

MGS2 : Optimisation multicritères de contours actifs par algorithmes génétiques

Nicolas CLADEL, Renaud SÉGUIER

Équipe SCEE (Supelec - IETR)

Supelec, avenue de la boulaie, BP 81127, 35511 Cesson-Sévigné Cedex, France

nicolas.cladel@supelec.fr, renaud.seguier@supelec.fr

Résumé – Dans cet article nous proposons une évolution de notre précédent travail sur l’optimisation multicritères de contours actifs. Cette approche permet une exploration globale de l’image et une gestion efficace de plusieurs énergies par la représentation de Pareto. Notre nouvel algorithme, le MultiObjective Genetic Snakes 2 (MGS2), associe l’algorithme multicritères NSGA2 [4] au codage des contours actifs des MGS [3]. Les MGS2 exploitent de nouvelles énergies d’attache à l’image afin d’assurer la convergence des contours vers un objet creux en environnement bruité à partir d’une initialisation aléatoire. Nous procédons ici à l’étude de ces énergies et à l’analyse de la convergence de l’algorithme à travers une série de tests sur une base d’objets synthétiques. Nous présentons également des résultats de l’application de MGS2 à la lecture labiale.

Abstract – We propose in this paper an evolution of our preceding work on the optimization multicriterion of active contours. This approach allows a global analysis of the image and an efficient use of several energy. Our new algorithm, MultiObjective Genetic Snakes 2 (MGS2) associates the multiobjective algorithm NSGA2 [4] the coding of active contours of the MGS [3]. The MGS2 exploit new energies of fastener to the image in order to ensure the convergence of contours towards a hollow object in disturbed environment starting from a random initialization. We carry out here the study of these energies and the analysis of the convergence of the algorithm through a series of tests on a basis of synthetic objects. We will also have results of the application of MGS2 at the labial reading.

1 Introduction

Notre équipe s’intéresse à l’analyse d’images en environnement réel pour des applications telles que la visiophonie et les interactions homme-machine. Nous nous focalisons dans cette étude sur la segmentation de lèvres en environnement réel. Nous souhaitons être peu dépendant des informations à priori sur la position des contours ou sur la teinte de l’objet. Compte tenu de la variabilité de la forme des bouches inter et intra locuteurs, nous avons fait le choix de ne pas prendre en compte un modèle à priori des contours recherchés. Dans un soucis de généralisation de l’application de notre algorithme, nous allons étendre le domaine d’application du MGS2 à d’autres objets. l’objectif de cette étude est donc l’extraction par contours actifs d’objets similaires aux lèvres, c’est à dire creux et peu convexes, dans un environnement bruité.

Le principe adopté est l’utilisation conjointe des énergies de type région et gradient afin d’obtenir une segmentation efficace en présence de bruit important. Nous avons proposé une telle approche de coopération énergétique [3] appliquée aux contours actifs dans le cadre de la lecture labiale. Cette méthode donnait de bons résultats mais était lente et complexe. Nous en présentons ici une amélioration en intégrant le formalisme des contours des MGS dans une structure génétique multicritères simple et rapide, le NSGA2 [4]. Ces améliorations nous permettent de gérer plus efficacement un plus grand nombre d’énergies. Nous proposons également de nouveaux potentiels d’attache à l’image que nous étudions à travers des tests sur une base d’objets synthétiques.

Cet article est découpé en trois parties. La première concerne l’adaptation des contours actifs aux algorithmes génétiques, la

seconde les énergies proposées et leur étude et la dernière décrit l’application des MGS2 à la lecture labiale.

2 Algorithme Multiobjective Genetic Snakes 2 (MGS2)

Cet algorithme associe le formalisme des contours actifs des MGS [3] à la structure génétique multicritères NSGA2. Les MGS permettent une exploration globale de l’image et une gestion efficace de plusieurs potentiels. Nous allons rappeler ici le codage et la déformation des contours actifs que nous effectuons dans MGS et MGS2.

Les double snakes sont représentés par un chromosome codé sous la forme d’un tableau de dimension $2 \times N$, N étant le nombre de nœuds d’un contour (figure 1). Un gène représente donc un nœud d’un contour et ne pourra prendre qu’un ensemble défini de valeurs dans l’image [1]. Les double contours découpent l’image en trois régions, la région centre à l’intérieur du contour interne, la région objet entre les contours interne et externe, et la région externe à l’extérieur du contour externe. Les contours ainsi modélisés sont initialisés aléatoirement puis

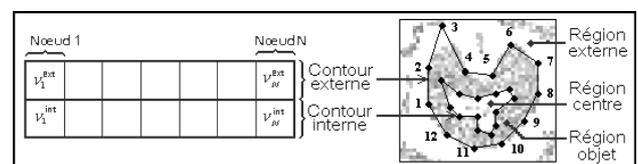


FIG. 1 – Codage des contours

vont être modifiés par les opérateurs génétiques de croisement et de mutation. La sélection d'une suite de gènes revient à découper une portion du contour (figure 2-a). Le croisement de deux contours P1 et P2 se traduit par l'échange de portions de contours (figure 2-b). Cette opération est appliquée simultanément aux contours interne et externe. La mutation d'un gène implique le déplacement d'un nœud (figure 2-c).

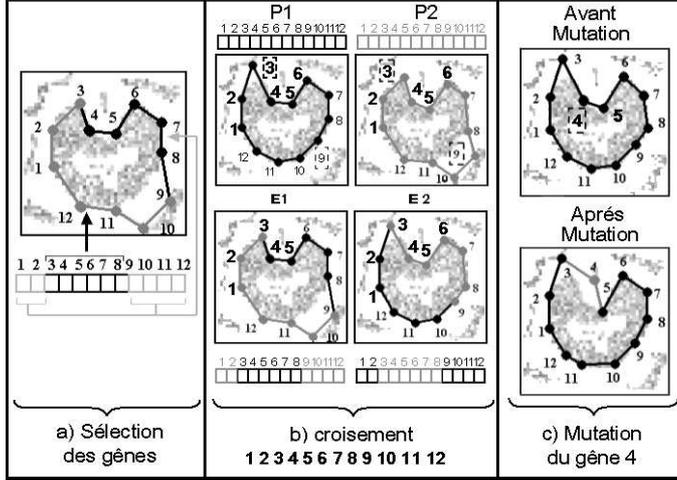


FIG. 2 – Déformation des contours par opérateurs génétiques

La représentation de Pareto attribue un rang [4] à chaque individu de la population en fonction de ses valeurs énergétiques. Les MGS fournissent ainsi en fin de traitement un ensemble de doubles contours potentiellement optimaux parmi lesquels un individu optimal doit être sélectionné en fonction des préférences de l'utilisateur.

3 Énergies

3.1 Définition des descripteurs

La segmentation doit prendre en compte de manière coopérative les énergies de type contour et de type région. Nous travaillons sur quatre images pour caractériser l'objet à extraire : l'image originale de l'objet (figure 3-a) et deux images issues des pré-traitements effectués sur cette image originale. La première est binaire (figure 3-b) et la seconde en niveaux de gris représente l'amplitude du gradient (figure 3-c). La quatrième image représente les contours orientés vis à vis du centre de l'image : les nœuds du contour externe sont attachés aux pixels noirs et ceux du contour interne aux pixels gris (figure 3-d). Nous proposons trois types de descripteur :

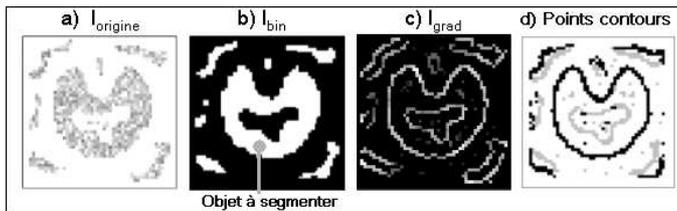


FIG. 3 – Images sources.

- **Un descripteur gradient** ($D_{gradient}$) proportionnel à l'amplitude du gradient le long des contours interne et externe

et au nombre de points contours appartenant aux courbes. Ce descripteur tend vers 0 lorsque les doubles snakes sont situées sur des pixels de forte amplitude gradient identifiés comme points contours :

$$D_{gradient}(C) = e^{-\frac{NbPtsContour(C)}{NbPtsTotal(C)} \sum_{\{x,y\} \in C} |\nabla I_{origine}(x,y)|}$$

avec :

- $C = C_{interne} \cup C_{externe}$
- $NbPtsContour(C)$: nombre de points contours appartenant aux contours interne et externe.
- $NbPtsTotal(C)$: nombre de pixels appartenant au contour.
- **Un descripteur variance** ($D_{variance}$ relatif à la variance des régions décrites par les contours interne et externe (figure 1) dans l'image originale. Ce descripteur tend ainsi vers 0 lorsque la région est homogène.

$$D_{variance} = \sigma^2(\Omega) = \frac{1}{|\Omega|} \int \int_{\Omega} (I_{origine}(x,y) - \mu(\Omega))^2 dx dy$$

avec :

- $|\Omega|$: taille de la région Ω courante (objet ou centre).
- $\mu(\Omega)$: moyenne des niveaux de gris des pixels appartenant à la région Ω .
- **Un descripteur d'agrégation** ($D_{agregation}$) calculé à partir de l'image binaire et proportionnel au nombre de pixels à 0 ou à 1 présents dans les régions décrites par les contours. Des coefficients pondérateurs permettent de privilégier l'ajout de pixels ayant le label de la région voulue par rapport à la suppression des autres pixels. Ce descripteur tend ainsi vers 0 pour une région constituée de nombreux pixels de la région recherchée :

$$D_{agregation}(\Omega) = NbPixelTotal - \left[\alpha_{agreg} \cdot NbPixels_{\Omega}^{label(\Omega)} - \alpha_{sup} \cdot NbPixels_{\Omega}^{label(\bar{\Omega})} \right]$$

avec :

- $NbPixelTotal$ est le nombre total de pixels de l'image : $NbPixelTotal = hauteur \cdot largeur$.
- $NbPixels_{\Omega}^{label(\Omega)}$ est le nombre de pixels ayant le label $label(\Omega)$ dans la région Ω .
- $label(\Omega)$ est le label associé aux régions d'intérêt :
$$\begin{pmatrix} \Omega \\ objet \\ centre \end{pmatrix} \rightarrow \begin{pmatrix} label \\ 1 \\ 0 \end{pmatrix}$$
- $\bar{\Omega}$ est le complémentaire $\bar{\Omega}$:
$$\begin{pmatrix} \bar{\Omega} \\ objet \\ centre \end{pmatrix} \rightarrow \begin{pmatrix} label \\ 0 \\ 1 \end{pmatrix}$$
- α_{agreg} et α_{sup} sont des coefficients pondérateurs. Si α_{agreg} est supérieur à α_{sup} , l'agrégation de pixels région sera privilégiée à la suppression des pixels non-région. Ces coefficients sont normés : $\alpha_{agreg} + \alpha_{sup} = 1$.

Les descripteurs de variance et d'agrégation induisent des comportements complémentaires : la variance permet de tendre vers une région compacte homogène alors que l'agrégation va autoriser la segmentation d'une région intégrant du bruit.

3.2 Étude des configurations énergétiques

Nous allons ici étudier les différentes configurations d'énergies possibles exploitées par les MGS2. Pour déterminer la configuration optimale, nous testons notre algorithme sur une base d'objets synthétiques. Cette base est constituée de différents objets et sources de bruits. Afin d'évaluer la qualité de la convergence induite par les configurations, nous calculons le nombre de pixels différents (N_{diff}) entre la surface de l'objet segmenté et celle de l'objet réel. La figure 4 montre l'évolution de cette valeur au cours du temps. Nous étudions ainsi la meilleure valeur (pointillés), la valeur moyenne (semi-pointillés) et la valeur correspondant à l'individu optimal (continu). Les résultats de la segmentation sont présentés dans la figure 5.

Les configurations énergétiques possibles peuvent prendre en compte cinq énergies : l'énergie gradient et les énergies région (variance ou agrégation) appliquées aux régions définies par le contour externe ou interne. Nous utilisons deux contours de 28 nœuds chacun. L'objet à segmenter est creux, situé au centre de l'image (figure 3).

Nous avons testé les cinq énergies ensemble (figure 4-E1). Les doubles snakes convergent vers une mauvaise segmentation car l'algorithme gère difficilement autant d'énergies. Pour analyser l'influence des deux descripteurs de type région, nous avons choisi deux configurations énergétiques : la première est composée du gradient et de l'agrégation appliqués aux régions centre et objet (figure 4-E2), la seconde est définie à partir de E2 en remplaçant l'agrégation par la variance (figure 4-E3). Nous observons que les valeurs optimales vers lesquelles tendent ces descripteurs au cours du temps ne correspondent pas à la segmentation cherchée. Ainsi E2 guide les doubles contours vers une région homogène restreinte (figure 5-R2) et E3 vers une région étendue au bruit voisin de l'objet (figure 5-R3). Afin de pallier à ces problèmes nous rajoutons à la configuration E3, le descripteur de l'agrégation appliqué au contour externe (figures 4-E4 et 5-R4). Les double snakes convergent bien vers l'objet et plus rapidement que dans les autres cas. La configuration énergétique que nous retiendrons est donc composée de quatre énergies : le gradient, la variance appliquée au contour interne, la variance sur le contour externe et l'agrégation sur ce même contour.

4 Application à la lecture labiale

Nous allons maintenant présenter l'application des MGS2 à la lecture labiale [6]. Les MGS2 sont ici associés à un opérateur de contours actifs variationnels développé dans [1] et implémentés en mode vidéo. Ces tests sont réalisés sur des séquences de prononciation de chiffres de la base M2VTS [5].

Le procédé de lecture labiale mis en œuvre recherche un mot unique dans une séquence. La base d'images M2VTS est composée de séquences de chiffres prononcées par dix personnes, avec quatre séquences par personne. Chaque séquence est prise à une semaine d'intervalle. L'apprentissage est réalisé sur 3 séquences et la reconnaissance est effectuée sur la quatrième.

Chacun des chiffres prononcé est associé à une signature composée d'informations sur la forme de la bouche, le nombre de pixels labellisés lèvres lors de pré-traitement région [3] et les dérivées de chacun de ces paramètres. Nous obtenons ainsi un vecteur de huit coordonnées pour chaque image. La classifi-

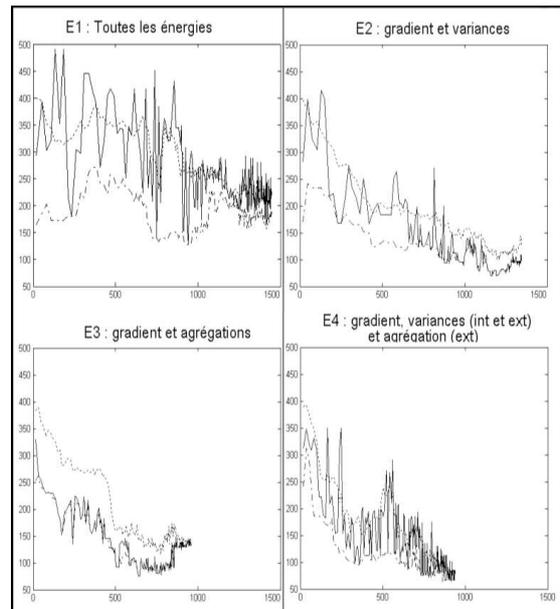


FIG. 4 – Courbes des tests énergétiques.

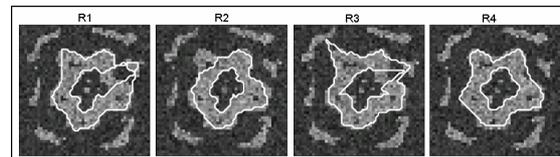


FIG. 5 – Contours des individus optimaux des tests énergétiques.

cation du mot est réalisée par la méthode du Dynamic Time Warping.

Les informations sur la forme de la bouche (hauteur et largeur des lèvres) sont extraites des contours segmentés en mode vidéo. Ce processus de segmentation exploite le MGS2 et un opérateur de contours actifs [1] au cours de deux étapes. La première étape est une phase d'initialisation (figure 6) : MGS2 est appliqué sur la première image de chaque séquence et le contour optimal obtenu est affiné par l'algorithme de contours actifs. La seconde étape est de type vidéo et utilise le résultat de chaque image pour initialiser la segmentation par contours actifs de l'image suivante (figure 7).

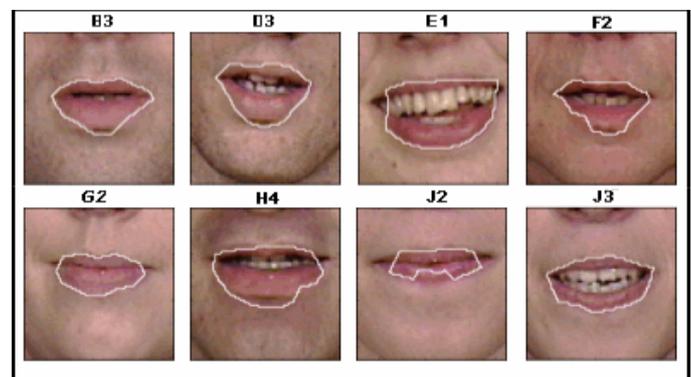


FIG. 6 – Exemples de segmentation d'images initiales de séquences de locuteurs

Nous obtenons ainsi 75% de bonne segmentation des images

initiales. Nous observons des imperfections au niveau des commissures des lèvres (E1-J2, figure 6). En effet les pixels appartenant à ces régions ne sont pas labellisés lèvre en raison de leur trop faible luminosité et sont donc rejetés par l'algorithme.

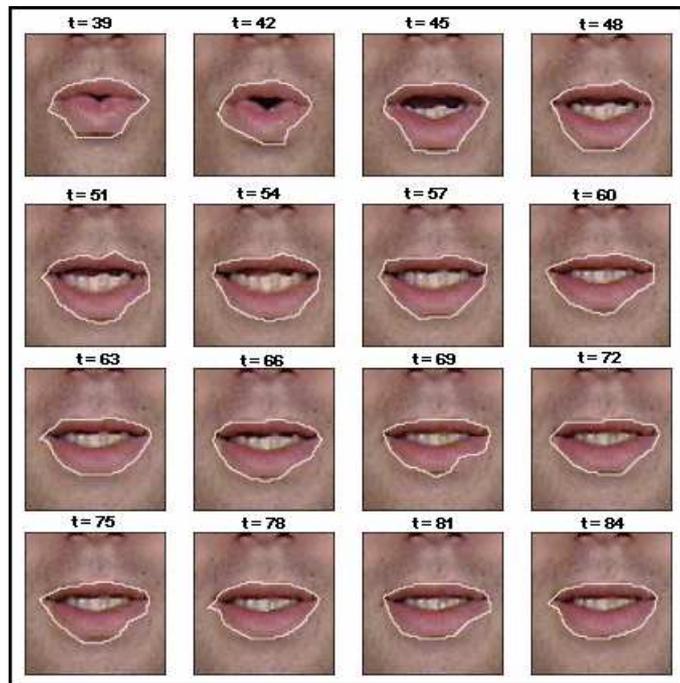


FIG. 7 – Extrait d'une séquence segmentée en mode vidéo

Le temps moyen de segmentation de l'étape d'initialisation est de 150 secondes sous matlab et celui de la seconde étape est de 25 secondes par image.

Nous obtenons alors une moyenne de 69% de bonne classification sur les dix personnes de M2VTS. Nous avons ainsi améliorés nettement nos précédents résultats qui étaient de 58% sur la première personne. En effet, notre ancienne méthode n'était pas assez robuste pour être appliquée sur les dix personnes.

5 Conclusion

Dans cet article nous avons présenté une nouvelle version de l'algorithme MGS [3] effectuant une optimisation multicritères de contours actifs par algorithmes génétiques. Les MGS2 permettent ainsi de segmenter plus rapidement un objet dans une image bruitée. Cet algorithme est également plus simple à paramétrer car il ne gère qu'une seule population de contours. Ces améliorations nous permettent aussi de manipuler des contours plus importants, 28 nœuds contre 8 pour les MGS.

Au cours de tests sur des objets synthétiques, nous avons pu observer l'influence des énergies d'attache à l'image sur la convergence des contours actifs et déterminer les paramètres énergétiques optimaux des MGS2. L'efficacité des MGS2 a été vérifiée à travers une application de lecture labiale.

Dans le cadre de travaux futurs, nous chercherons à améliorer la définition des énergies et à appliquer les MGS2 à l'extraction des contours cardiaques. Nous développerons également la spécialisation des MGS2 aux contours labiaux afin d'accélérer la segmentation des lèvres pour la lecture labiale, notamment

par l'ajout de modèles. Ces modèles devront être suffisamment flexibles pour s'adapter à la variabilité de forme des lèvres.

Remerciements

Cette étude a été soutenue par la région Bretagne.

Références

- [1] N. Cladel et R. Segulier. *Active contours multiobjective optimisation by hybrid algorithms*. Irish machine vision and image processing (IMVIP), 2004.
- [2] C.A. Coello Coello et G.T. Pulido. *Multiobjective optimization with a micro genetic algorithm*. Proceedings of the Genetic and Evolutionary Computation Conference (GECCO'2001), Morgan Kaufmann Publishers, San Francisco, California, pp 274–28, 2001.
- [3] Renaud Segulier et Nicolas Cladel. *Multiobjectives Genetic Snakes : Application on Audio-Visual Speech Recognition*. 4th EURASIP Conference focused on Video, Image Processing and Multimedia Communications, Zagreb, Croatia, 2003.
- [4] K. Deb. *A fast elitist non dominated sorting genetic algorithm for multiobjective optimization : NSGA II*. Parallel problem solving from nature - PPSN VI, Springer lecture notes in computer science, pp. 849-858, 2000.
- [5] S. Pigeon and L. Vandendrope. *The M2VTS Multimodal Face Database (Release 1.00)*. J. Bigun, G. Chollet, and G. Borgefors eds., Audio- and Video-Based Biometric Person Authentication, Springer, pp 403-409, 1997.
- [6] P. DAUBIAS and P. DELEGLISE. *Evaluation of an automatically obtained shape and appearance model for automatic audio visual speech*. Eurospeech, 2001.