

Segmentation interactive d'objets d'intérêt d'images par la coupe de graphes

Najd Abdmouleh, Victor Chen, Su Ruan

► **To cite this version:**

Najd Abdmouleh, Victor Chen, Su Ruan. Segmentation interactive d'objets d'intérêt d'images par la coupe de graphes. ORASIS - Congrès des jeunes chercheurs en vision par ordinateur, Jun 2011, Praz-sur-Arly, France. 2011. <inria-00596255>

HAL Id: inria-00596255

<https://hal.inria.fr/inria-00596255>

Submitted on 26 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Segmentation interactive d'objets d'intérêt d'images par la coupe de graphes

Interactive segmentation of interest objects image by graph cuts approach

Najd Abdmouleh
ISIM

Avenue de la Corniche
Monastir 5000 – Tunisie

najd.abdmouleh@gmail.com

Victor Chen, Su Ruan
CReSTIC

IUT Troyes, Université de Reims
9 rue de Québec, 10026 Troyes Cedex

{victor.chen-yuan, su ruan}@univ-reims.fr

RESUME. Nous nous intéressons ici à la segmentation d'objets d'intérêt contenus dans les images. Le but applicatif visé concerne le détourage d'images qui connaît un engouement croissant ces dernières années dans le multimédia. Nous proposons dans cette communication d'utiliser la méthode de coupe de graphes combinée à la technique de Lazy Snapping dont l'approche interactive permet de segmenter efficacement les images depuis sa résolution grossière au raffinement. Dans ce travail, afin de simplifier la problématique d'étiquetage de l'objet à segmenter et obtenir rapidement le rendu de la segmentation, nous nous sommes appuyés sur le graphe coupé de Boykov-Jolly qui, à travers leur modèle, permet de résoudre pour beaucoup d'utilisateurs la problématique de découpe d'images à des fins personnelles ou professionnelles. La procédure de découpage de graphes est organisée autour de la minimisation de la fonction d'énergie par l'intermédiaire de l'algorithme Max Flow/Min Cut. Les résultats de segmentation sont illustrés à partir de la base de données de Berkeley.

MOTS-CLES : coupe de graphes, segmentation, labellisation, Max Flow/Min Cut, détourage d'images.

SUMMARY. We are interesting on interest objects segmentation contained in the images. The focused application is dealing with the image cutting which knows an increasing infatuation these last years in the multimedia. We propose in this paper to use the combination of graph cuts method and Lazy Snapping technique which the interactive approach can treat effectively the coarse to fine processing. In this framework, in order to simplify the object labelling problem to segment and to obtain rapidly the rendering segmentation, we have recourse to the Boykov-Jolly graph cut model which can solve to many users the cutting image problem aiming personal or professional activity. The graph cutting process is organized about energy function minimizing by means Max Flow/Min Cut algorithm. Segmentation results are illustrated from data base of Berkeley.

KEY WORDS : graph cuts, image segmentation, labeling, Max Flow/Min Cut, image cutting.

1. Introduction

La segmentation interactive d'images est devenue cette dernière décennie un axe de recherche très actif et en plein essor de développement en particulier dans le multimédia. On recourt de plus en plus fréquemment à ce type d'approche pour contourner le problème inhérent à la segmentation automatique qui donne rarement des résultats parfaits. De nouvelles techniques d'approche apparaissent inévitablement qui, au lieu d'effectuer une segmentation globale et non supervisée sur l'image, se sont appropriées du créneau pour réaliser une segmentation avec une connaissance a priori, puisque dans la plupart des cas de traitements, l'objet à segmenter est en général connu à l'avance. C'est le cas par exemple d'un applicatif concernant le détourage d'image qui connaît un engouement croissant : réalisation de photomontages [1], changement d'arrière plan d'une scène [2], incrustation d'objets via diffusion d'image (image editing) [3], autocollage [4]. Ces résultats peuvent être obtenus à partir d'une plate forme qui utilise la combinaison des méthodes de coupe de graphes qui sépare l'objet d'intérêt du fond de l'image et de la technique de Lazy Snapping [5] qui permet de marquer les objets à segmenter depuis sa résolution grossière au raffinement jusqu'à l'obtention d'un rendu parfait. Un grand nombre d'algorithmes de coupes de graphes ont émergé afin de solutionner divers problèmes de traitement d'images (reconnaissance de formes, stéréovision, analyse de scènes,...). Les travaux de Shi et Malik [6] sur les coupes normalisées donnaient des résultats de segmentation satisfaisants pour des volumes de données généralement faibles (des tailles d'images inférieures à 128x128). La pertinence de l'approche concerne cependant les résultats qui peuvent être obtenus au moyen uniquement des calculs sur les valeurs propres issues d'un système d'équations de la forme de $Wx = \lambda Dx$ où W représente la matrice de similarité qui résulte de la mesure du poids des arcs entre les pixels voisins et D la matrice diagonale. La bipartition de l'image est obtenue en prenant le vecteur propre correspondant à la deuxième plus petite valeur propre du système. Cette technique qui favorise le regroupement des petits ensembles de nœuds isolés dans le graphe est exploitée pour réaliser la segmentation des images. Le coût élevé de temps de calculs reste néanmoins un handicap pour un traitement en temps réel. En parallèle, les travaux de Boykov-Jolly [7, 8] qui se distinguent par leur

formalisme interactif, permettent l'extraction facile des pixels d'intérêt (objet ou arrière plan), en raison de l'implantation des germes par l'utilisateur. Dans le cadre du traitement d'images, le modèle (S-T) coupe utilise l'algorithme du flot maximum (Max Flow) de Ford et Fulkerson qui recherche une chaîne augmentante de la source (S) vers le puits (T) tels que les arcs de la chaîne ne soient pas saturés. Il est montré dans cette procédure que, pour tout couple (S, T) de sommets du graphe et pour toute pondération positive, la valeur maximum du flot de (S) à (T) est égale aussi à la capacité minimum d'une coupe séparant (S) à (T). L'utilisation de ce résultat permet d'envisager une partition binaire optimale du graphe en minimisant une fonction d'énergie. Par ailleurs, en associant de la connaissance a priori au graphe, nous pouvons espérer obtenir une extraction d'objets d'intérêt suivant un processus de regroupement perceptuel.

L'objectif de notre travail est de proposer une approche de segmentation rapide et apte à isoler un objet quelconque de l'image à l'aide de la souris qui sélectionne les objets à mettre en évidence. L'application que nous visons concerne la découpe intelligente d'objets d'intérêt et son incrustation éventuelle dans une autre image via le mécanisme de poisson image editing [3].

La communication est organisée comme suit. Dans le paragraphe 2, nous rappelons les étapes de segmentation par coupe de graphe. La mise en œuvre ainsi que l'implantation du programme sont abordées dans le paragraphe 3. Les résultats expérimentaux obtenus sont présentés dans le paragraphe 4. Enfin, la section 5 conclut cet article et discute des perspectives que laisse envisager la méthode proposée.

2. Segmentation d'objets par la coupe de graphes

2. 1. Labellisation des pixels

La segmentation d'un objet contenu dans une image peut être formulée comme un problème de labellisation. Etant donné une suite de pixels notée $S = \{S_i \mid i=1, \dots, n\}$ à segmenter, une suite de rendus de la segmentation notée $A = \{A_i \mid i=1, \dots, n\}$ et une suite de labels notée $L = \{L_j \mid j=1, \dots, m\}$, la procédure de labellisation consiste à assigner une suite de rendus de la segmentation à une suite de labels telle que $A_i \in L$. Dans beaucoup de cas applicatifs qui visent à la binarisation d'une

image, le nombre de labellisation est égal à 2, soit $L = \{0, 1\}$ où 1 correspond à l'objet d'intérêt et 0 à l'arrière plan de l'image. La stratégie pour classifier les pixels implique une recherche de solution optimale sur les suites de labels. Celle-ci se ramène à un problème de minimisation d'une fonction d'énergie qui est décrite au paragraphe ci-après.

2. 2. Minimisation de la fonction d'énergie

L'objectif visé consiste à trouver une coupe optimale de graphes donnant une segmentation satisfaisante. Ce résultat implique à minimiser une fonction objective de la forme :

$$E(A) = \lambda \sum_{p \in P} E_p(A_p) + \sum_{(p,q) \in N} E_{pq} \delta_{A_p, A_q} \quad (1)$$

$$\text{avec } \delta_{A_p \neq A_q} = \begin{cases} 1 & \text{si } A_p \neq A_q \\ 0 & \text{sinon} \end{cases}$$

où $P \in \mathbb{R}^N$, $N \subset P \times P$ est un voisinage sur les pixels, le terme $E_p(A_p)$ est une fonction mesurant le coût d'attribution d'un label A_p à un pixel p , le terme $E_{pq}(A_p, A_q)$ mesure le coût d'assignation les labels A_p et A_q aux pixels adjacents p et q régularisation et le paramètre de pondération $\lambda \in [0,1]$ qui contrôle l'importance relative entre les termes de vraisemblance et d'a priori qui composent l'énergie.

Dans le cas d'une coupe de graphes, la minimisation de l'énergie $E(A)$ pourra être obtenue de façon optimale si la condition de régularité [8] présentée ci-après est respectée :

$$E_{p,q}(0,0) + E_{p,q}(1,1) \leq E_{p,q}(0,1) + E_{p,q}(1,0).$$

2. 3. Formulation du graphe dans l'image

Un grand nombre d'auteurs considèrent que la segmentation d'images peut être vue comme un problème de labellisation [5, 6]. Nous décrivons dans ce paragraphe la construction d'un graphe à partir d'une image illustrée à la figure 1. Chaque pixel de l'image est vu comme un nœud dans un graphe. Les arêtes E sont constituées de nœuds V adjacents et de leur capacité W dont la valeur dépend de la similarité entre deux pixels voisins.

Cette description peut être formalisée selon un graphe pondéré que l'on note $G = (V, E, W)$. Une grande (faible) capacité correspond à une faible

(grande) similarité entre les pixels voisins. Dans le cadre du modèle établi par Boykov et Jolly [7, 8], deux nœuds terminaux particuliers représentant respectivement la classe objet appelée source (S) et la classe arrière plan appelée cible (T) terminent le graphe. Suivant les labels assignés par l'utilisateur lors du marquage des germes, les coûts des capacités sont calculés à l'aide des modèles définis pour l'objet ou l'arrière plan de l'image comme le décrit le paragraphe suivant.

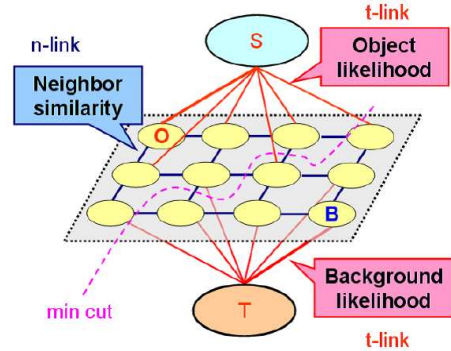


Figure 1 : Illustration d'une coupe de graphe sur une image de dimension 3x4 [9]

2. 4. Calcul des capacités pour une configuration terminale

En raison des contraintes de points de germes définis par l'utilisateur, l'attribution des poids initiaux aux arêtes est tributaire du point de germe affecté à la classe objet ("obj") ou à la classe arrière plan ("bkg") de l'image à segmenter.

- Calcul de la capacité $W_{\text{classe_obj}}$ liée à la classe objet

$$W_{\text{classe_obj}} = \begin{cases} K & \text{si } I \in I_{\text{obj}} \\ 0 & \text{si } I \in I_{\text{fond}} \\ \lambda \cdot W_{\text{obj}}(I) & \text{les autres cas} \end{cases} \quad (2)$$

- Calcul de la capacité $W_{\text{classe_bkg}}$ liée à la classe arrière plan

$$W_{\text{classe_bkg}} = \begin{cases} 0 & \text{si } I \in I_{\text{obj}} \\ K & \text{si } I \in I_{\text{fond}} \\ \lambda \cdot W_{\text{bkg}}(I) & \text{les autres cas} \end{cases} \quad (3)$$

$$\text{avec } \begin{cases} W_{\text{obj}}(I) = -\ln(\Pr(I/"\text{obj}")) \\ W_{\text{bkg}}(I) = -\ln(\Pr(I/"\text{bkg}")) \end{cases}$$

$$\text{et } K = 1 + \max_{p \in P} \sum_{q: \{p,q\} \in N} W_{\{p,q\}} \quad (4)$$

où $\Pr(I/"obj")$ et $\Pr(I/"bkg")$ sont des probabilités conditionnelles qui peuvent être déduites des histogrammes normalisés (Figure 2) de l'objet d'intérêt et de l'arrière plan de l'image.

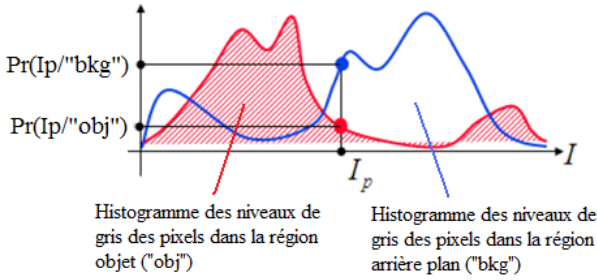


Figure 2 : Répartition des niveaux de gris des pixels suivant la classe objet et la classe arrière plan.

2. 5. Calcul des capacités pour une configuration de voisinage

La capacité des arêtes entre les nœuds du graphe est calculée à partir de la mesure de similarité W des pixels voisins en calculant le score fourni par :

$$W_{\{p,q\}} \propto \exp\left(-\frac{(I_p - I_q)^2}{2\sigma^2}\right) \cdot \frac{1}{\text{dist}(p,q)} \quad (5)$$

où I_p (resp. I_q) est l'intensité du pixel positionnée en p (resp. q), $\text{dist}(p,q)$ la distance Euclidienne entre les positions p et q et σ le paramètre qui contrôle l'affinité de teinte des pixels.

L'intérêt de cette mesure réside sur l'apport d'informations que nous fournissent le premier terme qui permet de discriminer les objets non semblables par la différence de teintes et le deuxième terme qui évalue les proches voisins des pixels par leur distance.

3. Implantation

3. 1. Apprentissage

A l'aide de la souris, l'utilisateur marque interactivement des pixels I_{obj} que nous considérons appartenir aux labels objet et d'autres pixels I_{bkg} que nous considérons appartenir aux

labels fond ou arrière plan. Dans le cadre d'un processus de coupe de graphe, ces marquages jouent un rôle double : ils sont utilisés à la fois comme germe de segmentation et comme base d'apprentissage pour l'attribution des poids à l'énergie de vraisemblance E_p via la construction des histogrammes décrits précédemment.

3. 2. Marquage suivant la technique de Lazy Snapping

L'imprécision du marquage précédent en particulier au niveau des régions de faible contraste ou dans les zones texturées nous a incité à conjuguer avec la technique de Lazy Snapping [5] qui propose une cartographie flexible et intuitive des pixels en opérant sur les régions. Cette technique nécessite néanmoins une rectification de la fonction objective. En reprenant les notations décrites précédemment, nous déterminons :

- Energie de vraisemblance :

Cette énergie $E_p(\cdot)$ a pour but de décoder la similarité des couleurs présentes sur des nœuds en indiquant si ceux-ci appartiennent à l'objet ou l'arrière plan. A partir des germes implantés, on fait appel à l'algorithme de K-means pour réaliser un partitionnement préalable de façon à faciliter l'évaluation de cette énergie.

$$\begin{cases} E_p(A_i = 1) = 0 & E_p(A_i = 0) = \infty & \forall i \in \text{objet ou (O)} \\ E_p(A_i = 1) = \infty & E_p(A_i = 0) = 0 & \forall i \in \text{fond ou (F)} \\ E_p(A_i = 1) = \frac{d_i^O}{d_i^O + d_i^F} & E_p(A_i = 0) = \frac{d_i^F}{d_i^O + d_i^F} & \forall i \in \text{incertain} \end{cases}$$

$$\text{où } d_i^O = \min_n \|C(i) - K_n^O\| \text{ et } d_i^F = \min_m \|C(i) - K_m^F\| \quad (6)$$

avec

$\{K_n^O\}$ et $\{K_m^F\}$ sont les partitions de la classe objet et de classe arrière plan obtenues respectivement par l'algorithme de K-means et $C(i)$ est la couleur du pixel positionné en i .

Nous illustrons sur la figure 3 les différentes traces de marquage qui participent à la réalisation de la coupe de graphes qui délimite la classe objet de la classe arrière plan.

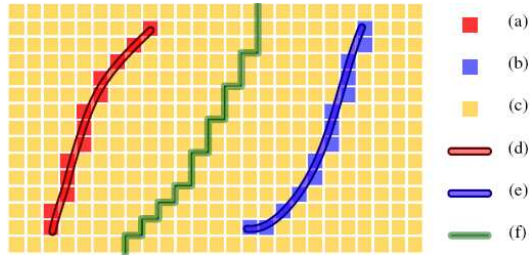


Figure 3 : Description du marquage des objets [5]

- (a) germes des objets ; (b) germes d'arrière plan ;
 (c) région incertaine ;
 (d) marqueur d'objet ; (e) marqueur d'arrière plan ;
 (f) frontière obtenue par la méthode.

- Energie d'a priori

On utilise cette énergie $E_{p,q}(\cdot)$ pour décoder le gradient des pixels colorés au voisinage de l'objet. Le gradient des couleurs entre deux nœuds i et j est calculé à partir de la relation suivante :

$$E_p(A_i, A_j) = |A_i - A_j| g(C(i, j)) \quad (7)$$

avec

$$g(\xi) = \frac{1}{1 + \xi} \text{ et } C(i, j) = \|C(i) - C(j)\|^2$$

où $C(k)$ représente la couleur du pixel à la position k .

3. 3. Classification des régions

Dans le but d'améliorer le temps de traitement et d'augmenter la finesse de la coupe de graphe, une sur-segmentation de l'image est opérée de façon à préserver les petites régions. Celles-ci seront reclassifiées à l'aide de l'algorithme de ligne de partage des eaux (watershed) [10]. En modulant l'inondation des minima locaux, il est ainsi possible de regrouper les petites régions avoisinantes pour créer des nouveaux nœuds à base de régions (figure 4) et les arcs qui les relient.

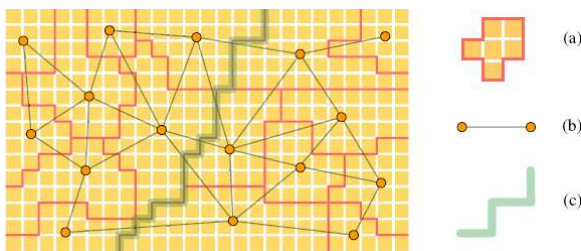


Figure 4 : Cartographie de frontières par Lazy Snapping [5]

- (a) régions obtenues par la technique de la ligne partage des eaux ;
 (b) représentation de nœuds et d'arcs ;
 (c) rendu d'une coupe de graphes obtenue.

3. 4. Segmentation d'image par minimisation d'énergie

La procédure de segmentation envisagée ici consiste maintenant à reformuler les énergies de vraisemblance et d'a priori de la fonction objective décrite dans le paragraphe 3. 2.

Nous commençons la mise à jour de la procédure en remplaçant les pixels par des nœuds qui correspondent à des petites régions segmentées par l'algorithme de ligne de partage des eaux. Ensuite, nous calculons la couleur $C(i)$ de la région i comme étant la moyenne des couleurs des régions présentant des similarités en termes de couleur. Afin d'extraire l'ensemble des arêtes, nous évaluons enfin le coefficient $C(i, j)$ qui lie les contrastes de couleur entre deux régions i et j .

Pour obtenir une segmentation optimale de l'image, nous recourons à l'algorithme MaxFlow/MinCut [5] qui est actuellement l'outil le plus efficace pour départager le graphe en deux sous ensembles en adéquation avec les contraintes établies par les marqueurs.

4. Résultats expérimentaux

Nous présentons dans ce paragraphe les résultats obtenus par l'algorithme de coupe de graphes qui utilise des contraintes rigides au moyen des points de germes sélectionnés par l'utilisateur. La méthode est testée sur une banque d'images de Berkeley afin de mieux cerner l'applicatif du résultat de la segmentation qui est le détournement d'images. Nous montrons à la figure 5 une illustration de différentes séquences d'extraction d'un objet d'intérêt. Les marqueurs rouges sont des points de germes pour obtenir l'objet d'intérêt tandis que les marqueurs bleus sont des points de germes pour réaliser l'arrière plan de l'image.

Nous observons durant la phase d'apprentissage que les choix et les positions des points de germes influencent grandement les résultats de la segmentation.

Cette dépendance peut être due à des erreurs de probabilités fournies par les histogrammes des points de germes qui ne sont modélisés que par deux gaussiennes. Sur le plan pratique, les expériences sont menées sur un ordinateur portable

muni d'un processeur Intel Dual Core2 de 2,26 GHz et d'une RAM de 4Go. Le temps de traitements utilisé pour segmenter les images de taille supérieure à 400x350 reste inférieur à 7 secondes. Ces résultats illustrent l'efficacité de la méthode pour extraire les objets d'intérêt de formes très variables. Pour cette étude, nous avons fixé a priori les valeurs de $\lambda=0.3$ et de $\sigma=4$. Cette performance permet dans le développement futur à la procédure d'être opérationnelle pour un détourage d'images en temps réel. Cependant, la segmentation présente encore à ce jour des défauts superficiels de coupe qui se caractérisent par des contours relativement perturbés au niveau des régions texturées ou de faible contraste. Nous comptons néanmoins résoudre ce problème en recourant aux transformées en ondelettes qui, grâce à ses propriétés de filtrage, atténueront ces perturbations.

Nous avons proposé dans cette communication une méthode de segmentation utilisant l'algorithme de coupe de graphes contrainte par une connaissance a priori. Cette technique d'approche a montré des intérêts réels à la fois sur l'efficacité et la rapidité du rendu de la segmentation.

Quelques imprécisions de segmentation sont encore observées dans les régions de faible contraste et dans les zones texturées. Nous comptons dans les travaux futurs contourner ces problèmes et améliorer la qualité de la segmentation en faisant appel aux schémas de décomposition en ondelettes [9] qui vont lisser l'image dans les basses fréquences du côté de la source S et atténuer les caractéristiques locales de texture en les filtrant dans les hautes fréquences du côté de puits T. Cette opération de lissage est pour nous quoiqu'il soit une nécessité pour mettre en application ce détourage dans un processus d'incrustation d'images.

5. Conclusions et perspectives

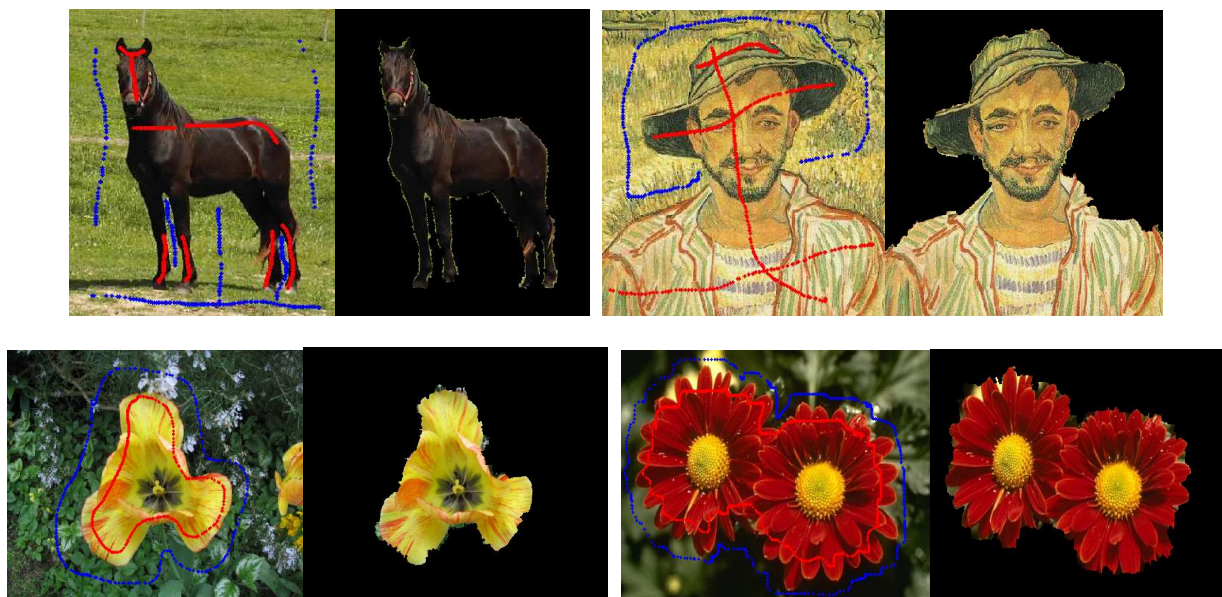


Figure 5 : Illustration des marqueurs (couleur rouge : objet, couleur bleu : arrière-plan) et des rendus de segmentation obtenus sur des images fortement texturées.

Références

- [1] C. Rother, V. Kolmogorov, A. Blake, Grabcut interaction foreground extraction using iterated graphics, ACM transactions on Graphics, 23(3), pp. 309-314, 2004.
- [2] J. Sum, W. Zhang, X. Tang, H. Y. Shum, Background Cut, Proceedings in 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, vol. 3954, pp. 628-641.

- [3] P. Perez, M. Gangnet, A. Blake, Poisson image editing, *ACM Transactions on Graphics*, 22(3), pp. 313-318, 2003.
- [4] C. Rother, L. Bordeaux, Y. Hamadi, A. Blake, Auto collage, *ACM Transactions on Graphics*, vol. 25, pp. 847-852, July 2006.
- [5] Y. Li, J. Sun, C.K. Tang, H. Y. Shum, Lazy Snapping, *ACM Transactions on Graphics*, vol. 23, pp. 303-308, 2004.
- [6] J. Shi, J. Malik, Normalized cuts and image segmentation, *IEEE PAMI*, 22(8), pp. 888-905, August 2000.
- [7] Y. Boykov, M. P. Jolly, Interactive graph cuts for optimal boundary and region segmentation, in *ICCV*, volume I, pp. 105-112, 2001.
- [8] Y. Boykov, V. Kolmogorov, An experimental comparison of min-cut/max flow algorithms for energy minimization in vision, *IEEE PAMI*, 26(9), pp. 1124-1137, 2004.
- [9] K. Fukuda, T. Takiguchi, Y. Ariki, Graph cuts using local texture features of wavelet coefficient for image segmentation, *IEEE International Conference on Multimedia and Expo*, June 23-26, 2008, Hannover, Germany.
- [10] L. Vincent, P. Soille, Watershed in digital spaces: an efficient algorithm based on immersion simulations, *IEEE PAMI*, 13(6), pp. 583-598, June 1991.