Vision Computing*, 1997, vol. 15, n° 3, pp. 161-166.
- [4] Littman E., Ritter E. : Adaptative Color Segmentation - A Comparison of Neural and Statistical Methods. *IEEE Transactions on Neural Networks*, 1997, vol. 8, n° 1, pp. 175-185.
- [5] Papamarkos N., Strouthopoulos C., Andreadis I. : Multithresholding of color and gray-level images through a neural network technique. *Image and Vision Computing*, 2000, vol. 18, pp. 213-222.
- [6] Oide M. : Discrimination of soybean leaflet shape by neural networks with image input. *Computers and electronics in agriculture*, 2000, vol. 29, pp. 59-72.
- [7] Zaidi M.A., Murase H., Honami N. : Neural network model for the evaluation of lettuce plant growth. *Journal of Agricultural Engineering Research*, 1999, vol. 74, pp. 237-242.

- [8] Mougin S. : Segmentation d'images couleur appliquée à un couvert végétal. *Rapport de stage ingénieur systèmes électronique et informatique industrielle*, Université de Nantes, IRESTE (1999).
- [9] Revollon P., Foucher P., Boumaza R. : Etude comparée de trois techniques de segmentation d'images de végétaux : seuillage d'histogramme, analyse discriminante et réseaux de neurones. *Communication soumise aux XXIIIèmes Journées de Statistiques, ENITIAA, Nantes, 14-18 mai 2001*.
- [10] Héroult J., Jutten C. : *Réseaux neuronaux et traitement du signal*. Hermès, Paris (1994).
- [11] Rumelhart D.E., Hinton G.E., Williams R.J. : Learning representations by back-propagating errors. *Nature*, 1986, vol. 323, pp. 533-536.
- [12] Vogl T.P., Mangis J.K., Rigler A.K., Zink W.T., Alkon D.L. : Accelerating the Convergence of the Back-Propagation Method. *Biological Cybernetics*, 1988, vol. 59, pp. 257-263.
- [13] Gonzalez R.C., Woods R.E. : *Digital Image Processing*. Addison-Wesley Publishing Company (1993).